Genomics and Personalised Medicine:

Diagnostics, Deep Phenotyping and Pharmacogenomics in Cohorts of Rare Disease Patients

A thesis submitted for the degree of Doctor of Philosophy

University College London

Dr Joanna Kenny

Genetics and Genomic Medicine

UCL Great Ormond Street Institute of Child Health

December 2018

Declaration of originality

I, Joanna Kenny, confirm that the research presented here is my own. When information has been derived from other sources, this is indicated in the thesis. Any assistance from other parties is acknowledged and research was done in accordance with the academic and ethical rules of UCL.

Joanna Kenny

December 2018

This thesis is dedicated to the memory of Professor Maria Bitner-Glindzicz, my mentor and my friend. She was a star who took others with her as she rose and a giant on whose shoulders I stood. Without her, I would probably not be a clinical geneticist and much of what I have achieved in my career is thanks to her. I will be forever grateful for the opportunities and support she gave me, the kindness she showed to me and how much she believed in me. She is so much missed.

Abstract

Introduction

Clinical genetics is a rapidly evolving specialty. Diagnostic testing is moving to more agnostic screening of whole exomes and genomes. This leads to challenges in variant interpretation, variants of unknown clinical significance and secondary findings. This study looked at the utilisation of phenotypic and whole genome sequence (WGS) data to increase understanding of rare disease and pharmacogenomics.

Methods

WGS was used to attempt to clarify the molecular basis of two cases of Bardet Biedl syndrome (BBS) by filtering variants in ingenuity, using bioinformatics methods to identify copy number losses or gains and examining known ciliopathy and other genes. The feasibility and accuracy of extracting pharmacogenomics data from WGS was assessed and results validated in a cohort of 84 people. Diplotypes or genotypes were determined for 18 actionable pharmacogenes and prescribing guidance was prepared. Diplotypes were also determined for 43 additional pharmacogenes. Results were validated using a commercial pharmacogenomics assay

Results

Candidate variants were identified in a number of BBS, ciliopathy and non-ciliopathy genes in each case. However no definitively pathogenic biallelic variants were identified. All patients had at least one actionable pharmacogenomic variant that could result in a change in drug dose or monitoring. The mean number of variants per patient was 3.8. Haplotype frequency data for the actionable pharmacogene haplotypes was not significantly different to population data. The majority of disagreements between WGS and SNP data were caused by poor clustering of samples during SNP genotyping resulting in ambiguous calls. Overall the discordance between WGS and SNP data for all tested pharmacogenes was less than 0.01%.

Conclusions

Singleton WGS was not sufficient to identify the cause of BBS and further work is required to identify this. Possible reasons include missed variants, variants in novel genes, deep intronic variants or non-Mendelian modes of inheritance. Pharmacogenomic variants can be identified using WGS with a similar success rate to a current commercial method, but has additional advantages including the ability to review data as pharmacogenomic prescribing guidelines change. There are many challenges to introducing population-level pharmacogenomic prescribing and many potential benefits.

Impact Statement

This study looks at the changing landscape of clinical genetics, how advances in diagnostics can be integrated into clinical practice and how additional information, such as pharmacogenomic prescribing information, can be obtained to maximise value for both patients and clinicians.

The confirmation that pharmacogenomic data can be extracted from whole genome sequences and prescribing advice generated is of particular benefit. As whole genome sequencing is introduced to the diagnostic test directory in the near future, more patients will have whole genome sequences available for analysis. The provision of pharmacogenomic prescribing advice has the potential to significantly reduce adverse drug effects, which will not only result in more efficacious treatment and a reduction of morbidity and mortality for patients, but will have many other benefits including cost savings for the NHS and wider society.

This thesis explores the issues in introducing pharmacogenomic testing to a system such as the NHS and in particular, has considered the challenges, risks and benefits of doing so. This should be of interest to bodies such as Genomics England, who are considering the introduction of such testing, and also of wider public interest, given the potential risks such as discrimination on ethnic or other grounds that are possible. The studies undertaken have also highlighted the need for further research and development of automated pipelines for the calling of pharmacogenomic variants and provision of prescribing advice.

Diagnostic data are of course of interest to the particular patients, but the wider discussion around risks and benefits is of interest to anyone who is involved in genetic testing, be they a patient or a professional. The development of the phenotyping database has already had applications for local researchers and is being integrated with analysis software for multiomics, something that has the potential to be of interest to researchers worldwide.

More widely, this thesis is relevant to anyone who has an interest in how clinical genetics is likely to develop in the coming years, and the challenges and advantages that will arise.

Table of Contents

List of Figures

List of Tables

Acknowledgements

Firstly, thanks must go to my supervisors, Dr Chiara Bacchelli and Professor Philip Beales, for giving me the opportunity to study with them. I am grateful for their support, their ideas, their encouragement and their belief in me. In particular, Dr Bacchelli set aside time to speak regularly which was invaluable and was also endlessly supportive through crises, both real and imagined.

Many people assisted when I was learning new techniques: Dr Lamia Boukhibar, Dr Rosalind Davies, Dr Elizabeth Forsythe, Dr Grace Freke, Dr Rasha Nour, Dr Louise Ocaka, Ms Jasmine Risvi and Dr Polona Stabej-De Quesne. The bioinformatics team of Dr Andrey Gagunashvili, Dr Chela James, Dr Nital Jani, Dr Federico Minnecci, Dr Georg Otto and Dr Xueting Wang were of great assistance, especially Dr Andrey Gagunashvili for help with Astrolabe and Lumpy. Many people provided advice and support, especially Dr Daniel Kelberman, Dr Suzanne Drury, Dr Elizabeth Forsythe and Professor Maria Bitner-Glindzicz. Miss Jasmine Gratton and Miss Jasmine Risvi were of great help in the lab, with sample preparation and pharmacogenomic genotyping respectively. Ms Deborah Ridout advised about appropriate statistical tests. I am grateful to the collaborators who made this and other projects possible: Professor Maria Bitner-Glindzicz, Professor Gudrun Moore, Dr Miho Ishido, Dr Jochen Kammermeier, Professor Lucy Wedderburn, Dr Claire Deakin, Dr Tomi Peltola, Professor Sami Kaski and the staff of Congenica Ltd. Special proofreading thanks are due to Christopher Bowen, Rosie Hennigan, Anne Marie Herron, Paul, Ann, Laura and David Kenny and Kathleen Murphy. I am also grateful to the many lovely librarians I have encountered over the years.

My parents were supportive as they always have been. I am very grateful to have been brought up in a house where life-long learning was encouraged, and where we were taught to believe that we could achieve anything. My darling children, Saoirse and Cara, were entirely supportive and uncomplaining of my absences and distractions. None of this would have been achieved without the love and support of my husband Christopher, who kept our lives running smoothly and was unfailingly positive and encouraging in addition.

Finally, special thanks must go to Betty and Phyllis Reilly, whose legacy made this possible. They didn’t find out what their gift was used for, but I know they would have been pleased and proud.

List of Abbreviations

A(2)	African (2) allele of *G6PD*
AA	Amino acid
AAC	Anthracycline-associated cardiotoxicity
ACE	Angiotensin-converting enzyme
ACGS	Association for Clinical Genomic Science
ACMG	American College of Medical Genetics
ACMGG	American College of Medical Genetics and Genomics
ADE	Adverse drug effect
ADHD	Attention deficit hyperactivity disorder
ADPKD	Autosomal dominant polycystic kidney disease
ADR	Adverse drug reaction
AIF	Assay information file
ALMS	Alstrom syndrome
APC	Activated protein C
ARPKD	Autosomal recessive polycystic kidney disease
ARVD	Arrhythmogenic right ventricular dysplasia
ASD	Atrial septal defect
ASO	Allele specific oligonucleotides
ATRA	All-trans retinoic acid
AV	Atrio-ventricular
BAM	Binary alignment map
BBS	Bardet-Biedl syndrome
BLAST	Basic Local Alignment Search Tool
BMS	Bristol Myers Squibb
BNF	British National Formulary
bp	Base pairs
BR	Broad range
BWA	Burrows Wheeler alignment
CADD	Combined annotation dependent depletion
cDNA	Complementary DNA
CED	Cranioectodermal dysplasia
CF	Cystic fibrosis
CFTR	Cystic fibrosis transmembrane receptor
CGH	Comparative genome hybridisation
CI	Confidence interval
CKD	Chronic kidney disease
ClinVar	Database of clinical variants
CNS	Central nervous system
CNV	Copy number variant
COACH	Cerebellar vermis hypo/aplasia, oligophrenia, congenital ataxia, ocular coloboma, hepatic fibrosis
CPIC	Clinical Pharmacogenetics Implementation Consortium
CPNDS	Canadian Pharmacogenomics Network for Drug Safety
dbSNP	Single Nucleotide Polymorphism Database
DDD	Deciphering Developmental Diseases
ddNTPs	Di-deoxynucleotidetriphosphates
DECIPHER	DatabasE of genomiC varIation and Phenotype in Humans using Ensembl Resources
DNA	Deoxyribonucleic acid
dNTPs	Deoxynucleotidetriphosphates
DPWG	Royal Dutch Pharmacists Association - Pharmacogenetics Working Group

DSM	Diagnostic and Statistical Manual of Mental Disorders
EDTA	Ethylene diamine tetra acetic acid
eGFR	Estimated glomerular filtration rate
EHR	Electronic health record
EM	Extensive metaboliser
EMA	European Medicines Agency
ePGA	Electronic Pharmacogenomics Assistant
ERG	Electroretinography
ESRF	End-stage renal failure
EU	European Union
EURO-WABB	European Wolfram, Alstrøm and Bardet-Biedl syndrome study
EVC	Ellis van Creveld syndrome
ExAC	Exome Aggregation Consortium
FDA	Food and Drug Administration
FVL	Factor V Leiden
GATK	Genome Analysis Toolkit
gDNA	Genomic DNA
GDPR	General Data Protection Regulation
GI	Gastro-intestinal
GnomAD	Genome Aggregation Database
GO	Gene Ontology
GORD	Gastro-oesophageal reflux disease
GOSH	Great Ormond Street Hospital NHS Foundation Trust
GP	General practice
GPCR	G protein-coupled receptor
GRCh37	Genome Reference Consortium human genome (build 37)
GRCh38	Genome Reference Consortium human genome (build 38)
GWAS	Genome-wide association study
HCSC	Health Canada Santé Canada
HGNC	HUGO gene nomenclature committee
HGMD	Human Gene Mutation Database
HGVS	Human Genome Variation Society
HH	Hedgehog
HIV	Human immunodeficiency virus
HLA	Human leucocyte antigen
HPO	Human Phenotype Ontology
HUGO	Human Genome Organisation
hURECs	Human urine-derived renal epithelial cells
IBD	Inflammatory bowel disease
IC	Imprinting centre
ICD	International Statistical Classification of Diseases and Related Health Problems
ICH	Institute of Child Health
ID	Identification
IDHS	Data Safe Haven, University College London
IFT	Intraflagellar transport
IGV	Integrative genomics viewer
IHTSDO	International Health Terminology Standards Development Organisation
IM	Intermediate metaboliser
indel	insertion/deletion
INR	International normalised ratio
IPEX	X-linked immune dysregulation, polyendocrinopathy and enteropathy syndrome
iPS	Induced pluripotent stem cells

IUGR	Intra-uterine growth retardation/restriction
IVA	Ingenuity Variant Analysis™
IWPC	International warfarin pharmacogenetics consortium
JATD	Jeune asphyxiating thoracic dystrophy
JBTS	Joubert syndrome
JDM	Juvenile dermatomyositis
JDRG	Juvenile Dermatomyositis Research Group
JSON	Java Script Object Notation
kg	Kilogram
LCA	Leber congenital amaurosis
LOINC	Logical Observation Identifiers Names and Codes
LRWGS	Long-read whole genome sequencing
MAF	Mean allele frequency
MELAS	Mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes
MERRF	Myoclonic epilepsy, ragged red fibres
MESH	Medical Subject Headings Thesaurus
mg	Milligram
MHC	Major histocompatibility complex
MKKS	McKusick-Kaufman syndrome
MKS	Meckel syndrome
MNGIE	Mitochondrial neurogastrointestinal encephalopathy syndrome
MORM	Mental retardation, truncal obesity, retinal dystrophy, and micropenis syndrome
MPO	Mammalian Phenotype Ontology
MRC	Medical Research Council
MS	Multiple sclerosis
MS Access	Microsoft Access
NADPH	Nicotinamide adenine dinucleotide phosphate (reduced form)
NGHRI	National Human Genome Research Institute
NGS	Next-generation sequencing
NHGRI	National Institutes of Health Genome Research Institute
NHLBI	National Heart, Lung and Blood Institute
NHLBI-ESP	National Heart, Lung and Blood Institute exome sequencing project
NHS	National Health Service
NIH	National Institutes of Health
NLM	National Library of Medicine
NM	Normal metaboliser
NPH	Nephronophthisis
NRI	Noradrenaline (norepinephrine) reuptake inhibitor
NRTI	Nucleoside analogue reverse transcriptase inhibitor
NSAIDs	Non-steroidal anti-inflammatory drugs
NSCLC	Non-small cell lung cancer
NTC	No template control
OCD	Obsessive compulsive disorder
OFDS	Oro-facio-digital syndrome
OMIM	Online Mendelian Inheritance in Man
PATO	Phenotype, Attribute, and Trait Ontology
PBMC	Peripheral blood mononuclear cell
PCD	Primary ciliary dyskinesia
PCR	Polymerase chain reaction
PDF	Portable document format
PETIT	Patients with Early-onseT Intestinal inflammaTion (PETIT) Study
PGRN	Pharmacogenomics Research Network

PGx	Pharmacogenomics
PharmGKB	Pharmacogenomics Knowledge Base
PharmVar	Pharmacogene Variation Consortium
PhenIX	Phenotypic Interpretation of Exomes
PI	Principal investigator
PM	Poor metaboliser
PPCA	Pigmented paravenous chorioretinal atrophy
PPI	Proton pump inhibitor
PRO	Professional society
PROVEAN	Protein variation effect analyser
QALY	Quality-adjusted life year
REC	Research ethics committee
RIMA	Reversible inhibitor of monoamine oxidase A
RNA	Ribonucleic acid
RNAseq	RNA sequencing
RP	Retinitis pigmentosa
RSV	Respiratory syncytial virus
RTI	Respiratory tract infection
RTPCR	Reverse transcription polymerase chain reaction
RVOTT	Right ventricular outflow tract tachycardia
SAM	Sequence alignment tools
SCA	Spino-cerebellar ataxia
SCAR	Severe cutaneous adverse reactions
SCID	Severe combined immunodeficiency
SGB	Simpson-Golabi-Behmel syndrome
SIFT	Sorting Intolerant From Tolerant
SJS	Stevens-Johnson Syndrome
SLS	Senior-Loken syndrome
SNHL	Sensorineural hearing loss
SNOMED-CT	Systemised Nomenclature of Medicine- Clinical Terms
SNOP	Systemized Nomenclature of Pathology
SNP	Single nucleotide polymorphism
SNRI	Serotonin noradrenaline (norepinephrine) reuptake inhibitor
SNV	Single nucleotide variant
SRS	Silver-Russell syndrome
SSRI	Selective serotonin reuptake inhibitor
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins
SVR	Sustained virologic response
SVT	Supraventricular tachycardia
TBE	Tris-Borate-EDTA
TCS	Treacher Collins syndrome
TCSDD	Tuft's Centre for the Study of Drug Development
TEN	Toxic epidermal necrolysis
THE	Trichohepatoenteric syndrome
TNF	Tumour necrosis factor
TOF	Tetralogy of Fallot
UC	Ulcerative colitis
UCL	University College London
UCSC	University of California Santa Cruz
UK	United Kingdom
UM	Ultra-rapid metaboliser
UMLS	Unified Medical Language System

UPD	Uniparental disomy
U-PGx	Ubiquitous pharmacogenomics
USA	United States of America
UTI	Urinary tract infection
UTR	Untranslated region
VACTERL	Vertebral defects, Anal atresia, Cardiac malformations, Tracheoesophageal fistula with Esophageal atresia, Radial or renal dysplasia, Limb anomalies
VCF	Variant call format
VEOIBD	Very early onset inflammatory bowel disease
VIP	Very important pharmacogene
VHL	Von Hippel Lindau syndrome
VSD	Ventriculoseptal defect
VT	Ventricular tachycardia
VUS	Variant(s) of unknown significance
WAGR	Wilm's tumour, aniridia, genital anomalies, retardation
WES	Whole exome sequencing
WGS	Whole genome sequencing
WHO	World Health Organisation
XML	Extensible Markup Language

Chapter 1 Introduction

1.1 Clinical genetics in the NHS

1.1.1 What is clinical genetics?

Clinical genetics is the specialty that deals with the diagnosis of genetic disease. It has been a National Health Service (NHS) clinical service since 1980, when the Department of Health and Social Services established it as a stand-alone specialty. It was, of course, practised in the UK long before this, for example in the laboratory of Professor Lionel Penrose, who worked on phenylketonuria and tuberous sclerosis(1). Limited testing options existed initially, although karyotyping, Sanger sequencing and linkage mapping were available, and most diagnoses were clinical, with empirical recurrence risks being given. Testing was done primarily to confirm a clinical diagnosis determined in advance. It is a specialty that has changed and grown very rapidly since then, both because of the development of new testing methods and the discovery that genetics has a role to play in every specialty and most disease types. Clinical genetics differs from other specialties in several regards. One of these is the emphasis on the family, with the documentation of multi-generation pedigrees and a focus on determining the implications for other family members including siblings, offspring and more distant relatives. Another is the longer appointment times and emphasis on preparation for clinic. A third and major one is the approach to testing and consent for testing and the realisation that the impact may be far beyond a diagnosis, as is seen in the predictive test setting.

1.1.2 Current diagnosis and testing in clinical genetics

Clinical genetics has always had phenotyping as a core skill. This was necessary when little testing was available, and giving the correct recurrence risk depended entirely on making the correct “gestalt” diagnosis based on dysmorphological features. This is something that is being challenged by the advent of the newer next-generation testing methods, where there is less pressure to focus testing on a specific gene.

Another change in the clinical genetics landscape is the recent trend for mainstreaming, where tests are done by other specialties, with patients attending genetics clinics for discussion of results and recurrence risks, rather than for diagnosis, or indeed not seeing a clinical geneticist at any point. Mainstreaming is important for ensuring equitable access to genetic testing for all patients, as well as utilising the expertise of other clinicians in their specialist areas. However, disadvantages may include the fact that some clinicians may not have the knowledge or time to deal with secondary findings of clinical significance (mutations found incidentally during testing which have medical consequences for patients or family members such as *BRCA1* mutations that increase the risk of breast and ovarian cancer) or the implications for other family members and provision of family or cascade testing. In addition, with high-throughput next-generation sequencing (NGS), huge numbers of variants are being discovered in patients and categorising these, which is discussed further in Chapter 3, requires skills in phenotyping as well as familiarity with the scientific principles of genetics, including mutation types and their consequences.

(Chapters 3 and 4). This puts clinical geneticists, along with their laboratory colleagues, at the forefront of delivering genetic testing to the population.

NGS is now firmly established in the NHS laboratory setting, with most patients having a panel of genes sequenced simultaneously, rather than individual genes sequenced sequentially as was the practice until recently. Panel use has been both driven by, and confirmatory of, the realisation that there is a spectrum in the presentation of most genetic diseases and that not all cases will present in the classic manner described in the literature. This has led to the diagnosis of many individuals who would not have received testing for mutations in a particular gene previously, as they did not meet established criteria for diagnosis. However, more widespread sequencing has led to the discovery that some mutations described in the literature as pathogenic are instead likely to be benign, something that has profound medical and psychological implications for patients and family members(2). Variant classification is assisted by the use of guidelines, but these need to evolve as more is learned(3-5). Clinical exome testing is becoming more widely used in the NHS setting and is common in the research setting, though it is not without difficulties, including the large numbers of variants found and the possibility of secondary findings (Chapter 3). Exome sequencing, as well as driving the discovery of novel causative genes, has led to the possibility of completely agnostic testing, with no prior diagnostic hypothesis. However, this leads to real challenges in the interpretation of variants of unknown significance (VUS) and highlights the importance of accurate phenotyping.

1.1.3 Future NHS testing directions

The 100,000 genomes project is a clinical transformation project funded by the Department of Health that has been running since 2012. It was set up to use whole genome sequencing (WGS) in the diagnosis of rare disease and cancer, and while recruitment closed in September 2018, results continue to be returned to clinicians and patients. Following on from this, a new NHS test directory has been developed and WGS is intended to be part of NHS genetic testing in future. There are advantages and disadvantages to this, again discussed in Chapter 3, but there will undoubtedly be challenges in result interpretation and dealing with secondary findings. In addition, consenting for any genetic test is complex, but obtaining properly informed consent for WGS, including allowing the patient to think about whether they wish to be informed of secondary findings such as carrier status or disease-causing mutations will be challenging. Policy decisions, such as which secondary findings should be reported and to whom, will also be necessary. As genetics knowledge changes and develops rapidly, there will also be a need to revisit test results when no diagnosis has been made (Chapter 3). At present relatively little is known about non-coding sequence but this is likely to change.

There are other benefits to using WGS. For example, results can be revisited if another diagnosis is made in the future or in another family member. A second advantage is that additional information can be provided, for example about disease risk or pharmacogenomics (Chapters 5 and 6). A third is that the more genomes that are sequenced and published, the better our knowledge of variant frequencies and consequences becomes.

1.1.4 Personalised medicine

Personalised medicine is defined in a number of ways, but all describe the greater understanding of the characteristics of patients, be it their diagnosis, genetic background or other features, in order to provide targeted advice and treatments that will be effective and minimise adverse effects(6). Traditional medical practice, which is empirical and operates on the assumption that patients with a particular diagnosis will respond to treatment in a similar manner, is evolving rapidly. Genomics has a role to play in this. It is likely that in the near future many patients will have their whole genome sequenced, and from this will obtain not only diagnostic information, but their carrier status for rare diseases, profiles of their risks of various conditions such as cardiovascular or psychiatric disease, profiles of therapies they will respond to and benefit from, therapies that will cause problems and more. Already commercial companies are providing much of this information as well as information on ancestry, parentage and less important medical traits such as the presence of the photic sneeze reflex(7). Clinical genetics has always been a specialty of personalisation, with the selection of tests depending on the patient's phenotype, but now tests tell us more than diagnosis- they also give information on prognosis and treatment. The 100,000 Genomes Project has begun the process of increasing personalisation in the NHS setting, but is still in the early stages.

1.2 Genetic testing in the research setting

As WGS moves into the clinical setting, other techniques are becoming more widely used for research. One of these is multiomics. Multiomics or omics is an umbrella term describing any biological study ending in the suffix omics. These include genomics, metabolomics, proteomics, transcriptomics, methylomics and microbiomics. Omics studies have expanded rapidly over the last decade and involve the integration of multiple sets of omics data, often but not always including genomics, to answer a research question. Large data sets in multiple modalities increase the power of studies, enabling the discovery of biomarkers and therapy targets, molecular disease-stratification and more. It is a powerful tool for studying complex disease, where single datasets, such as genomics, may not yield results(8). It is also a proven tool for use in small cohorts. In 2014, the Snyder group analysed the genome of a healthy individual and carried out proteomics, transcriptomics, metabolomics and immune profiling over a 24-month period(9). During this time the subject had several viral infections and developed type II diabetes. Rising glucose levels were identified and treated early. The course of viral infections could be followed through the multiomics data, showing that multiomics may be the way in which truly personalised medicine will be implemented. Multiomics may also provide a way around the need for expensive and time-consuming functional studies to prove the pathogenicity of variants. For example, transcriptomics or proteomics could be used to prove haploinsufficiency in an individual with a mutation in a gene of interest. Multiomics is also being used to identify drug targets and biomarkers and to increase our knowledge of common and rare disease(10-13). However, integration and interpretation of these complex datasets is challenging and clinical data is important in interpreting multiomics results.

1.3 The HIGH-5 project

The HIGH-5 project, running at the UCL Great Ormond Street Institute of Child Health, is one of the earliest multi-cohort multiomics projects in rare disease. Its objectives are:

- To establish an expert team for the implementation of multiomics studies in rare disease
- To establish protocols for the collection, extraction and processing of samples
- To establish methods for the collection, recording and utilisation of clinical data
- To establish minimum clinical data sets for multiomics research
- To develop methods of data integration of omics and clinical data
- To use machine learning to interrogate data sets
- To discover modifiers, biomarkers and treatment targets in rare disease cohorts
- To make the multiomics pipeline scalable
- To identify and overcome challenges and issues in the implementation of multiomics projects

Members of the HIGH-5 team include laboratory scientists, clinicians and bioinformaticians who provide expertise in all areas of data generation, processing and interpretation. The work undertaken in this PhD was done under the auspices of the HIGH-5 project.

1.3.1 Disease cohorts in HIGH-5

All the cohorts in the HIGH-5 project are examples of rare diseases. Rare diseases are defined by the European Union as those which affect fewer than five in 10,000 people(14). They are individually rare but collectively common, with between 6,000 and 8,000 diseases affecting six to eight percent of the population(15). Rare Disease UK suggests that up to 3.5 million people in the UK may be affected and up to five novel rare diseases are described weekly(16). They impose a huge burden of individual morbidity and mortality and also a significant financial burden(17). There were seven patient cohorts included in HIGH-5 and the cohorts, along with patient numbers, available samples and multiomics data obtained are listed in Table 1.1. The cohorts included adults and children with Bardet-Biedl syndrome (BBS), children with juvenile dermatomyositis (JDM), children with mitochondrial disease (MIT), children with Silver-Russell syndrome (SRS), adults with Usher syndrome (USH), children with Wilm's tumours (WIL) and children with very early-onset inflammatory bowel disease (VEOIBD or IBD). Of these cohorts, clinical information was available for BBS, JDM, USH and IBD cohorts and this was collected in the phenotyping databases described in Chapter 4. Five cohorts had WGS performed (BBS, IBD, JDM, SRS, USH) and so pharmacogenomic profiling could be done (Chapters 5 and 6). Three individuals from the BBS cohort did not have a molecular diagnosis and they are discussed further in Chapter 3.

Cohort	Size	Samples available	Omics
BBS	15	Whole blood	Whole genome sequencing (WGS)
		Fibroblasts	mRNAseq transcriptomics, proteomics
		Plasma	Proteomics, metabolomics
		Urine	Proteomics
		Peripheral blood mononuclear cells (PBMC)	mRNAseq transcriptomics
IBD	20	Gut biopsy	mRNAseq transcriptomics, proteomics, metabolomics
		Peripheral blood leucocytes	mRNAseq transcriptomics, proteomics, metabolomics
		Plasma	Proteomics
		Whole blood	WGS
JDM	13	$CD4^+$, $CD8^+$, $CD14^+$, $CD19^+$ cells	mRNAseq transcriptomics
		Plasma	Proteomics, metabolomics
		Whole blood	WGS
MIT	37	Fibroblasts	mRNAseq transcriptomics proteomics
		Whole blood	Exome sequencing (WES)
SRS	30	DNA	WGS, methylomics
USH	3	Multiple clones of differentiated iPS cells at 3 time-points	mRNAseq transcriptomics, proteomics, methylomics
		Whole blood	WGS
WIL	50	Tumour DNA	WES, methylomics

Table 1.1 Disease cohorts, sample numbers and multiomics in HIGH-5 project

1.3.1.1 Bardet Biedl Syndrome

Bardet Biedl syndrome is a ciliopathy affecting between 1 in 100,000 and 1 in 160,000 in non-consanguineous individuals of European extraction although the prevalence is higher in other populations(18). It is characterised by obesity, hypogonadism, polydactyly, rod-cone dystrophy resulting in blindness, learning difficulties and renal and genitourinary malformations in addition to other features. Patients are diagnosed clinically against major and minor criteria(18). BBS is inherited in an autosomal recessive manner. At least 21 genes are known to cause BBS (appendix 1, table A1.1) and even among patients of the same family there can be significant phenotypic variability(19). The commonest gene causing BBS is *BBS1*, accounting for approximately 42% of cases(18, 20, 21). Currently there is no treatment but there is ongoing research into therapies including gene and gene-editing therapies(22). BBS is discussed in Chapter 3.

1.3.1.2 Juvenile Dermatomyositis

Juvenile dermatomyositis is the commonest idiopathic inflammatory myopathy in children. A rare disease, it affects about two to four children per million per year(23). Its main features are

progressive weakness of the proximal muscles and skin changes of the face, hands and extremities. Serious complications including calcinosis and lung and cardiac disease may occur(24). It is believed to be an autoimmune disease, with females affected more often than males(25). The age of onset ranges from very early childhood to 18, but the median age of onset is seven. Diagnostic criteria were proposed by Bohan and Peter in 1975(26, 27). Consensus based guidelines for diagnosis and management were released in 2016(24).

The disease is associated with significant morbidity and mortality and sequelae can continue into adulthood(28). The exact cause is not known, though risk factors have been identified. A recent genome-wide association study showed that human leucocyte antigen (HLA) region was the most strongly associated and identified other loci known to be important in autoimmunity also including one at *PTPN22*(29). It is likely that combinations of genetic risk factors increase the likelihood of JDM to a greater or lesser extent. There is also evidence that myositis-specific antibody subgroups may be useful in stratifying patients and predicting outcomes(30). Environmental factors such as exposure to UV light and pollutants have also been proposed as risk factors(31, 32). The mainstay of treatment is immunosuppressants, both high dose corticosteroids and other disease-modifying drugs(24).

1.3.1.3 Silver-Russell Syndrome

Silver-Russell syndrome (SRS), also known as Russell-Silver syndrome, is characterised by intra-uterine growth retardation and post-natal proportionate short stature. Affected individuals are said to have a normal head size, a triangular face and may have hemi-hypotrophy(33). Cognitive impairment and developmental delay are seen in some affected individuals. The incidence varies significantly depending on how the syndrome is defined, but is estimated to be between 1 in 3000 and 1 in 10,000(34). A clinical scoring system has been developed to try to resolve this(35-37).

The commonest cause of SRS is hypomethylation of the first imprinting centre of the paternal allele of chromosome 11p15.5, accounting for between 35 and 50% of cases(33). 11p15.5 is imprinted and paternal hypomethylation causes biallelic loss of expression of IGF2 and biallelic expression of H19. This results in growth restriction(38). Another relatively common cause is maternal uniparental disomy of chromosome 7 (UPD7), which accounts for a further 10%, but rarer causes have been reported(33, 39). Recurrence and offspring risks depend on the cause, but in most cases are low(33). No specific treatment is available but growth may be improved by growth hormone administration(40).

1.3.1.4 Usher syndrome

Usher syndrome is a rare, inherited ciliopathy whose main features are deafness and progressive visual loss caused by retinitis pigmentosa (RP)(41). It is the commonest inherited cause of deafblindness. Three types of Usher syndrome have been identified, each of which may be caused by a number of genes (appendix 1, table A1.2)(42).

Usher syndrome type I is characterised profound congenital sensorineural deafness, blindness secondary to RP and vestibular areflexia and is inherited in an autosomal recessive manner(43).

Unless treated with cochlear implantation, the hearing loss is such that affected individuals do not develop speech. Six genes causing Usher syndrome type I have been identified(44). Mutations in *MYO7A* account for between 53-63% of cases, with *CHD23* and *USH1C* accounting for the majority of the remainder. In approximately 10-15% of patients, causative mutations are not found in any of the six known genes(43).

Usher syndrome type II is characterised by deafness that is moderate at lower frequencies and severe to profound at higher frequencies and blindness secondary to RP, with normal vestibular reflexes. Again it is an autosomal recessive condition, with three genes, *USH2A*, *WHRN* and *ADGRV1*, identified to date. Mutations in *USH2A* account for up to 80% of cases(43).

Non-congenital progressive deafness, RP and abnormal vestibular function characterise Usher syndrome type III, with two causative genes, *CLRN1* and *HARS*, identified(45). In all types there is significant variability in severity and age of onset of RP and in types 2 and 3, variability in severity of hearing loss. Estimates of prevalence vary from approximately 1 in 6000 to 1 in 20000(43, 46). Treatment is supportive, with cochlear implants or hearing aids being used. The use of sign language may be complicated by visual impairment.

1.3.1.5 Very early-onset Inflammatory Bowel Disease

Inflammatory bowel disease (IBD) consists of a group of conditions causing chronic inflammation in the lower gastro-intestinal tract. It presents with weight loss, pain, diarrhoea, which may contain blood or mucus, and fatigue as well as extra-intestinal symptoms such as mouth ulcers. Overall, approximately 20-25% of patients with IBD develop symptoms as a child or adolescent(47). Very early-onset inflammatory bowel disease (VEOIBD) is defined as the onset of IBD in children of under six years. It can be further divided into neonatal (birth to 28 days), infantile (one month to two years) and early childhood (ages two to six) onset(48, 49). A retrospective cohort study in 2009 determined that the incidence of IBD in children was 13.2/100000 and the incidence of VEOIBD 3.4/100000. Both had increased since 1994, and this trend appears to have continued(50, 51).

Individuals can present with features of ulcerative colitis (UC), Crohn's disease or both, in which case it is termed unclassified IBD, a presentation more common in children than in adults(49, 52). In addition to the symptoms of IBD, children may have issues with growth and pubertal development as well as having more extensive disease and rapid progression(53-55). Delays in diagnosis are common, with up to 1 in 5 children having symptoms a year before diagnosis(53). Paediatric disease may also be more resistant to treatment(56).

There is evidence to show that VEOIBD is more likely to be monogenic than other forms of IBD, with 31% of children with an age of onset under 2 years having mutations in a causative gene(56). To date, 60 genes causing monogenic VEOIBD have been identified (appendix 1, table A1.3) and many other single nucleotide polymorphisms (SNPs) have been implicated in the development of both paediatric and adult-onset IBD(57). Treatment of children with VEOIBD is similar to that of older children and adults, with antibiotics, anti-inflammatory drugs, immune suppressants and

biological agents being used, as well as surgical and nutritional therapies. Monogenic forms of VEOIBD may lend themselves to specific approaches such as the use of stem cell transplant for patients with IL-10R deficiency(58, 59).

1.3.1.6 Other cohorts

No data from the MIT or WIL cohorts were utilised in this thesis and they are not discussed further.

1.4 Objectives of this thesis

This thesis summarises work undertaken as part of the HIGH-5 project, and falls into three parts. As the parts are complementary, each chapter has an introductory section exploring the current practice in that area. The overall aim is to increase personalisation in the diagnostic and treatment settings and to analyse how whole genome sequencing will help with this.

1.4.1 The use of whole genome sequencing for diagnostics

Chapter 3 looks at the use of WGS in the diagnostic setting using three patients, a pair of monozygotic twins and a singleton, all with a clinical diagnosis of Bardet-Biedl syndrome. All met the current diagnostic criteria and had had testing for mutations in 19 of the currently known BBS genes performed by the North East Thames Regional Genetics Laboratory (*BBS20* and *BBS21* were not included on the panel)(60). In this chapter, the following are explored:

- The heterogeneous nature of BBS and other ciliopathies and the difficulties in their diagnosis
- Whether a diagnosis could be made for each patient using current variant interpretation guidelines to interpret WGS
- If not, why this might be the case and what further steps might be taken to obtain a diagnosis
- Whether WGS offers advantages over other methods of genetic testing including panel-based and whole exome testing
- What the disadvantages of WGS are and how they can be mitigated
- Important considerations for the introduction of WGS to the NHS diagnostic test directory and the challenges that may be faced

1.4.2 Deep phenotyping and its place in the research setting

The HIGH-5 project enrolled multiple cohorts of rare disease patients. One of the challenges of studying rare disease is the difficulty of enrolling homogeneous cohorts of patients. Deep phenotyping was used to try to stratify the patients and to make the interpretation of the multiomics data more straightforward. In addition the construction of a database to store clinical data securely and in a standardised manner allowed the HIGH-5 project to look at integrating clinical and multiomics data for analysis, to develop software for the analysis of multiomics data and to utilise clinical data and multiomics data for machine learning. In Chapter 4 the following are explored:

- Why deep phenotyping is important
- What tools are currently available for phenotyping and how they can be used
- What are the most useful data types for deep phenotyping and how they can be recorded, standardised and output for maximum utility
- How a purpose-built database was built and utilised
- The requirements for keeping such data secure and how this could be achieved
- The challenges of recording phenotypic data and how these could be met or improved

1.4.3 Extraction of pharmacogenomic data from whole genome sequencing

Pharmacogenomics is an area of significant current research and is not yet in widespread use in the NHS. However, the Department of Health is planning to extract pharmacogenomic data from WGS in the future. Chapters 5 and 6 consider the following:

- Whether it is feasible to extract pharmacogenomic data from WGS and how reliable this is, particularly for genes with multiple haplotypes or copy number variants
- How the use of WGS compares to other testing methods for pharmacogenomics
- Which genes it is useful to extract data for and how this might be translated into clinical use
- How various sets of guidelines differ from one another and which guidelines might be adopted in the UK in future
- The benefits and limitations of pharmacogenomic testing and the current practice
- How the NHS might introduce pharmacogenomic testing and what considerations might be important for this

Overall, this thesis uses WGS and clinical data to investigate how a more personalised diagnostic and therapeutic service might be provided to NHS patients.

Chapter 2 Materials and Methods

2.1 Samples

2.1.1 Cohorts

There were five cohorts from which samples were obtained. Numbers of individuals are detailed in Table 2.1. Detailed clinical information is available in an anonymised form in Supplementary Information S2.1 and S2.3 (CD-ROM).

Cohort name	Cohort details	Number of individuals	Additional information
BBS	Individuals with Bardet-Biedl syndrome	18	All affected clinically. The cohort included two sets of monozygotic twins. 15 of 18, including one set of twins, had two pathogenic mutations including at least one p.Met390Arg mutation in *BBS1*. Three, including the second set of twins, did not have a molecular diagnosis
IBD	Individuals with very early-onset inflammatory bowel disease	20	Histologically confirmed. Age of onset less than six years
JDM	Individuals with juvenile dermatomyositis	12	All affected, clinically diagnosed
SRS	Individuals with Silver-Russell syndrome	30	Historical samples. The cohort included ten affected individuals and their parents i.e. ten parent-child trios. No information apart from ethnicity available
USH	Individuals with Usher syndrome	3	The cohort included two affected individuals with confirmed pathogenic mutations and one healthy control

Table 2.1 HIGH-5 cohort details

2.1.2 Ethics

All participants were part of the HIGH-5 cohort. Each individual cohort had ethics approval for WGS, data analysis and collection of clinical information as follows. The BBS cohort samples had ethics approval granted by the West Midlands Research Ethics Committee (REC) as part of the EU rare disease registry for Wolfram, Alstrom, and Bardet-Biedl Syndromes (EURO-WABB) study(61). The IBD cohort samples had ethics approval granted by the London Bloomsbury REC

as part of the Patients with Early-onseT Intestinal inflammaTion (PETIT) study. The JDM cohort samples were collected under ethics permission from the UK Northern & Yorkshire REC with approval for inclusion in the HIGH-5 project given by the UK Juvenile Dermatomyositis Research Group (JDRG). SRS samples were obtained before 2006 and were anonymous. Consent for entry into the High-5 project was from the London South East NHS REC, as was the consent for the USH samples.

2.1.3 Sample collection

Sample collection was done at various times and in different places. However, all BBS samples were collected at Guys and St Thomas' or Great Ormond Street Hospital (GOSH) NHS Foundation Trusts by Professor Philip Beales, Dr Elizabeth Forsythe and Dr Joanna Kenny, all IBD samples were collected at GOSH by Dr Jochen Kammermeier, all SRS samples were collected by Professor Gudrun Moore at GOSH and all USH samples were collected by Prof Maria Bitner-Glindzicz at GOSH. The JDM samples were collected by various clinicians in the UK. All patients had samples collected for DNA extraction. A minimum of 3.5mls of whole blood in the case of adult samples and 1-2mls in the case of paediatric samples was collected in EDTA and mixed well. They were transported to the laboratory for DNA extraction. Samples collected at GOSH, with the exception of SRS historical samples, where collection method is not known, used the Starstedt S-Monovette® system, while samples collected at Guys and St Thomas NHS foundation trust were collected in Beckton Dickenson Vacutainer® tubes.

2.1.4 Sample naming

All samples were allocated a unique HIGH-5 identification number when the samples were received by the HIGH-5 project. The first three letters indicated the cohort (BBS- Bardet-Biedl syndrome, IBD- very early-onset inflammatory bowel disease, JDM- juvenile dermatomyositis, SRS- Silver-Russell syndrome and USH- Usher syndrome) and the numbers indicated the order in which they were registered. For example the first Bardet-Biedl sample registered became BBS-001 and the fifth Silver-Russell sample became SRS-005. This allowed anonymisation of the samples. Details of the samples including date and time of collection, date of extraction, date of use, sample concentration and additional sample details were stored anonymously under the HIGH-5 identifier in a Microsoft Access (MS Access) database. No clinical or personal details were stored with the sample details.

2.2 DNA extraction and quantification

Genomic DNA was extracted from most BBS samples by the North East Thames Regional Genetics Laboratory, Great Ormond Street NHS Foundation Trust or by the method detailed below. Additional BBS samples and parental samples were extracted by the method detailed below. DNA was provided already extracted for IBD, JDM, SRS and USH samples.

2.2.1 DNA extraction method

The QIAamp® DNA Blood mini-kit (spin protocol) was used for genomic DNA extraction and extraction was performed according to the manufacturer's instructions (www.qiagen.com)(62). Samples and appropriate reagents were brought to room temperature and a heating block prepared (56^{o}C). Mixing was done using a mini-vortex and centrifugation using a microcentrifuge unless otherwise specified. 20μL of QIAGEN Protease was placed in a 1.5μL microcentrifuge tube, 200μL of whole blood in EDTA was added and the sample mixed. 200μL of lysis buffer AL was added and the sample mixed. The sample was incubated for 10 minutes in a heat block heated to 56^{o}C. The sample was then centrifuged for 30 seconds.

500 μL of buffer AW1 of 100% ethanol was added and the sample mixed for fifteen seconds and centrifuged for 30 seconds. The mixed sample was placed in a 2ml QIAamp Mini spin column which was held in a collecting tube and centrifuged at 6000g for 1 minute. The collecting tube was replaced and the filtrate discarded. 500μL of wash buffer AW1 was added to the Mini spin column. This was centrifuged at 6000g for 1 minute. The collecting tube was replaced and the filtrate discarded. 500μL of wash buffer AW2 was added and the sample was centrifuged at 20000g for 3 minutes. The collecting tube was replaced and the filtrate discarded and the sample centrifuged for a further minute at 20000g. The Mini spin column was placed in a clean 1.5μL microcentrifuge tube and 500μL of elution buffer AE was added. The sample was incubated at room temperature for 1 minute and then centrifuged for 5 minute at 6000g. A further 500μL of buffer AE was added and the incubation and centrifugation repeated. DNA was quantified and stored at -20 ^{o}C.

2.2.2 DNA quantification

2.2.2.1 DNA quantification and assessment using NanoDrop™

The ThermoScientific™ NanoDrop1000™ spectrophotometer was used to quantify DNA immediately after extraction according to the manufacturer's instructions(63). The instrument was cleaned and a 1μL aliquot of blank (elution buffer AE) was loaded onto the pedestal and the instrument closed. The blank was measured using the blank option and recorded and the instrument wiped. This step was repeated using a second aliquot of blank, but this time using the measure sample option. This step was repeated until the measurement difference between the blank and measured blank was no more than 0.04A (absorbance units.) A 1μL aliquot of the DNA sample was loaded onto the cleaned pedestal, the instrument closed and DNA selected as the nucleic acid to be measured. The sample was measured using the measure sample option. The concentration (in ng/μL), absorbances and absorbance ratios were recorded for each sample. The instrument was reblanked if in use for more than 30 minutes. Samples were considered to be of sufficient quality if the 260/280 and 260/230 ratios were >1.8.

2.2.2.2 DNA quantification using Qubit

The Thermofisher™ Qubit™ 3 fluorometer and Invitrogen™ Qubit™ dsDNA broad range (BR) assay kit were used to quantify DNA before experimental use as per the manufacturer's instructions. Reagents and DNA samples were brought to room temperature. A working solution

was prepared by adding 1μL of Qubit dsDNA BR reagent into 199μL Qubit dsDNA BR Buffer for each sample being quantified plus the two standards. Standards one and two were made up by adding exactly 10μL of standard to 190μL of working solution in labelled Qubit™ assay tubes. Samples were made by mixing 1μL of DNA to be quantified into 199μL of working solution in labelled Qubit™ assay tubes. All samples and standards were mixed by vortexing and spun for 30 seconds. Samples and standards were incubated at room temperature for 2 minutes.

The broad-range assay was selected in the Qubit™ programme. Run new calibration was selected and the first standard and then the second standard were assayed. Once the concentration standard curve had been calculated, the samples were assayed and their concentration plotted on the curve and recorded in ng/mL and μg/mL. Samples that were too concentrated were diluted until they fell within the standard curve.

2.3 DNA sequencing

2.3.1 Whole genome sequencing

WGS was carried out by BGI (www.bgi.com) for all samples in the cohort. Samples were diluted to a concentration of 60μg/mL, with a minimum sample volume of 15μL. BGI carried out sample quality control checks both before and during processing. Library preparation was performed with Illumina™ TrueSeq DNA PCR-free Preparation kit as specified by the manufacturer. Samples were sequenced using the Illumina™ HighSeq X Ten, which uses sequencing by synthesis (SBS) technology, again as per instructions. Samples were sequenced to a depth of 30x and fastq files (text files that store sequence and sequence quality data) were generated.

2.3.2 Sanger sequencing of *CEP290*

Sanger sequencing of *CEP290* variants in exon 38 and the promoter region was done for patient BBS-018, her mother and a control (Chapter 3).

2.3.2.1 *CEP290* Primer design

Primers were designed using Ensembl and the primer design function of the University of California Santa Cruz (UCSC) genome browser(64-66). Specificity was confirmed by using the UCSC genome browser Basic Local Alignment Search Tool (BLAST) function. In the case of coding variants, primers were designed to amplify the whole exon. In the case of non-coding variants, primers were designed at approximately 150 base pairs (bp) either side of the variant. Primers were designed to have a minimum length of 18bp and a maximum length of 24bp. Forward and reverse primers were synthesised by Sigma-Aldrich™ and are listed in Table 2.2.

2.3.2.2 Polymerase chain reaction (PCR) of *CEP290*

Touchdown PCR was performed using a t100™ thermal cycler (Bio-Rad). The PCR mix is shown in Table 2.3 and the PCR programme in Table 2.4. All PCRs were performed according to manufacturer's instructions using the primers as described in section 2.3.2.1 and Promega® reagents. The promoter fragment was 323 base pairs and the exon 38 fragment was 576 base pairs.

Gene	Location	Direction	Melting temperature	Sequence
CEP290	Exon 38	F	64.0^{0}C	CACTTGAATCTGGGAGGCAG
CEP290	Exon 38	R	61.0^{0}C	CACAAATCAGATTGACGAAAAC
CEP290	Promoter	F	60.9^{0}C	CTTGCACGAGTAAGAGTGGTAA
CEP290	Promoter	R	64.1^{0}C	GATAGTTAGAGTGAGAGCCGCG

Table 2.2 Primers for Sanger sequencing of CEP290

Reagent	Volume
Distilled water	18.7 µL
5x Buffer	6µL
Magnesium chloride	2.4µL
Deoxyribonucleotide Triphosphate (dNTPs) (10mM)	0.6µL
Forward primer (20mM)	0.3µL
Reverse primer (20mM)	0.3µL
GoTaq G2 Flexi Polymerase	0.15µL
Genomic DNA	1µL

Table 2.3 PCR mix

Temperature	Duration	Additional instructions	Number of cycles
105ºC	Before start	Preheat lid	n/a
95 ºC	3 minutes		10 cycles
95 ºC	30 seconds		
64 ºC	30 seconds	Reduce temp by 1 ºC each cycle	
72 ºC	1 minute		
95 ºC	30 seconds		24 cycles
54 ºC	30 seconds		
72 ºC	1 minute		
72 ºC	7 minutes		
4 ºC	Indefinitely		

Table 2.4 Touchdown PCR programme

2.3.2.3 Checking of *CEP290* PCR products using gel electrophoresis

PCR products were run on a 2% agarose gel to check primer specificity. 4g of UltraPure agarose (Invitrogen™) was dissolved in 200ml of Tris-Borate-EDTA (TBE) buffer (BioRad™) by heating in a microwave. Once it had cooled slightly, 4μL of ethidium bromide (Sigma-Aldrich™) was added and it was mixed well by swirling. It was poured into a casting tray, and combs were added to give a 10 well gel which was left to set. Once set it was placed in an electrophoresis tank which was then filled with 1x TBE buffer. A 100 bp ladder (Bioline™) was added to the first and last wells. 4μL of PCR product was mixed with 1μL of TrackIt™ loading buffer (Thermofisher™). Samples were loaded as in Table 2.5 and run at 100V for 50 minutes. The gel was then photographed using a UVP BioDoc-It™ imaging system (Analytik Jen AG). The promoter fragment was 323 base pairs and the exon 38 fragment was 576 base pairs. Results are shown in Figure 2.1.

Well	Contents
1	100 bp ladder
2	BBS-018 promotor region *CEP290*
3	Mother of BBS-018 promotor region *CEP290*
4	Control promotor region *CEP290*
5	No template control (NTC) promotor region *CEP290*
6	BBS-018 Exon 38 *CEP290*
7	Mother of BBS-018 Exon 38 *CEP290*
8	Control Exon 38 *CEP290*
9	No template control (NTC) Exon 38 *CEP290*
10	100 bp ladder

Table 2.5 Order of loading of agarose gel

Figure 2.1 Agarose gel showing CEP290 PCR products in the order described in Table 2.5

2.3.2.4 Cleaning and quantification of *CEP290* PCR products

Exo-SAP-IT™ (Affymetrix™) was placed on ice. For every 5μL of PCR product, 2μL of Exo-SAP-IT™ was added. Samples were incubated at 37 °C for 15 minutes and then at 80 °C for 15 minutes and held at 12 °C indefinitely. Samples were quantified using NanoDrop™ and Qubit™ as described in section 2.2.2.

2.3.2.5 Sanger sequencing of CEP290 PCR products

Sanger Sequencing was carried out by Source Bioscience™. 5μl of DNA at a concentration of 1ng/μL and 5μL of primers at a concentration of 3.2pmol/μL were sent. Results were returned electronically and visualised in Sequencher® (Gene Codes Corporation).

2.3.3 SNP genotyping using Thermofisher 12KFlex™

Discussed in Chapter 6, SNP genotyping was done by me using a custom QuantStudio™ 12K Flex (Applied Biosystems™) assay designed by Congenica Ltd. All samples in the cohorts were genotyped with the exception of those listed in Chapter 6 which were excluded because of a shortage of DNA. Details of SNPs included can be seen in Chapter 6. SNP genotyping was done as per manufacturer's instructions. Samples were run in batches of up to a maximum of 16 with at least one being a no template control (NTC). Each plate had 16 sets of three sub-arrays containing 180 assays.

2.3.3.1 Sample quantification

DNA was quantified using Qubit™ as previously described and diluted to 50ng/μL using nuclease-free water. A minimum volume of 8.5μL was required for each sample.

2.3.3.2 Plate setup

The correct plate setup file for the particular plate being used was downloaded from the Thermofisher website (www.thermofisher.com). QuantStudio™ OpenArray® AccuFill™ software was used to integrate this with a sample file, linking the samples to the plate and assays.

2.3.3.3 Sample preparation and plate loading

8.5μL of DNA at a concentration of 50ng/μL was mixed with an equal volume of 2x TaqMan® OpenArray® genotyping master mix and vortexed until well mixed. 5μL of sample plus master mix was placed in 3 positions of a 384 well plate according to the layout dictated by the AccuFill™ software. In the case of a 16 sample experiment this was columns 1-12 of rows A, B, C and D. The plate was covered with an adhesive foil cover.

The QuantStudio™ OpenArray® AccuFill™ instrument was prepared and a system test performed. The 384 well plate was placed in the AccuFill™ and the TaqMan® OpenArray® assay plate corresponding to the plate set-up file was placed in the instrument, having first been removed from the freezer and defrosted for 15 minutes. The foil over the part of the 384 well plate containing the samples to be loaded was removed. The instrument was closed and load selected. Once the plate was loaded it was sealed with an adhesive lid using the TaqMan® OpenArray®

Plate Press 2.0 within a maximum of 90 seconds. The OpenArray® assay plate was filled with immersion fluid loaded in a pre-primed syringe within 60 seconds. The fluid was loaded in a single slow, smooth action to avoid air bubbles. The plate was then sealed with the OpenArray® plug.

2.3.3.4 Sample genotyping

The sealed OpenArray® assay plate was loaded into the QuantStudio™ 12K Flex instrument and the plate information entered. The assay was started by selecting run. Once the assay was completed results were exported and uploaded to the Thermofisher™ cloud for analysis and QC data were checked for problems (see user guide available at www.thermofisher.com). The assay works by running real-time PCR based on TaqMan™ chemistry. Probes and primers for each allele of a SNP are present on the assay plate. Probes are labelled and the amount of each fluorescent dye present is used to call the genotype (section 2.4.2.3).

2.3.4 Copy number confirmation using TaqMan® *CYP2D6* copy number assay

Only 2 samples identified as having a possible copy number variant (CNV) had sufficient DNA for confirmation (IBD-007- duplication, IBD-013- deletion). *CYP2D6* copy number confirmation was performed by Ms Jasmine Risvi, Congenica Ltd., and done according to manufacturer's instructions.

2.3.4.1 Sample quantification

Sample quantification was done using Qubit™ as previously described and diluted to 5ng/μL using nuclease-free water. A minimum volume of 4μL was required for each sample.

2.3.4.2 Sample preparation and plate loading

At least 1 sample with known copy number and 1 no-template control (NTC) were used in each 96-well plate. 3 replicates of each were used. For a 96-well plate, reagents were prepared as described in Table 2.6 and mixed well. 4μL of vortexed DNA or NTC was placed in the 96-well MicroAmp® optical reaction plate, 16μL of reaction mix was added and mixed thoroughly by pipetting up and down. It was sealed with MicroAmp® optical adhesive film and centrifuged.

2.3.4.3 *CYP2D6* copy number assay reaction and results

The plate was loaded into an Applied Biosystems™ StepOnePlus™ 96 well real-time PCR system and the reaction was run as described in Table 2.7. Results were analysed using Applied Biosystems® Copy Caller software as per manufacturer's instructions.

Reagent Mix	Volume
TaqMan® genotyping Master Mix	10μL
TaqMan® *CYP2D6* copy number assay	1μL
TaqMan® reference copy number assay	1μL
Nuclease-free water	4μL
Total	16μL

Table 2.6 Reagent quantity for CYP2D6 TaqMan® copy number assay

Temperature	Duration	Number of cycles
95 ºC	10 minutes	At start
95 ºC	15 seconds	40 cycles
60 ºC	60 seconds	

Table 2.7 qPCR programme

2.4 Bioinformatics and sample analysis methods

2.4.1 WGS processing

Processing of BGI fastq files for all samples was done by the HIGH-5 bioinformatics team at the UCL Great Ormond Street Institute for Child Health, London (Dr A. Gagunashvili, Dr C. James, Dr N. Jani, Dr F. Minnecci and Dr G. Otto). Fastq files were checked using fastQC (Babraham Bioinformatics Ltd), a quality control programme that summarises read quality among other indices. The sequence was then aligned to GRCh38 using the Burrows-Wheeler Alignment (BWA) tool, which has been shown to be accurate for both short and long read alignment, resulting in sequence alignment (SAM) files(67). As binary alignment map (BAM) files were required for the analysis software, these were created using SAMtools(68). Genome analysis toolkit (GATK) was used to recalibrate the BAM files after duplicate reads had been highlighted using the Broad Institute Picard tools, and variant call format (VCF) files were generated(69, 70).

2.4.2 Variant visualisation and filtering

2.4.2.1 Integrative genomics viewer (IGV)

IGV (version 2.3) was used for visualising possible pathogenic variants in BBS-016, BBS-017 and BBS-018 (Chapter 3) and for calling pharmacogenomic variants in all samples (Chapters 5 and 6). Developed by the Broad Institute (Cambridge, MA, USA), it is a freely available tool that allows direct visualisation of a variant and surrounding sequence(71, 72). BAM files were uploaded, as was the reference genome to which they are aligned, in this case GRCh38. The genomic coordinates were entered and IGV highlighted the number of reference and non-reference reads and the overall read-depth. Several samples or positions could be visualised simultaneously. Tools were available within IGV to allow visualisation of paired-end alignment, helpful in identifying possible CNVs.

2.4.2.2 Ingenuity Variant Analysis™ (IVA)

IVA (QIAGEN) is a web-based application used to filter variants. Variant call format (vcf) files for BBS-016, BBS-017 and BBS-018 were uploaded to IVA, and filters were set to reduce the numbers of variants shown. Initially, filters were set to show coding variants and variants in immediately surrounding areas (up to 20 bp into the intron) only, but later expanded to show non-coding variants. Variants with a frequency of >5% in ExAC, GnomAD and NHLBI were excluded from analysis before filtering. It should be noted that data in population databases are based on GRCh37, rather than GRCh38 as these data are and frequencies quoted are from 28/06/2016.

Filter settings for IVA are shown in Table 2.8. Only variants with a frequency of <1% remained after filtering. For comparison, the most frequently seen pathogenic variant causing Bardet-Biedl syndrome is p.Met390Arg in *BBS1,* whose frequency in ExAC is 0.15%. The metrics of the variant examined are shown in Table 2.9 and tools and databases used within it in Table 2.10. Filters were added to select genes associated with Bardet-Biedl syndrome, ciliopathies and cilia (Appendix 2, Tables A2.1-A2.4). In addition, a filter containing known OMIM morbid genes was applied(73). Following this all nonsense and frameshift mutations in all genes were checked.

Intronic variants were considered only if they were situated within 20 base pairs of an intron-exon boundary or had a probable functional effect such as affecting splicing, the promotor or a transcription factor binding site identified by the ENCODE project(74). Variants with a read depth of less than 10x were excluded as artefacts are difficult to distinguish from genuine variants below this level. A read depth of 15x has been shown to be sufficient to give 97% coverage and call 98.7% of heterozygous variants(75). Variants with an allele fraction of <5% were excluded. Heterozygous disease-causing variants would be expected to have an allele fraction of approximately 50%. A cut off of 5% allowed the possibility of identifying mosaic variants and reduced the possibility of missing any true heterozygous variants. Variable genes and regions were not excluded. Predictions of pathogenicity by computational methods were not used to filter variants. When a single candidate variant was identified, the gene was visualised in IGV to see if a small copy number variant could be identified that would constitute a second pathogenic variant.

In the case of BBS-018, 4,878,324 variants were identified. 449,231 remained once variants seen at a frequency of more than 1% in 1000 Genomes, ExAC, GnomAD or NCLBI-ESP were removed. When only coding variants were considered, 4,725 were identified, 3,135 of which remained once common variants were removed. Any coding variants in OMIM morbid or cilia genes were examined, as were non-coding variants with possible functional effect or in ciliopathy genes with a single coding variant identified. With OMIM genes, they were considered if more than one coding variant was found, or the disease is inherited in an autosomal dominant or X-linked dominant manner.

In the case of BBS-016 and BBS-017, monozygotic twins, all discussed variants were seen in both patients. As both patients were male, hemizygous variants in genes on the X chromosome were considered. 4,691,841 shared variants were identified, 449,231 of which remained once variants seen at a frequency of more than 1% in 1000Genomes, ExAC, GnomAD or NHLBI-ESP databases were removed. When only coding variants were considered, there were 3,887 shared variants, 2,608 of which remained when common variants were removed. Any coding variants in OMIM morbid or cilia genes were examined, as were non-coding variants with possible functional effect or in ciliopathy genes with a single coding variant identified. With OMIM genes, they were listed if more than one coding variant was found, or the disease is inherited in an autosomal dominant or X-linked manner. Read depths and allele fractions detailed in Chapter 3 are averaged between the two patients unless they differed by more than 5 or 5% respectively, in which case they are given for each patient.

Filter	Metric	Minimum
Confidence	Call quality	20
	Genotype quality	20
	Read depth	10
	Allele fraction	5%
Common variants	1000 Genomes	>1%
	ExAC	>1%
	GnomAD	>1%
	NHLBI	>1%
Predicted deleterious	All	Nothing filtered

Table 2.8 Filter settings for Ingenuity Variant Analysis™

Metric	Meaning
Genomic position	Location of variant
Gene symbol	Gene identifier
Gene region	Exonic, intronic, splice site, promoter, UTR
Transcript variant	DNA sequence change
Protein variant	Protein sequence change
Zygosity	Heterozygous, homozygous or possible compound heterozygous
Read depth	Number of reads covering variant position
Allele fraction	Percentage of reads showing non-reference genotype
Translational impact	Frameshift, missense, nonsense, synonymous
In silico prediction	In silico tools to predict pathogenicity- PolyPhen2, SIFT
CADD score	Measure of deleteriousness
Regulatory site	Whether variant is at promotor, splice or other regulatory site
Population frequency	Frequency of variant in 1000 Genomes, ExAC, GnomAD, NHLBI
dbSNP	Whether the variant is listed as a known SNP
Disease databases	Whether the variant is listed as a pathogenic or possibly pathogenic variant- ClinVar, HGMD, OMIM

Table 2.9 Metrics analysed for variants in Ingenuity Variant Analysis™

Tool/ database	Function	Reference/ site
1000 Genomes	Database of variants from 1000 Genomes Project	http://www.internationalgenome.org/(76)
ExAC	Database of variants from 60,706 unrelated individuals- various disease-specific and population genetic studies. Subset of GnomAD	http://exac.broadinstitute.org/(77)
GnomAD	Database of variants from 123,136 exome sequences and 15,496 whole genome sequences- unrelated individuals, various disease-specific and population genetic studies	http://gnomad.broadinstitute.org/(77)
NHLBI	Database of variants from 6503 samples from multiple exome sequencing cohorts	http://evs.gs.washington.edu/EVS/(78)
ClinVar	Categorisation of relationship between variants and phenotypes with supporting evidence	https://www.ncbi.nlm.nih.gov/clinvar(79)
HGMD	List of all known disease-causing variants	http://www.hgmd.cf.ac.uk(80)
OMIM	Database of information about genes including function and phenotype	https://www.omim.org(73)
PolyPhen2	Predicts functional effect of amino acid changes	http://genetics.bwh.harvard.edu/pph2/(81)
SIFT	Predicts functional effect of amino acid changes	http://sift.bii.a-star.edu.sg/(82)

Table 2.10 Databases and tools used within Ingenuity Variant Analysis™

2.4.2.3 Thermofisher Connect™ cloud-based genotyping analysis

The data from the genotyping assay described in section 2.3.3 were uploaded to the Thermofisher Connect™ site and the assay information file (AIF) containing details of the assays was imported from the Thermofisher™ website. Cluster analysis was carried out automatically and genotypes called. Each data point in the real-time PCR cycle had 3 lines corresponding to FAM™, VIC® and ROX™ fluorophores. Homozygotes for allele 1 were called when there was a low level of FAM™ fluorescence and a high level of VIC® fluorescence. Allele 2 homozygotes were called if FAM™ fluorescence levels were high and VIC® fluorescence levels were low. Heterozygotes were called when FAM™ and VIC® levels were equivalent. An example of good clustering is shown in Figure 2.2. Examples of real time amplification plots are shown in Figure 2.3. Which allele was the reference allele was determined from the data in the AIF file.

Each assay was then visualised manually, in both amplification and cluster plots, to ensure that calls were correct. If calls were incorrect, they were corrected manually. Calls were determined to be incorrect if they clearly clustered with a different sample set to the one it was called as. Ambiguous results were examined to see if they could be determined, for example if they were

clearly tracking along the same trajectory as a cluster but more slowly or if clustering was clearer in an earlier cycle. This was visualised by rewinding to an earlier cycle in the multicomponent plots (Figure 2.4). Results were manually compared to WGS and discrepant calls flagged.

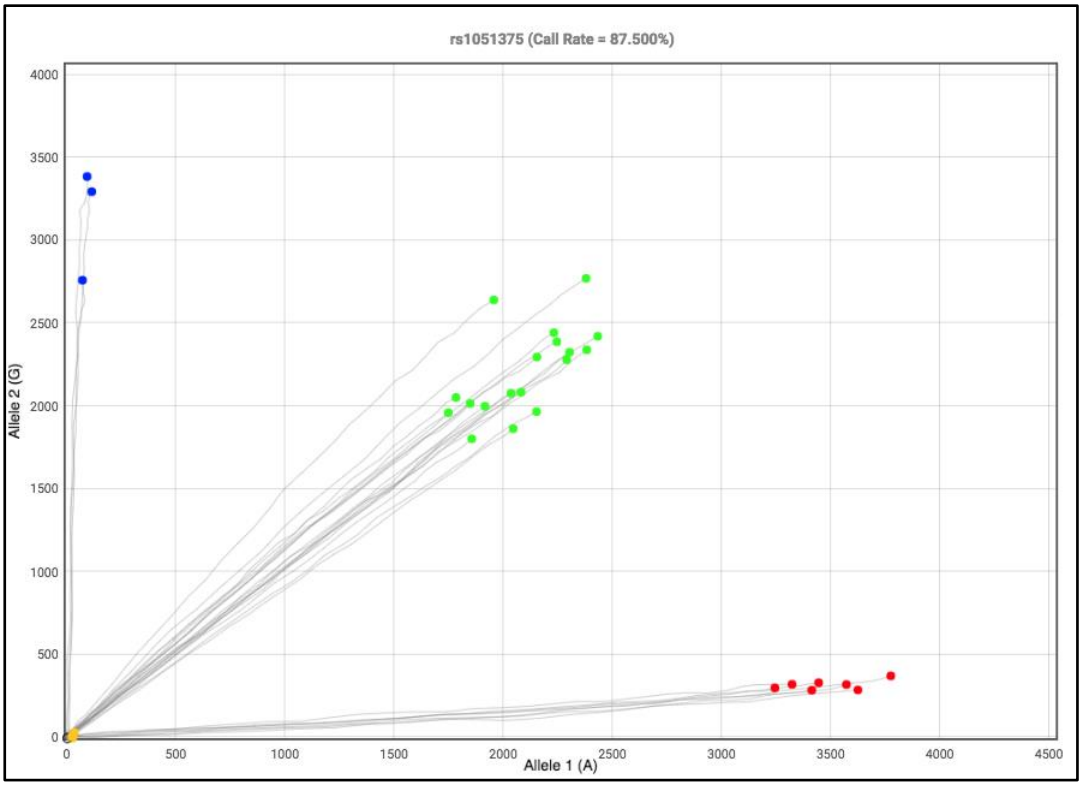

Figure 2.2 Good clustering of genotypes visualised in ThermofisherConnect™ showing homozygotes for allele 1 (red), homozygotes for allele 2 (blue) and heterozygotes (green)

2.4.3 Astrolabe

Astrolabe, previously known as Constellation, is freely available software developed at the Children's Mercy Hospital, Kansas, USA which uses unphased WGS data to identify pharmacogenomic diplotypes (https://www.childrensmercy.org/genomesoftwareportal/)(83). It was used to identify possible CNVs in *CYP2D6* and also identify possible missed haplotypes in *CYP2C9*, *CYP2C19* and *CYP2D6* (Chapter 5). BAM files were uploaded and depth of coverage analysis was used to detect duplications and deletions by comparing depth of coverage of *CYP2D6* to depth of coverage of a control region outside the gene. To call haplotypes, variants present in the BAM file were compared to sets of variants present in each of 7140 diplotypes and scored. This was then adjusted to take account of the sensitivity and specificity of scoring for each variant. The diplotype with the highest score was then called. Astrolabe calls were then checked manually in IGV (section 2.4.2.1) and a final decision was made as to whether the Astrolabe diplotype or that originally called from WGS data was upheld. Astrolabe analysis was performed with the assistance of Dr A. Gagunashvili.

2.4.4 Lumpy

Lumpy, a freely available bioinformatics programme, was used to detect possible CNVs(84). BAM files of discordant read pairs and split reads were prepared and Lumpy was run. Resulting vcfs were analysed for CNVs in candidate genes of interest using IGV. Lumpy analysis was performed with the assistance of Dr A. Gagunashvili.

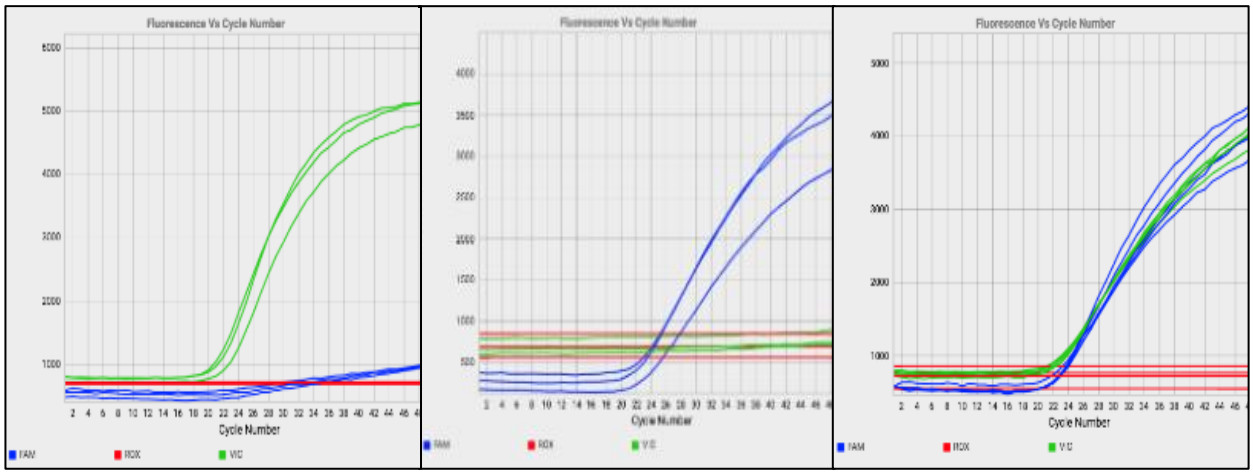

Figure 2.3 Real time plots visualised in ThermofisherConnect™ showing allele 1 homozygotes (left), heterozygotes (centre) and allele 2 homozygotes (right) labelled with FAM (blue) and VIC (green). Fluorescence shown on x axis, number of cycles on y axis

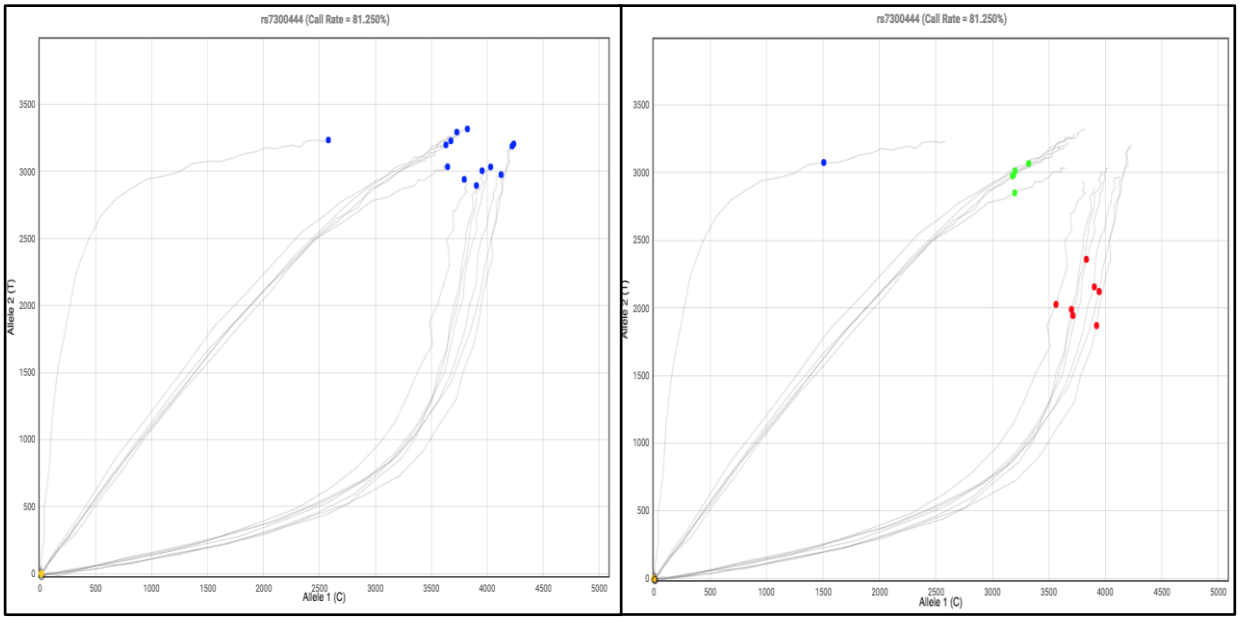

Figure 2.4 Differences in clustering depending on cycle number. Cycle 40 (left) and cycle 36 (right). Allele 1 on x axis, allele 2 on y axis

2.5 Determination of pathogenicity

2.5.1 American College of Medical Genetics (ACMG) guidelines

In the case of samples BBS-016, BBS-017 and BBS-018, after filtering (section 2.4.2.2) a shortlist of coding and non-coding variants with possible functional effects in genes known to be associated with BBS, ciliopathies or cilia structure or function was drawn up. The ACMG guidelines were used to assess the likely pathogenicity of each variant(85). These have been described and discussed in Chapter 3. To be considered, variants had to be at least as rare as the commonest BBS mutation, p.Met390Arg (section 3.4.1.2).

2.5.2 Literature search

A literature search using PubMed (www.pubmed.com) and Google (www.google.com) search engines was carried out to see if the variants had been seen before. Variants were searched for using both forms of protein variant nomenclature (e.g. p.M390R and p.Met390Arg).

2.5.3 Protein networks

Protein networks were examined using STRING(86). Gene names were entered by uploading a text file containing the HGNC-approved gene symbol. The output was saved as a jpg file.

2.6 Pharmacogenomic analysis and guidelines

2.6.1 Prescribing guidelines

Prescribing guidelines were extracted manually from www.pharmGKB.org in September 2015. A review was done in January 2018 to update changed guidelines or any newly added ones, and results were updated accordingly. Guidelines were mainly from the Clinical Pharmacogenetics Implementation Consortium (CPIC) and the Dutch Pharmacogenomics Working Group (DPWG), although there were some additional guidelines from the Canadian Pharmacogenomics Network for Drug Safety (CPNDS). In total, 100 different guidelines related to 72 different drugs were extracted (section 5.1.2.2). Comparison of the guidelines was done and later checked with the aid of a comprehensive comparison publication(87). A table was made of guidelines by haplotype and drug for ease of analysis (Supplementary Information S2.1 (CD-ROM)).

2.6.2 Haplotype and genotype data

Haplotype and genotype data were extracted from www.pharmGKB.org haplotype definition tables for each pharmacogene that had associated guidelines in September 2015. A review was done in January 2018 to update any changed haplotype definitions (section 5.3.5.7.2).

2.6.3 Haplotype and genotype extraction from WGS data

Only haplotypes or genotypes with associated pharmacogenomic prescribing guidance were checked (section 5.2.1). This was done using IGV as previously described. All SNPs identified were recorded and the haplotypes and diplotypes determined (Tables 5.1-5.11). Data were extracted for all 84 individuals in the cohort. *CFTR* was excluded as it would constitute a carrier test (see Chapter 5). *HLA-B* *44 and *HLA-B* *58:01 were also excluded as they are sequence

rather than SNP-based (see Chapter 5). Further data were later extracted to include additional pharmacogenes on the Congenica panel.

2.6.4 Individual prescribing guidance

Genotypes and diplotypes for each individual were compared to published pharmacogenomic prescribing guidelines and individual prescribing advice prepared. Two forms of the advice were prepared, a long form as an Excel spreadsheet and a summary form (section 5.2.2). Prescribing recommendations were analysed to see how many patients had variants in actionable pharmacogenes. The phenotyping database was interrogated to determine whether patients had been prescribed any drugs metabolised by actionable pharmacogenes in which they had variants, and the probable effect of this was determined.

2.6.5 Haplotype frequency calculations

Haplotype frequencies were obtained from PharmGKB or from the literature. Frequencies for the cohort were obtained by excluding one of each of the pairs of monozygotic twins and the children from the SRS trios, leaving a total of 72 individuals. Haplotype or allele frequency was compared to the published data. The overall data for each haplotype or allele were compared with a Fisher's exact test, while individual haplotype or allele frequencies were compared with 95% confidence intervals (CI) and a two proportion z-test. Fisher's exact test is used to analyse contingency tables. If the null hypothesis is true, i.e. if the tables are different, the p value will be less than 0.05. If the tables are not significantly different, the p value will be greater than 0.05. 95% confidence intervals give a range of values into which the observed value will fall with 95% certainty. In this case, if the observed frequency were the same as the published frequency, one would expect the published figure to lie within the 95% confidence interval calculated for the observed figure. The two proportion z-test compares an observed proportion to an expected proportion and sees whether they are significantly different (p value less than 0.05) or not.

2.7 Patient databases and data collection

2.7.1 Patient data collection

Patient data were collected by me from a number of sources. In the case of the BBS, IBD and USH patients these included patient notes, both paper and electronic, held at GOSH and Guy's and St Thomas' Hospital NHS Foundation Trusts and Moorfields Eye Hospital. It also included pathology, radiology and e-prescribing databases at those hospitals. In the case of the JDM patients, clinical data were obtained from a clinician in summary form, and included phenotypic descriptors and pathology and radiology results. In the case of the SRS patients, no clinical data were available beyond ethnicity.

2.7.2 Patient data recording and output

All patient data were recorded on the UCL Data Safe Haven (IDHS) in accordance with NHS Information Governance Toolkit and the ISO27001 information security standard, and which is approved for the storage of sensitive clinical data belonging to NHS patients. The IDHS is

protected by pre-allocating storage to a named researcher. They have a password and a secure key to access data. Individual applications within IDHS were also password protected.

2.7.2.1 Phenotips database

Clinical descriptors, identified from the sources described in section 2.7.1, were entered as Human Phenotype Ontology (HPO) terms into the PhenoTips® database which was installed on the UCL IDHS with the assistance of UCL information technology staff following approval by the UCL IDHS committee. Updates to HPO terms were checked for every 6 months and the IDHS version of PhenoTips® updated accordingly. With the entry of each phenotypic descriptor an attempt was made to find the most specific HPO term to describe the clinical feature. Significant negative features were recorded as not present. Additional data such as age, height and weight were also recorded. Only patients from the BBS, IBD, JDM and USH cohorts had data entered into PhenoTips® as the SRS cohort had no clinical data available. JDM patients were recorded as having or not having a list of important features of JDM. These were also entered in the MS Access database along with the dates when they were recorded. Only data for patients USH-001 and USH-002 were recorded. USH-003 was not included in phenotyping as they were unaffected and no clinical data were available.

Outputting the PhenoTips® database was done using the export data option, which allowed data to be exported as a JSON or text file, options which are listed in the “other actions” tab of PhenoTips®. At this point it was possible to select which data types were to be exported and this was the point at which anonymisation occurred, with the patients being identified only by their unique cohort number e.g. BBS-001. Other data, such as names and dates of birth, were not included. Before transfer out of IDHS, the files were checked for anonymisation. Again, these spreadsheets and text files were distributed to a single, named researcher using the IDHS file transfer tools and could not be deanonymised except by the original researcher.

2.7.2.2 Microsoft Access (MS Access) database

A custom MS Access database was built by me to contain all clinical information apart from HPO terms. It was built as a relational database with a patient information table as the central table. It contained 21 static lookup tables and 15 tables of patient specific data (Tables 4.2 and 4.3). Forms were created for accurate and efficient data entry and queries were created to output data as required in anonymised and non-anonymised forms. Data were entered from the sources described in section 2.7. In the case of blood and pathology results, data were recorded only when they were associated with a clinician appointment or another investigation such as a colonoscopy. This is further discussed in section 4.3. Reference ranges were recorded as these were variable. Abnormal results were flagged using the formula “IIf([ResultValue]>= [MinNormalValue] And [ResultValue]<= [MaxNormalValue],1, IIf ([ResultValue]< [MinNormal Value], 2, IIf([ResultValue]> [MaxNormalValue],3,0))). This was then converted into a tick box normal/high/low result using the formula IIf(IsNull([MinNormalValue])OrisNull ([MaxNormal Value]), IIf([IsHigh]=-1,3,IIf([IsLow]=-1,2, IIf([IsNormal]=-1,1,0))),0) where values of one, two and three were converted into a tick in the normal, low and high columns respectively.

For transfer to other researchers data were outputted anonymously as required by IDHS. This was done in several ways. In general, researchers received outputted text files with the patient represented solely by their cohort number. Dates of birth were removed and replaced with an age if that information was deemed necessary. Hospital numbers and other identifiers were also removed. As data were generally outputted as text files of MS Access queries (section 4.1.4.1), it was easy to anonymise information as any patient identifiable information was excluded at the query design stage. Data were exported using the MS Access external data tools, where there are options to export as an Excel spreadsheet, a text file, an XML file or a PDF. In the case of the HIGH-5 study, data were given to researchers as anonymised comma delimited text files.

Had the entire database been required by a researcher, the plan was to anonymise by removing unnecessary identifiers such as hospital numbers and dates of birth and replace both the first and last name with the cohort number so that a patient called Joe Bloggs who was enrolled as the 19th BBS patient would be anonymised as BBS-019. However, this was not necessary. Data were transferred out of the IDHS by a single researcher using the IDHS secure file transfer service to a single individual. All data were checked for complete anonymisation first. The data could not be deanonymised without access to either PhenoTips® or the MS Access Phenotyping database within IDHS. These were protected within IDHS by a password and digital secure key to access IDHS and then each with their own password for additional security.

An anonymised and an empty version of the MS Access database can be found in Supplementary Information S2.1 and S2.2 (CD-ROM). The anonymised database contains clinical details for the BBS, IBD, JDM and USH patients. The SRS patients are not included as no data were available, nor was USH-003, who was a control. Several extra patients were included (IBD-021 and JDM-014- JDM-018). These patients did not have WGS so were not included in the pharmacogenomics chapters, but were included here as the data were important for other multiomics analyses. Not all data in the lookup tables were required for this part of the project but was put there for the long-term utility of the database as part of the HIGH-5 project. Tables prefixed with tbl_L e.g. tbl_L_BloodMarkers are look-up tables and contain background data that can be incorporated into other tables. Tables prefixed with tbl_ e.g. tbl_BloodTestResults contain entered data about patients. Tables prefixed with x_ e.g. x_BBS_PatientInformation are the results of queries, while qry_ e.g. qry_BBS_BloodTestResults denotes queries used to extract these data. The prefix frm e.g. frmBloodTestResult denotes forms that allow the easy entry of data. In the case of forms, all data that have been entered by this method is visible and one must use the arrows beneath to scroll forward to an empty form to enter data. The empty database has had all patient information and queries removed but information in the look-up tables and the forms used for data entry remain.

2.7.3 Data use

The databases were interrogated for clinical information for the diagnostics chapter, prescribing information for the pharmacogenomics chapter and by researchers including Dr Rosalind Davies and Dr Jochen Kammermeier for burden analysis, patient stratification and multiomics analyses.

Chapter 3 Utilisation of whole genome sequencing for diagnostics

3.1 Introduction

There are many reasons to seek a molecular diagnosis for a rare disease. It informs management and may give information about prognosis, treatment and recurrence risk for parents, siblings or offspring of affected individuals. Families may find it useful in less concrete ways such as accessing support, advocating for the individual or reducing feelings of guilt(88-90). The process of attaining a molecular diagnosis has changed considerably in cost and complexity in recent years and obtaining a molecular diagnosis is now cheaper, faster and more likely to be successful than ever before. It is estimated that up to 50% of genes causing rare monogenic diseases have been discovered, and diagnosis is now possible antenatally(91, 92). However, patients with rare diseases can still wait many years for clinical and molecular diagnoses and may see many different clinicians prior to this(93). Improved methods of molecular diagnosis are shortening these “diagnostic odysseys” but they still have a great financial, physical and psychosocial cost(94, 95).

3.1.1 Methods of molecular diagnosis

Many molecular diagnostic methods are available for single gene disorder diagnosis. They vary in cost, scope and labour-intensiveness, and each may be appropriate in specific clinical situations.

3.1.1.1 Sanger Sequencing

Sanger sequencing was developed in 1977 by Frederick Sanger et al.(96). It utilises DNA polymerase to selectively incorporate dideoxynucleotidetriphosphates (ddNTPs) in place of normal deoxynucleotidetriphosphates (dNTPs). Primers bind to the specific section of DNA whose sequence is required. Each time a ddNTP is incorporated by the polymerase in place of a dNTP, which are present in higher concentrations, the polymerase drops off and the chain terminates. Originally four separate reactions were run, but later, fluorescently labelled ddNTPs were developed so that automated Sanger sequencing is possible(97). Sanger sequencing is used regularly in the NHS for predictive testing, confirmation of pathogenic variants found in a research setting and first line testing in diseases where small numbers of variants are responsible for most cases of the disease. It remains the gold standard for variant confirmation.

3.1.1.2 SNP arrays

A SNP array is a DNA-based microarray which can range from condition-specific to genome-wide. The method involves applying fragmented nuclear DNA from a patient to an array containing allele-specific oligonucleotide (ASO) probes for both the major and minor alleles which are

immobilised on a solid surface. Hybridisation of the patient DNA to the target causes a signal to be released which can be detected and called automatically(98). This is rarely used in diagnostic testing of single gene disorders but has many other applications, for example in genome-wide association studies and the study of malignancies(98-100).

3.1.1.3 Next-generation sequencing

Next-generation sequencing (NGS) methods, which are also known as high-throughput, deep or massively-parallel sequencing, are new and rapid methods of nucleic acid sequencing. There are multiple platforms for performing NGS, but the principle is of “sequencing by synthesis” where millions of parallel DNA fragments are synthesised simultaneously based on template patient sequence. The resultant overlapping reads allow the determination of longer sequences(101). Raw data are run through bioinformatics pipelines to align reads to a reference genome and highlight any deviation from it. A higher number of reads, or better read depth, improves the accuracy of sequencing and variant calls. While NGS was expensive, costs have fallen rapidly. The NIHGR put the cost of sequencing the first human genome at over one billion dollars. In 2001 a genome cost $100,000,000. By 2014 this had fallen to $4,000 and to $1500 by 2015(102). It is now possible to obtain whole genome sequences for under $1000 per genome and in 2017 Illumina announced that it expects to reduce the cost to $100(103).

3.1.1.3.1 Panel-based next generation sequencing

Panel-based NGS involves sequencing a pre-selected number of genes, such as those known to cause a condition or a group of phenotypically overlapping conditions(104-106). The advantages of a panel-based approach include cost-effectiveness and a reduction in the risk of finding variants of uncertain significance (VUS) in genes not known to be causative of the disorder in question and significantly reduces the risk of discovering secondary findings, for example, pathogenic mutations in genes causing adult-onset disease. However, gene discovery is not possible with this approach, and panels need alteration if a new gene is discovered.

3.1.1.3.2 Whole exome sequencing

Whole exome sequencing (WES) was first used to identify the cause of a Mendelian disorder in 2010, when, following the sequencing of just four affected individuals from three unrelated families, variants in *DHODH* were identified as causing Miller syndrome(107). WES involves the sequencing of the coding regions of the human genome, amounting to approximately 180,000 exons or less than 2% of the total genome(108). An advantage of this approach is that it does not require prior knowledge of causative genes, allowing for gene discovery or review of results when other genes are discovered. Disadvantages include the large number of VUS seen in genes that may or may not relate to the disorder being investigated and the possibility of identifying pathogenic variants causing diseases unrelated to the disorder, often referred to as incidental or secondary findings, and the difficulties that arise in relation to informing patients of these(109-111). The American College of Medical Genetics (ACMG) has released guidance for dealing with incidental findings and recommends seeking and reporting variants in a number of genes, mainly for hereditary cancers and cardiac syndromes(5). It also highlights the importance of informing

patients of the possibility of incidental findings before testing is carried out. Studies have shown that patients are broadly in favour of the disclosure of incidental findings(112-114).

3.1.1.3.3 Whole genome sequencing

Whole genome sequencing (WGS) involves sequencing both coding and non-coding regions of the genome, approximately 3 billion base pairs. WGS has advantages over WES including increased detection of coding, copy number and mosaic variants, and the ability to look at pathogenic non-coding variants(115-118). Its main disadvantages are those of WES though at greater scale, the computational requirements for analysis and storage and the increased cost. The increased number of variants being detected means that validation and/or functional studies cannot be done on all variants and there is evidence that variants of indeterminate pathogenicity have made their way into the literature as pathogenic variants(111.). In addition, because costs are higher, a lower read depth may be chosen, which may lead to the mutant allele being missed and a false negative result(101).

3.1.2 Variant interpretation and classification

Once a variant has been identified by NGS, its pathogenicity needs to be determined. This is vital as otherwise relevant results may be unreported or non-pathogenic variants reported as disease-causing, with consequences for family screening, recurrence risk, treatment options and more.

3.1.2.1 Categories of variant

When determining pathogenicity of variants, priority is given to functional and coding variants and various systems, both general and disease-specific, have been proposed to help with classification(85, 119-122). The ACMG guidelines are widely used and divide variants into pathogenic, likely pathogenic, variant of unknown significance, likely benign or benign when a variant has not previously been determined to be pathogenic(85).

3.1.2.1.1 Pathogenic and likely pathogenic variants

Table 3.1 sets out the ACMG guidelines for determining pathogenicity and Table 3.2 shows how they can be combined for an overall likelihood of pathogenicity. A similar set of guidelines exist for determining benignity(85). Contradictory lines of evidence automatically result in a variant being called as a VUS. Despite the widespread adoption of the guidelines, they have not yet led to uniformity in the determination of pathogenicity by laboratories(4, 123).

3.1.2.2 Tools for determining pathogenicity

3.1.2.2.1 Databases

Various population and disease specific databases can be used to determine whether a variant is rare or common and whether or not it has previously been associated with disease(76, 77, 79, 80). Some of these databases, such as ExAC, 1000 genomes and GnomAD, are useful for determining whether a variant is rare or common in a population of interest, while others, such as ClinVar and the Human Gene Mutation Database (HGMD) collate information about the

pathogenicity of a variant. In addition, databases such as GeneMatcher, where researchers publish variants of unknown significance along with phenotypic information, may be of use(124).

Very strong evidence of pathogenicity	
PVS1	Null variant where loss of function is a known disease mechanism
Strong evidence of pathogenicity	
PS1	Same amino acid change as a previously established pathogenic variant
PS2	*De novo* in a patient with no family history (dominant)
PS3	Well established functional studies supportive of deleterious effect on gene/product
PS4	Prevalence of variant significantly increased in affected compared to controls
Moderate evidence of pathogenicity	
PM1	Located in a well-established hot spot and/or critical functional domain
PM2	Absent from controls (dominant) or at extremely low frequency (recessive)
PM3	In *trans* with a pathogenic variant (recessive)
PM4	Change in protein length
PM5	Novel missense amino acid change where a different missense change has been determined to be pathogenic
PM6	Assumed *de novo* but without confirmation
Supporting evidence of pathogenicity	
PP1	Co-segregation with affected family members in a gene known to cause disease
PP2	Missense variant in a gene with a low rate of benign missense variation and in which missense variants are a common mechanism of disease
PP3	Multiple lines of computational evidence support a deleterious effect
PP4	Phenotype/history is highly specific for a disease with single genetic aetiology
PP5	Reputable source recently reports variant as pathogenic but evidence unavailable

Table 3.1 Interpretation of variants- adapted from Standards and Guidelines for the Interpretation of Sequence Variants, ACMG(85)

Class Five- Pathogenic	
One very strong (PVS) AND	One or more strong (PS) OR
	Two or more moderate(PM) OR
	One moderate (PM) and one supporting (PS) OR
	Two or more supporting (PS)
OR two or more strong (PS)	
OR one strong (PS) AND	Three or more moderate (PM) OR
	Two moderate (PM) and two or more supporting OR
	One moderate (PM) and four or more supporting (PS)
Class Four- Likely Pathogenic	
One very strong (VS) AND	One moderate (PM)
OR one strong (PS) AND	One or two moderate (PM)
OR one strong(PS) AND	Two or more supporting (PS)
OR three or more moderate (PM)	
Or two moderate (PM) AND OR	Two or more supporting (PS)
Or one moderate (PM) AND	Four or more supporting (PS)
Class Three- Variant of unknown significance	
Criteria for pathogenic, likely pathogenic, likely benign or benign not met OR	
Conflicting benign (see paper) and pathogenic criteria	

Table 3.2 Classification of variants- adapted from Standards and Guidelines for the Interpretation of Sequence Variants, ACMG(85)

3.1.2.2.2 *In silico* prediction software

In silico prediction tools such as PolyPhen2, PROVEAN and SIFT use algorithms to determine how likely variants are to be pathogenic(82, 125). They operate by looking at evolutionary conservation of DNA and the probable effect of amino acid changes on protein structure and function. Their accuracy is estimated at between 60 and 80%(81, 126, 127). Currently prediction tools are not very useful for non-coding sequence, although specific splice-site prediction software is available(128).

3.1.2.2.3 Family studies

Family studies are important for determining whether potentially pathogenic alleles are in *cis* or in *trans* and whether variants segregate with a disease in a kindred. However, this is often complicated by non-availability of samples and may not always be possible.

3.1.2.2.4 Functional studies

Functional studies take many forms, from quantification of RNA or protein to looking at cellular and animal models containing the variant or variants of interest. The main limitation of these studies are that they are time-consuming and expensive and cannot be done for more than a few variants of interest, even in small scale studies of a single patient. They are much more difficult for large-scale studies where many variants of interest may be identified in multiple genes in a cohort.

3.1.3 Types of disease-causing variant

Disease causing variants, also known as pathogenic variants or mutations, are commonly found in coding regions of the DNA, but can also be found in non-coding regions if they affect regulatory elements such as promotors or splice sites. There are three main categories of variant.

3.1.3.1 Substitutions

Substitutions occur when one base is substituted for another. Owing to redundancy in the genetic code, a substitution doesn't necessarily affect the amino acid sequence in which case the substitution is termed a synonymous variant. These rarely cause disease but can do so, for example when they affect an existing splice site or result in the formation of a novel splice site(129-131). If the amino acid sequence is altered, the substitution is termed non-synonymous. Missense variants cause one amino acid to be substituted for another. These may have no or minor effects, or may cause disease. They are the most commonly implicated type of variant in inherited disease and exert their effect in many ways(132). Occasionally a substitution will result in a codon for an amino acid being replaced with a premature stop codon, in which case they are termed nonsense variants(133, 134). These usually cause disease, often but not always by a mechanism known as nonsense-medicated decay, in which mRNA containing a premature stop codon is degraded rather than being translated to form a shortened polypeptide. However caution should be exercised when interpreting nonsense variants found at the extreme 3' end of genes, those affecting splice sites or when the gene has multiple transcripts(85).

3.1.3.2 Insertions

Insertions involve the addition of one or more nucleotides to a DNA sequence. They can occur alone or in combination with deletions in which case they are referred to as indels. As with substitutions, they may or may not cause disease. If the number of nucleotides inserted can be divided by 3, the open reading frame of the gene is not disrupted and the insertion may have no effect, even if the insertion is large. However, if the number is not divisible by three, the open reading frame will be disrupted. This is termed a frameshift and is likely to result in the formation of a premature stop codon and may result in nonsense mediated decay(135, 136).

3.1.3.3 Deletions

Deletions involve the removal of one or more nucleotides from a DNA sequence. As with insertions, the consequences are more severe when numbers of nucleotides not divisible by three are involved, resulting in a frameshift with the consequences described above, though in-frame deletions are also described in human disease(137, 138). One of the best known deletion variants is p.Phe508del in *CFTR*, which results in abnormal maturation and abnormal transport of the cystic fibrosis transmembrane regulator(139).

3.1.3.4 Structural changes and expansions

There are various other ways in which DNA can be disrupted. These include gene deletions and duplications, larger copy number variants (CNVs), insertions, where a piece of DNA is inserted into a gene disrupting the normal reading frame, inversions, where a piece of DNA is inserted in the correct location but the incorrect direction and expansions, where repetitive sequences of DNA are repeated more times than they should be.

3.1.3.5 Variant nomenclature

Although there are multiple systems for naming variants, the standardised system proposed by the Human Genome Variation Society (HGVS) in 2000 and updated in 2016 has been widely adopted(140, 141). This uses a prefix to indicate whether the sequence is DNA (genomic, coding, non-coding or mitochondrial), RNA or protein. A suffix or symbol is used to indicate the variant type, such as > for a substitution, ins for an insertion and del for a deletion. Clear rules are set out for numbering nucleotides. At the protein level the use of three letter amino acid codes are preferred, for example p.Met390Arg instead of p.M390R.

3.1.4 Cilia and ciliopathies

3.1.4.1 Cilia structure

Cilia may be categorised as motile or non-motile. Motile cilia are microtubule-based structures found on the surface of certain cells and which move with a beating motion. Multiple motile cilia are found on the surface of a cell and the cilia in a particular region often coordinate their movements. They contain dynein arms which allow their movement. Examples of motile cilia are those found in the trachea which move mucus back towards the oral cavity and those in the middle ear(142). Defects in motile cilia-associated genes cause problems including lung problems

secondary to mucociliary clearance failure, sub-fertility, problems with laterality and central nervous system defects.

Single non-motile or primary cilia are found on the apical surface of almost all vertebrate cells, with the exception of cells derived from bone marrow and the intercalated cells of the collecting duct of the kidney(143). They are microtubule-based structures that play multiple roles in cell signalling. Non-motile cilia have a 9+0 structure, where there are 9 outer tubule doublets with no central microtubules (Figure 3.1). This differs from the 9+2 structure of the motile cilium where there is a central microtubule pair(144). Non-motile cilia also lack the dynein arms seen in motile cilia and are immobile. The microtubule structure, known as the axoneme, extends from the basal body and is enveloped by a specialised plasma membrane known as the ciliary membrane(145). The basal body, which also contains a ring of microtubules, is formed from the original centriole inherited during mitosis. As well as providing a template for the formation of the microtubules of the axoneme, it is involved in anchoring the cilium(146). Cilia form when cells are in the G_0 phase of the cell cycle(147). Cilia cannot synthesise proteins, but instead transport them in from the Golgi apparatus and endoplasmic reticulum by a process known as anterograde intraflagellar transport (IFT). As well as transporting proteins into the cilium, IFT is involved in transporting them out, which is known as retrograde IFT. At or around the basal body are docking sites for IFT proteins, which are then directed to the cilium(148). Several genes mutated in Bardet Biedl syndrome (BBS) encode proteins that form a complex known as the BBSome, which consists of 8 proteins and is required for ciliary formation (Figure 3.2)(149, 150). These proteins are BBS1, BBS2, BBS4, BBS5, BBS7, BBS8, BBS9 and BBIP1, which is also known as BBS18 or BBIP10, and variants in any of them can cause BBS(151). The BBSome plays an important role in IFT(152). There are many genes involved in maintaining the structure and function of normal cilia(153, 154).

3.1.4.2 Cilia function

Cilia are involved in several cell signalling pathways. Huangfu et al. showed that IFT proteins play a vital role in the Hedgehog (HH) signalling pathway, with disruption of IFT proteins resulting in lower levels of HH(155). HH signalling plays important roles in embryonic development including in embryonic polarity and tissue differentiation and also in post-embryonic tissue regeneration(156). It has been shown that the proteins required for cilia formation are also required for HH signalling(157).

Cilia are important in Wnt signalling. Three Wnt signalling pathways have been identified- the canonical Wnt pathway (WNT/β-catenin pathway) and two non-canonical pathways, the Wnt planar cell polarity pathway and the Wnt/calcium pathway. Canonical Wnt signalling results in β-catenin accumulating and being transported into the nucleus where it activates other transcription factors, resulting in cell proliferation. Non-canonical wnt signalling is independent of β-catenin and has roles in tissue differentiation and cell polarity(158-160). Cilia appear to contribute to the regulation of Wnt signalling by mediating switching from the canonical to non-canonical pathways, though are not essential for it(161-164).

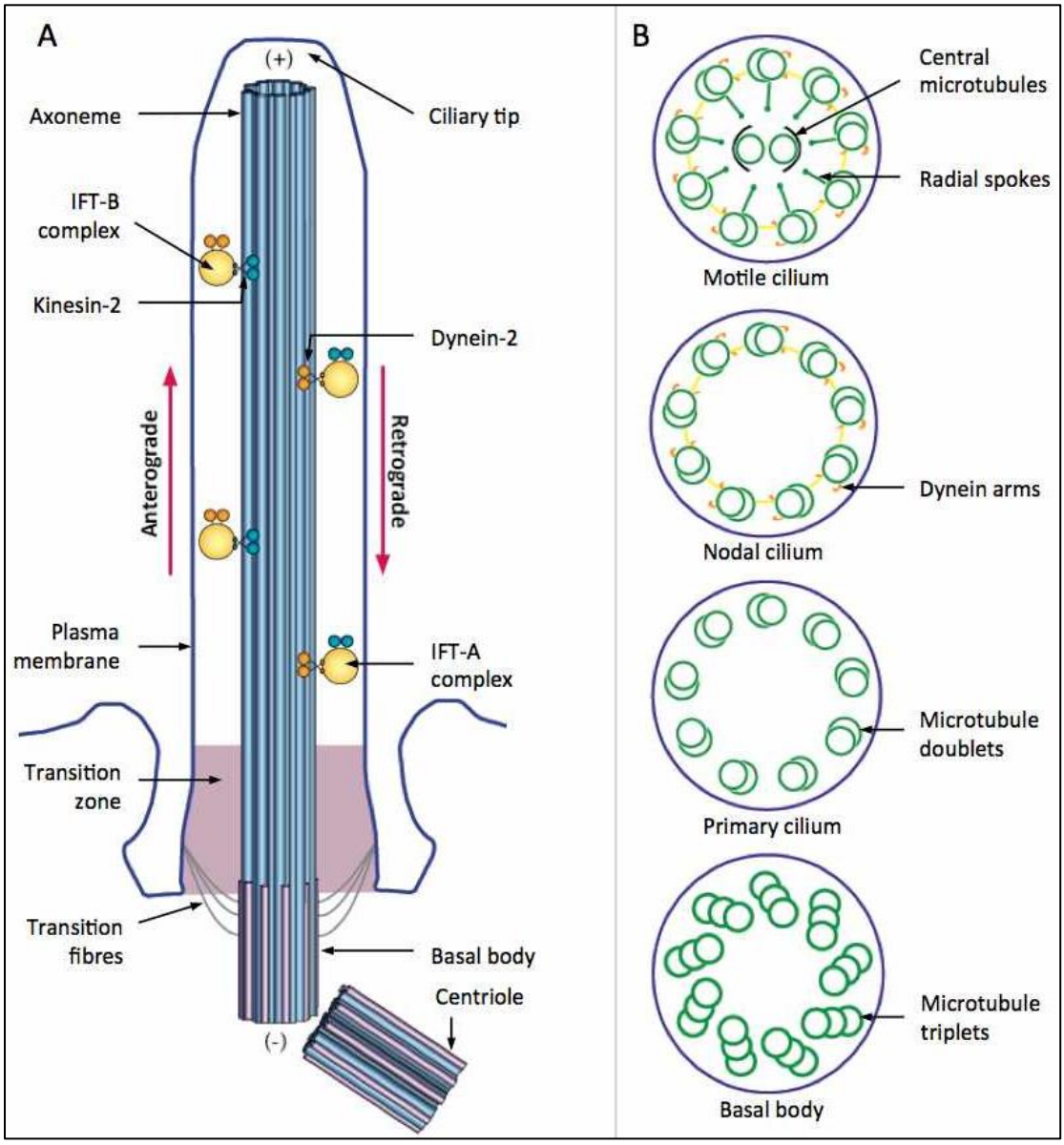

Figure 3.1(A) Structure and function of the primary cilium and (B) Motile and non-motile cilia in cross-section. Image courtesy of Dr Rosalind Davies, UCL Great Ormond Street Institute of Child Health

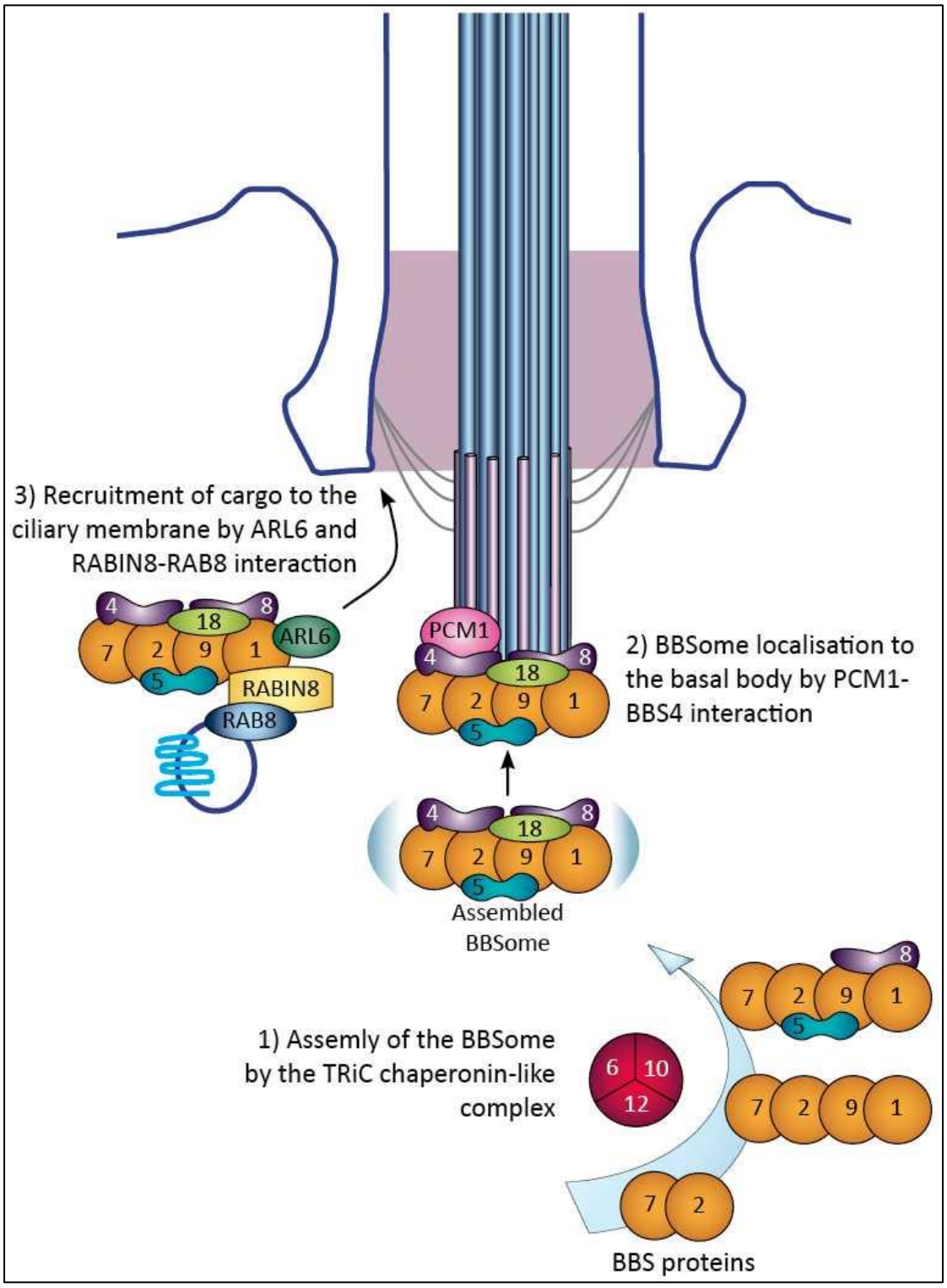

Figure 3.2 Structure and assembly of the BBSome. Image courtesy of Dr Rosalind Davies, UCL Great Ormond Street Institute of Child Health

Cilia appear to play a role in G protein-coupled receptor (GPCR) signalling. GPCRs have an extracellular domain which is activated by a ligand. Ligand binding activates internal signal transduction pathways. There are about 800 different GPCRs, many of unknown function. They are activated by a huge range of substances including many drugs. Cilia are important in the trafficking of GPCRs and some GPCRs are specifically targeted to cilia, such as the MCHR1 GPCR which is found on the cilia of the hypothalamus(165, 166). Some non-motile cilia, especially in the kidney, are thought to have a mechanosensory role, for example when bending of cilia appears to cause increased intracellular calcium levels although recent research has cast some doubt on this(167-169).

3.1.4.3 Ciliopathies

The ciliopathies are a heterogeneous group of inherited diseases caused by variants in genes coding for ciliary proteins(142). They share clinical features and causative genes. The advent of molecular diagnosis has shown that any of a number of ciliopathies may be caused by a variant in a single gene. Rare families have been reported with members having different ciliopathies despite sharing the same variants(19). The wide range of clinical features seen in ciliopathies demonstrates the many and varied roles of non-motile cilia (Figure 3.3).

3.1.4.4 Features of ciliopathies

Cilia are important in the development of the embryonic brain and have roles in determining cell differentiation, migration and maturation. Both structural and functional abnormalities of the brain are seen in ciliopathies, such as cerebellar vermis hypoplasia in Joubert syndrome (JBTS) and the intellectual disability common in BBS and MORM (mental retardation-obesity-retinal dysplasia-micropenis) syndrome(170, 171).

The outer segment of the photoreceptor of the eye is a specialised primary cilium, containing many proteins involved in photosensitive signal transduction, such as opsin and transducin(172). Retinal dystrophy is frequently seen in ciliopathies, from the congenital retinal dystrophy of MORM syndrome to the later onset retinal dystrophy of BBS and Usher syndromes.

Many ciliopathies involve the kidneys and cyst formation is frequently seen(173). This varies enormously in age of onset and severity. Renal problems range from mild to severe. End-stage renal failure (ESRF) requiring transplantation may result. Structural renal problems such as horseshoe kidneys are also seen(174).

Cilia are involved in the development of bone and cartilage, so skeletal abnormalities and dysplasia are an unsurprising feature of ciliopathies such as Jeune syndrome (JATD) and orofaciodigital syndrome (OFD)(175). Polydactyly is another frequent feature, and is presumed to be related to HH signalling disruption. Other organs such as the liver, pancreas and lungs may also be affected(142).

	ALMS	BBS	EVC	JATD	JBTS	OFD	LCA	MKKS	MKS	MORM	NPHP	SLS	USH
Nervous system abnormalities													
Craniofacial abnormalities													
Deafness													
Genital abnormalities													
Intellectual disability													
Obesity													
Polydactyly													
Renal abnormalities													
Retinal dystrophy													
Skeletal abnormalities													

ALMS- Alström syndrome, BBS- Bardet Biedl syndrome, EVC- Ellis van Creveld syndrome, JATD- Jeune asphyxiating thoracic dystrophy, JBTS- Joubert syndrome, OFD- orofaciodigital syndrome, LCA- Leber congenital amaurosis, MKKS- McKusick-Kaufman syndrome, MKS- Meckel syndrome, MORM- mental retardation-obesity- retinal dystrophy-micropenis, NPH- nephronophthisis, SLS- Senior-Løken syndrome, USH- Usher syndrome

Figure 3.3 Features of ciliopathies. Paler shading indicates a more rarely seen feature

3.1.4.5 Genes implicated in ciliopathies

Table A1.4, Appendix 1 shows that the majority of ciliopathies can be caused by variants in a number of genes and that each gene can cause a number of ciliopathies. Variants in some genes in particular can cause a wide variety of phenotypic presentations. For example *MKKS* causes Bardet Biedl, McKusick-Kaufman and Meckel syndromes, while *MKS1* causes Bardet Biedl, Joubert and Meckel syndromes(142). *CEP290* has been implicated in Leber congenital amaurosis and Bardet Biedl, Joubert, Meckel and Senior-Løken syndromes(176-180). This variation can be a function of the variant type or location, or may be due to additional modifiers(181-183).

3.1.5 Bardet-Biedl syndrome

Bardet-Biedl syndrome (BBS) is a rare, autosomal recessive ciliopathy. It is seen at prevalence of between 1 in 100000 and 1 in 160000 in Europe and North America, but at a much higher prevalence in some genetically isolated or consanguineous populations(18). It was first described in 1866 by Laurence and Moon, and then independently by both Bardet and Biedl in the early 1920s. From the mid-1920s the term Laurence-Moon-Bardet-Biedl was used as the patients had overlapping clinical features. However, those described by Bardet and Biedl had polydactyly in addition to the retinitis pigmentosa (RP), obesity and intellectual impairment described by Laurence and Moon. Biedl's kindred had hypogonadism in addition. Generally, the syndrome is now referred to as Bardet-Biedl syndrome(184-186).

3.1.5.1 Features of BBS

The main features of BBS are rod-cone dystrophy, obesity, polydactyly, renal problems and genitourinary abnormalities, hypogonadism and learning difficulties.

3.1.5.1.1 Rod-cone dystrophy

More than 90% of patients with BBS develop a pigmentary rod-cone dystrophy, often called retinitis pigmentosa (RP). It generally presents in mid-childhood, with the majority of patients registered blind by their mid-teens or early twenties. The presenting feature is night blindness, followed by loss of central and colour vision, with the macula being affected relatively early in the process. It is diagnosed by electroretinography (ERG) and is often the presenting feature for individuals with BBS. Up to 75% of patients will be declared legally blind(18, 187). Photoreceptors contain a modified primary cilium known as the connecting cilium, which is anchored in the inner segment of the photoreceptor and extends to the outer segment. As in other primary cilia, IFT occurs in a bidirectional manner along the connecting cilium. IFT is necessary for the maintenance and function of cilia. In photoreceptors, the movement of important phototransduction proteins such as transducin, rhodopsin and arrestin along the cilium in response to stimuli is vital to the phototransduction cascade and normal vision. In BBS, the photoreceptors appear to form normally, but are gradually lost, leading to blindness(188, 189).

3.1.5.1.2 Obesity

Obesity is a very common feature of BBS with between 72% and 86% of patients developing it, often in early childhood(187). It increases the risk of complications such as diabetes and cardiovascular disease(190). BBS mouse models have been shown to have increased appetite and decreased activity, and although the mechanism for obesity is complex, leptin resistance and abnormal localisation of neuropeptide Y receptors appear to play a role(191-193). Recent research has implicated loss of function variants in *ADCY3* in severe obesity(194). ADCY3 localises to the primary cilia in neurons(195).

3.1.5.1.3 Polydactyly

Polydactyly, which refers to the presence of extra digits, affects approximately 60-80% of patients with BBS and is most commonly post-axial(187). It may affect all four limbs or hands or feet only, and may be anything from vestigial skin tags to fully formed digits. Syndactyly, brachydactyly and clinodactyly are also seen. Polydactyly is often the only feature of BBS present at birth, but rarely prompts diagnosis in the absence of a family history(196). The molecular basis of polydactyly appears to be disruption of Hedgehog (HH) signalling, which is known to be instrumental in limb bud development(197-199).

3.1.5.1.4 Renal disease and renal tract malformations

Renal abnormalities are found in approximately 50-80% of people with BBS(187, 200). The renal defects are variable and can include structural abnormalities of the renal tract, cystic disease of the kidney, renal dysplasia, glomerulosclerosis and others. In addition, diabetes and cardiovascular diseases such as hypertension can result in secondary kidney disease with onset in adulthood. The 2017 study by Forsythe et al. showed both adult and paediatric patients had similar rates of ESRF, 8% and 6% respectively, suggesting that the majority of BBS patients with ESRF develop it in childhood. Another study which looked at an international patient registry found a paediatric transplant rate of 10%(201). The products of the nephronophthisis and polycystic kidney disease genes have been localised to the cilia, confirming the role of the cilium in renal disease. The pathophysiology of renal disease in ciliopathies is not yet well understood, but cyst formation is likely to be a function of abnormal Wnt and HH signalling(154, 202).

3.1.5.1.5 Hypogonadism and genital anomalies

Estimates for genital anomalies range from 58-98%(187). Hypogenitalism may be present in males at birth, whereas females tend to have structural genital abnormalities of various types, which may not become apparent until later in life(18, 203). Delayed puberty is common, with some children requiring hormone replacement therapy to complete puberty. Males with BBS are usually infertile, while female fertility is variable and dependant on the structural genital abnormalities seen. Male infertility may result from the inability to form flagella during spermatogenesis(204). Again, HH signalling may play a role in the development of structural genital abnormalities(205).

3.1.5.1.6 Developmental delay and cognitive difficulties

When proposing the current diagnostic criteria, Beales et al. determined that about 50% of BBS patients, from a sample size of 112, had developmental delay. 62% of patients had learning

difficulties, which were generally in the mild to moderate range and 33% had behavioural problems(206). Cilia have been implicated in central nervous system (CNS) development, and more recently in learning and memory(207, 208).

3.1.5.1.7 Additional features of BBS

Many other features of BBS have been identified. These include a typical facial appearance with a high-arched palate and dental crowding, neurological abnormalities, speech problems, congenital heart disease and others(206).

3.1.5.2 Clinical diagnostic criteria for BBS

The diagnostic criteria currently in use for a clinical diagnosis of BBS are those proposed by Beales et al. in 1999(206). Features were divided into two categories, primary and secondary, and for a diagnosis to be made, patients require the presence of four primary or three primary and two secondary features (Table 3.3).

Primary features	Secondary features
Hypogonadism and/or genital abnormalities	Ataxia or poor coordination
Learning difficulties	Anosmia or hyposmia
Polydactyly	Brachydactyly and/or syndactyly
Obesity	Craniofacial dysmorphism
Renal abnormalities	Congenital heart disease
Retinal dystrophy	Dental abnormalities
	Diabetes mellitus
	Developmental delay
	Eye abnormalities
	Hepatic abnormalities
	Hirschsprung disease
	Hypotonia
	Speech delay

Table 3.3 Primary and secondary features of Bardet-Biedl syndrome (BBS)

3.1.5.3 Genetic basis of BBS

At present, biallelic variants in 21 genes are known to cause BBS (Table 3.4)(18, 174, 209, 210). However, as only about 80% of patients receive a molecular diagnosis, it appears that other genes may be involved. *BBS1* is the gene most often implicated, with a common biallelic variant in exon 12, p.Met390Arg accounting for up to 80% of *BBS1*-related BBS. Some genes appear to cause BBS only rarely and many are implicated in the causation of other ciliopathies.

Gene name	Other name	% of Cases	Other associated conditions
BBS1		23	
BBS2		8	Retinitis pigmentosa
BBS3	*ARL6*	0.4	Retinitis pigmentosa
BBS4		2	
BBS5		0.4	
BBS6	*MKKS*	6	McKusick-Kaufman syndrome
BBS7		2	
BBS8	*TTC8*	1	Retinitis pigmentosa
BBS9	*PTHB1*	6	
BBS10		20	
BBS11	*TRIM32*	0.1	Limb-girdle muscular dystrophy type 2H
BBS12		5	
BBS13	*MKS1*	4.5	Joubert syndrome, Meckel syndrome
BBS14	*CEP290*	1	Joubert syndrome, Meckel syndrome, Leber congenital amaurosis, Senior-Løken syndrome
BBS15	*WDPCP*	1	Congenital heart disease, hamartomas of tongue and polysyndactyly (CHDCHP)
BBS16	*SDCCAG8, NPHP10*	1	Senior-Løken syndrome
BBS17	*LZTFL1*		
BBS18	*BBIP1, BBIP10*		
BBS19	*IFT27*		
BBS20	*IFT74*		
BBS21	*C8orf37*		

Table 3.4 Genes causing Bardet-Biedl Syndrome., adapted from Forsythe and Beales, 2013(187)

3.1.5.3.1 Function and localisation of BBS genes

The products of *BBS1, BBS2, BBS4, BBS5, BBS7, TTC8 (BBS8), PTHB1 (BBS9)* and *BBIP1 (BBS18)* co-purify and together form the BBSome (Figure 3.2)(149, 151). The BBSome regulates IFT assembly and turnaround at both the ciliary base and tip, a process vital to the assembly, maintenance and signalling functions of cilia(152, 211). The BBSome appears to travel up and down the cilium in association with IFT cargoes. BBS1 controls interaction with RABIN8, the

guanine nucleotide exchange factor for the RAB8, which promotes ciliary membrane growth. BBS5 mediates binding to phospholipids, while BBS9 appears to be vital for complex organisation. BBIP1 (BBS18, BBIP10) appears to be responsible for acetylation and polymerisation of microtubules as well as forming part of the BBSome(151). Unlike the other components of the BBSome, knockout of *BBS1* or *BBIP1* mean that cilia do not form. The BBSome is recruited to the ciliary membrane by ARL6 (BBS3*)*(212). BBS6, BBS10 and BBS12 act as chaperones and are vital for correct assembly of the BBSome(213). TRIM32 (BBS11) is a ubiquitin ligase and is part of the ubiquitin/proteasome system, suggesting that proteasome degradation is important in BBS(214). *MKS1 (BBS13)* is essential for the migration of the centriole to the apical membrane, an early step in cilium formation(215). CEP290 (BBS14) is required for the recruitment of RAB8A to the centrosome and knockdown of CEP290 causes significant disruption to cilia formation. It may also have a role in controlling ciliary trafficking along with NPHP5(216, 217). WDPCP (BBS15) is involved in planar cell polarity and recruitment of septins and may play a role in the stabilisation of the plasma membrane. Knockout causes defective ciliogenesis(218). SDCCAG8 (BBS16, NPHP10) is a component of the centrosome and may have a role in DNA damage repair(219). LZTFL1 (BBS17) appears to regulate BBS complex trafficking(220). IFT27 (BBS19*)* is essential for normal IFT retrograde transport and knockout results in the accumulation of IFT proteins and abnormal cilia formation(221-223). IFT74 (BBS20) is involved in tubulin binding which is essential for ciliogenesis(224). The function of C8orf37 (BBS21) is not yet known but a recent study suggested that it is not a ciliary protein and may instead have a role in homeostasis of photoreceptor proteins(225).

3.1.5.3.2 Pathogenic variants in BBS

Missense, nonsense, frameshift and splice-site variants have all been reported to cause BBS, along with small and large deletions(18). Insertions are rare but have been reported, for example in *BBS16*(226). A common variant in *BBS1,* p.Met390Arg, causes about 80% of *BBS1*- related cases, and 30% of cases overall(227). There is also a recurrent one base pair insertion in *BBS10*, c.271dupT, which leads to a premature stop codon, and accounts for about 40% of *BBS10*-related cases(228). These variants are seen worldwide and so may either represent an ancient variant or a site of recurrent mutation.

3.1.5.3.3 Complex inheritance in BBS

While most cases of BBS seem to be inherited in an autosomal recessive manner, occasional families with more complex inheritance patterns have been identified. Katsanis et al. identified two families with biallelic variants in *BBS2* who did not appear to manifest the disease unless they also had a variant in *BBS6*(229). Beales et al. identified individuals homozygous for the *BBS1* p.Met390Arg variant who did not appear to have any features of BBS and further families have been reported who appeared to exhibit non-Mendelian inheritance of BBS(229, 230). One explanation for these findings is incomplete penetrance, a phenomenon seen in many Mendelian diseases. Other studies have not replicated the findings of triallelic inheritance, but it is possible that triallelism is a rare mechanism of BBS inheritance(227). It is certainly difficult to test for and at present genetic counselling for BBS assumes an autosomal recessive model of inheritance. It

has also been hypothesised that mutational load may affect the phenotypic presentation in BBS and other ciliopathies(176).

3.1.5.3.4 Genetic modifiers in BBS

An alternative explanation for the examples of triallelic inheritance and an explanation for the marked phenotypic heterogeneity seen in BBS and other ciliopathies is the presence of additional variants acting as genetic modifiers(231, 232). Modifiers are variants distinct from the causative gene that alter the phenotype of the condition. Zaki et al. reported two consanguineous families where members of the same family displayed the features of different ciliopathies, in one case BBS and Joubert syndrome(19). The genetic cause was not identified, although linkage analysis was used to exclude *CEP290*, *INPP5E* and *TMEM67* as the cause of the BBS and Joubert presentations. A possible explanation for this is the presence of modifier variants. A good example of a condition where modifiers may be relevant is Joubert syndrome. Joubert syndrome can be caused by mutations in *AHI1*, *NPHP1* and *NPHP6*. However, most people with *NPHP1* mutations will have the much milder condition nephronophthisis. In a study by Tory et al., seven of 13 patients with homozygous or compound heterozygous mutations in *NPHP1* had additional variants in *AHI1* or *NPHP6*(233). Modifiers are difficult to identify for several reasons. Firstly, they can be in genes that are not known to be related to the condition in question or in a non-coding sequence at some distance to the causative gene, their effect may be subtle, they may not segregate with disease in a family, phenotypic variation may be difficult to quantify, multiple variants might have additive or opposite effects and, in rare disease, it may be impossible to generate sufficiently homogeneous cohorts large enough to identify small effects. An additional challenge is the identification of controls, as modifier variants may have no effect in the absence of causative variants and may be seen at relatively high frequency in the normal population.

3.1.5.3.5 Genotype-phenotype correlation in BBS

There is little clear genotype-phenotype correlation in BBS with some exceptions. Patients with variants in *LZTFL1* (BBS17) appear more likely to have mesoaxial polydactyly(234). Two studies by Forsythe et al. appeared to show that patients with missense variants in *BBS1* had lower levels of cardiovascular risk markers than patients with variants in *BBS10* or other types of variants in *BBS1*(190, 200). Patients with variants in *BBS10* also appear to have greater levels of leptin and insulin resistance and adiposity compared to patients with *BBS1* variants while patients with *BBS1* variants may have a less severe visual phenotype(235, 236).

3.1.5.3.6 Genetic diagnosis of BBS

In the UK at present, genetic diagnosis of BBS is undertaken at the North East Thames Regional Genetics Laboratory, where 19 of the 21 known genes are sequenced using next-generation sequencing technology (NGS), though this may change with the introduction of new genetic testing services by NHS England(237). If a diagnosis cannot be made, samples may be submitted to the Genomics England 100,000 Genomes Project (recruiting until September 2018), where whole genome sequencing will be undertaken to attempt to identify a genetic diagnosis.

3.1.6 Aims

This chapter looks at the use of whole genome sequence to make a diagnosis in a pair of monozygotic twins and a singleton patient, all of whom meet the current diagnostic criteria for a clinical diagnosis of Bardet-Biedl syndrome. It looks at whether this strategy was successful and explores the reasons why it might or might not have been. It considers the potential benefits, disadvantages and challenges of introducing diagnostic WGS to the clinical setting.

3.2 Results for patient BBS-018

3.2.1 Next generation sequencing results

In patient BBS-018, WGS was performed as described in Chapter 2, sections 2.1 to 2.3. Data were analysed as described in Chapter 2, section 2.4. Common variants and low confidence variants were excluded and remaining variants filtered. All remaining coding variants in genes causing BBS, other ciliopathies and known ciliary genes identified in patient BBS-018 are listed in Table 3.5. Non-coding variants with possible functional effect in genes causing BBS, other ciliopathies and known ciliary genes are listed in Table 3.6. Variants in other genes of interest are listed in Table 3.7.

3.2.1.1 Variants in known BBS, ciliopathy and ciliary genes in patient BBS-018

The only BBS gene with a coding variant identified was *CEP290.* Variants were identified in several genes on the ciliopathy panel, which are known to cause motile and non-motile ciliopathies. In addition, several variants were identified in known ciliary genes that have never had a morbid phenotype associated with them(153, 238-240). Details can be seen in Tables 3.5 and 3.6.

3.2.1.1.1 Variants in *CEP290*

A heterozygous missense variant, p.Met1723Val, was identified in exon 38 of *CEP290*. This was the only coding variant seen in a known BBS gene, although it appears to be an extremely rare cause of BBS. *CEP290* is involved in ciliary assembly and trafficking. It is implicated in multiple ciliopathies in addition to BBS, including COACH, Meckel, Joubert and Senior-Løken syndromes(142). It is almost ubiquitously expressed, found in all tissues except smooth muscle and placenta. Missense variants in *CEP290* have been reported to cause disease, but in the single reported case of BBS caused by variants in *CEP290*, the patient was homozygous for a premature termination codon mutation and has a heterozygous *TMEM67* variant in addition(176, 241).

Gene	Amino acid change	Protein change	Position	Read depth	Allele fraction %	Panel	Hom/ Het	Disease	Polyphen	SIFT	1000G %	ExAC %	Gnom AD %	NHLBI %	CADD score	ACMG/ HGMD
ALMS1	c.8641A>G	p.Ile2881Val	2:73490597	39	51.3	S	Het	ALM	None	None	0	0	0	0	22	ACMG 3
ALMS1	c.1577_1579 delCTC	p.Pro526del	2:73448098	23	60.5	S	Het	ALM	None	None	0	0	0	0	None	ACMG 2
ALMS1	c.75_77 dupGGA	p.Glu28dup	2:73385909	14	51.3	S	Het	ALM	None	None	0	0	0	0	14.9	ACMG 2
CDHR1	c.713delA	p.Asp238fs*29	10:84203053	31	41.4	CC	Het	RP	None	None	0	0	0	0	None	ACMG 4
CEP290	c.5167A>G	p.Met1723Val	12:88080241	29	36	B	Het	BBS, JS, COACH, MKS, SLS	Possibly damaging	Damaging	0	0	0	0	24.2	ACMG 3
DNAAF5	c.718G>A	p.Val261Ile	7:740819	35	45.7	S	Het	PCD	Possibly damaging	Possibly damaging	0.02	0.041	0.062*	0.046	25.3	ACMG 3
DNAH5	c.2195C>T	p.Ser732Phe	5:13900270	36	56.5	C	Het	PCD	Possibly damaging	None	0	0	0	0	28.4	ACMG 3
DNAH9	c.9542C>T	p.Pro3181Leu	17:11854037	35	48	C	Het	PCD	Probably damaging	None	0	0	0	0	28.8	ACMG 3
DNAH12	c.1369A>G	p.Ile457Val	3:57489654	31	61.2	CC	Het	None	None	None	0.319	0.121	0.133	0	13.5	ACMG 3
DNAH12	c.2311A>G	p.Met771Val	3:57468774	33	45.5	CC	Het	None	None	None	0	0	0.003	0	<10	ACMG 3
NPHP4	c.3364A>C	p.Thr1122Pro	1:5867848	34	35.3	C	Het	NPH, SLS	Possibly damaging	Damaging	0	0.008	0.01	0.007	23.2	HGMD path
PCM1	c.4378G>A	pVal1514Met	8:17991550	35	57.14	S	Het	None	Probably damaging	None	0	0.002	0.003	0	32	ACMG 3
PKHD1	c.143G>A	p.Gly48Asp	6:52082550	23	43.8	C	Het	ARPKD	Benign	Tolerated	0.02	0.002	0	0	11	ACMG 3
PKHD1	c.8110T>A	p.Ser2704Thr	6:51836467	26	50	C	Het	ARPKD	Benign	Tolerated	0	0	0	0	13.4	ACMG 3
PKHD1	c.8521A>G	p.Met2841Val	6:51775841	53	64.2	C	Het	ARPKD	Benign	Tolerated	0.819	0.316*	0.311*	0.915	<10	ACMG 2

ALM- Alström syndrome, ARPKD- autosomal recessive polycystic kidney disease, BBS- Bardet Biedl syndrome, COACH- cerebellar vermis hypo/aplasia, oligophrenia, ataxia, coloboma, hepatic fibrosis, JBTS- Joubert syndrome, MKS- Meckel syndrome, NPH- nephronophthisis, PCD- primary ciliary dyskinesia, RP- retinitis pigmentosa, SLS- Senior-Løken syndrome. Panels: B-BBS, C- Ciliopathy, CC- CiliaCarta, S-Syscilia (see Appendix 2). *= homozygotes seen. CADD- Combined Annotation Dependent Depletion, ExAC- Exome Aggregation Consortium Browser, GnomAD- Genome Aggregation Consortium Browser

Table 3.5 Protein coding variants in BBS, ciliopathy and ciliary genes in patient BBS-018

Gene	Amino acid change	Position	Predicted functional effect	Read depth	Allele fraction %	Panel	Hom/ Het	Disease	Polyphen	SIFT	1000G %	EXAC %	Gnom AD %	NHLBI %	CADD	ACMG/ HGMD
CEP290	c.-641T>C	12:88142513	Promoter, transcription factor binding site	33	36.6	B	Het	BBS, COACH, JS, MKS, SLS	None	None	0.359	0	1.017*	0	<10	ACMG 1
CEP290	c.-1003T>C	12:88142875	Promoter, transcription factor binding site	26	57.7	B	Het	BBS, COACH, JS, MKS, SLS	None	None	0	0	0	0	None	ACMG 3
CCDC39	c.2406+6G>A	3:180616820	Splice site loss	28	53.5	C	Het	PCD	None	None	0.02	0.001	0.002	0	<10	ACMG 3
NPHP4	c.1282-2A>T	1:5875102	Splice site loss	35	57.1	C	Het	NPH, SLS	None	None	0.157	0.164	0.831	0	None	HGMD benign
PCM1	c.3584+8T>C	8:17969756	Intronic	28	35.7	S	Het	No known disease	None	None	0	0.006	0.003	0.008	<10	ACMG 3
TTC8	c.-1004_-998 del ATTATTA	14:88823907	Promoter loss	22	27.3	B	Het	BBS, RP	None	None	0	0	0.043	0	6.4	ACMG 3
WDR19	c.2250-1G>A	4:39253145	Slice site loss	33	48.5	C	Het	NPH	None	None	0	0	0	0	26.5	ACMG 3

ALM- Alström syndrome, ARPKD- autosomal recessive polycystic kidney disease, BBS- Bardet Biedl syndrome, COACH- cerebellar vermis hypo/aplasia, oligophrenia, ataxia, coloboma, hepatic fibrosis, JBTS- Joubert syndrome, MKS- Meckel syndrome, NPH- nephronophthisis, PCD- primary ciliary dyskinesia, RP- retinitis pigmentosa, SLS- Senior-Løken syndrome. Panels: B-BBS, C- Ciliopathy, CC- CiliaCarta, S-Syscilia (see Appendix 2). TFBS- transcription factor binding site. *= homozygotes seen. CADD- Combined Annotation Dependent Depletion, ExAC- Exome Aggregation Consortium Browser, GnomAD- Genome Aggregation Consortium Browser

Table 3.6 Non-coding variants at promoter, splice or transcription factor binding sites in BBS, ciliopathy and ciliary genes in patient BBS-018

Gene	Amino acid change	Protein change	Position	Read depth	Allele fraction %	Panel	Hom/ Het	Disease	Polyphen	SIFT	1000G %	EXAC %	Gnom AD %	NHLBI %	CADD	Notes
LAMB1	c.2915C>T	p.Ser972Leu	7:107953694	34	55.6	OMIM	Het	LIS	Benign	Tolerated	0	0.005	0.003	0	10.3	ACMG 3
LAMB1	c.3470T>C	p.Val1157Ala	7:107940280	46	56.5	OMIM	Het	LIS	Benign	Tolerated	0	0.002	0.001	0	23.0	ACMG 3
TCOF1	c.814A>G	p.Ser272Gly	5:150372180	31	61.3	OMIM	Het	TCS	Benign	Tolerated	0	0	0	0	24.1	ACMG 3
TCOF1	c.742_768 del	p.Lys322_Thr 330del	5:150374627	27	59.3	OMIM	Het	TCS	None	None	0	0	0	0	None	ACMG 3
TCOF1	c.2684A>T	p.Glu956Val	5:150383745	35	51.4	OMIM	Het	TCS	Benign	Tolerated	0	0	0.001	0	<10	ACMG 3
TCOF1	c.3010G>A	p.Ala1004Thr	5:15039160	30	50	OMIM	Het	TCS	Benign	Damaging	0	0.012	0.01	0.069	23.2	ACMG 3
TTC37	c.536A>G	p.Gln179Arg	5:95540697	38	52	OMIM	Het	THE	Benign	Damaging	0	0	0	0	13.52	ACMG 3
TTC37	c.623C>G	p.Phe208Cys	5:95537062	27	40.7	OMIM	Het	THE	Possibly damaging	Damaging	0	0.003	0.003	0	26.7	ACMG 3

Diseases: LIS- Lissencephaly, TCF- Treacher Collins syndrome, THE- Tricho-hepato-enteric syndrome. *= homozygotes seen. CADD- Combined Annotation Dependent Depletion, ExAC- Exome Aggregation Consortium Browser, GnomAD- Genome Aggregation Consortium Browser

Table 3.7 Coding variants in OMIM morbid genes seen in patient BBS-018

Residue 1723 is very highly conserved and situated in a coiled-coil domain. The other disease-causing variants reported in this area are frameshifts(217). Both methionine and valine are very hydrophobic amino acids and are similar in structure. The Grantham distance, a measure of physicochemical distance between amino acids, is 21. However methionine to valine variants have been reported as disease-causing. Indeed, a valine to methionine substitution at position 339 in the *BBS1* gene has been reported as pathogenic(18). It is not a null variant, an amino acid change in the same position has not been reported and due to the absence of a paternal sample, inheritance of the variant cannot be determined and so the highest level it can be classified as is PM2, as it is absent from controls. This is considered moderate evidence of pathogenicity. It is also predicted to be deleterious in multiple lines of computational evidence (PP3- supportive) but as no additional strong, moderate or supportive criteria can be confirmed at this time it does not meet the criteria for classification as a pathogenic or likely pathogenic variant and is considered a class three VUS under ACMG guidelines. It also has a CADD score of 24.2, suggesting it is in the top 1-0.1% of deleterious variants(242).

In addition, *CEP290* was the only known BBS gene in which a second rare variant, albeit non-coding, could be identified. Two promoter variants at ENCODE transcription factor binding sites were seen. The first of these was the heterozygous variant, c.-641T>C. However this was seen at a frequency of 0.359% in 1000Genomes and 1.017% in GnomAD, including in three homozygotes, two ACMG strong criteria for classifying variants as benign, making it class one. This variant was too common to cause BBS, as it is commoner than the most frequently seen BBS-causing variant, p.Met390Arg in *BBS1*. The second was the heterozygous variant c.-1003T>C. This variant was very rare, being absent from the population databases. No promoter variants are found in the *CEP290* database, but intronic variants have been reported(241). As with the other variant, this meets a single moderate evidence criterion (absent in controls) so must be classified as a VUS.

There were several additional rare intronic variants in *CEP290* which were not sited at splice sites, promotors or possible transcription factor binding sites. These had no predicted functional effect. It is known that intronic variants can be pathogenic(243). None were in any population database and none have previously been reported as pathogenic. Therefore, while it was impossible to rule these out as pathogenic, there was no evidence to suggest pursuing them at present. There was no evidence of a small CNV in *CEP290* that would constitute a pathogenic variant.

Sanger sequencing of the proband revealed that both variants were present (section 3.2.2). Maternal sequencing showed neither. Paternal samples were unavailable and so the inheritance of the variants could not be determined. It may be that the variants are in *cis* and paternally inherited or in *trans* with one variant occurring *de novo* or present in the mother as a mosaic.

3.2.1.1.2 Variants in *NPHP4*

A heterozygous missense coding variant and a heterozygous non-coding variant were identified in *NPHP4*. *NPHP4* is a gene known to cause Senior-Løken syndrome and nephronophthisis, is localised to the ciliary transition zone and may mediate attachment of the basal body to the

transition zone membrane(244). The major features of Senior-Løken syndrome are nephronophthisis (medullary cystic kidney disease) and retinal dystrophy. Features such as obesity, polydactyly and learning difficulties have not been reported, but genes causing Senior-Løken syndrome (*CEP290, SDCCAG8*) have been reported to cause BBS. The coding variant identified, p.Thr1122Pro, has previously been reported as a pathogenic variant causing nephronophthisis and is listed in HGMD (CM110621)(245, 246).The single patient reported by Otto et al. had a second frameshift variant and had no extra-renal features.

The second *NPHP4* variant identified in BBS-018, is a splice site variant, c.1282-2A>T, that affects the invariant acceptor splice site of intron 20. It is listed in dbSNP (rs1287637). It is also listed in HGMD, where it is described as a functional polymorphism(247, 248). It was found to be associated with a reduced glomerular filtration rate (eGFR), although the results were only borderline significant (p=0.054). There is also a single submission in ClinVar (RCV00153587.2) where it is listed as benign. Its frequency in ExAC is 0.1638% and 0.1567% in 1000Genomes making it an unlikely candidate for a pathogenic variant in BBS. There were no other rare coding or non-coding variants and no evidence of a small CNV in *NPHP4*.

3.2.1.1.3 Variants in *CDHR1*

A rare heterozygous frameshift variant leading to a premature termination codon, p.Asp238fs29, was identified in *CDHR1*. *CDHR1* is mainly expressed in the retina and is necessary for photoreceptor survival(249). Deletions in this gene are known to cause isolated rod-cone dystrophy but as yet, it has not been associated with additional features such as polydactyly or renal cystic disease and has never been identified as a cause of BBS(250). Under ACMG guidelines this variant would be classified as likely pathogenic (one very strong and one moderate piece of evidence). However, no additional rare coding or likely functional non-coding variants or small CNVs were identified, so this was not pursued as a causative variant.

3.2.1.1.4 Variants in *WDR19*

A rare heterozygous splice site variant, c.250-1A>G was identified in *WDR19* in BBS-018. Under ACMG guidelines, this is classified as a variant of unknown significance. *WDR19* is known to be associated with multiple ciliopathies including nephronophthisis, Senior-Løken syndrome, short-rib thoracic dysplasia with or without polydactyly and cranioectodermal dysplasia. A homozygous missense variant was also identified in a patient with Jeune syndrome and was thought to be causative(251). The reported patient in this case required a renal transplant. In addition, Halbritter et al. identified a patient with nephronophthisis, retinal dystrophy and hepatic cysts(252). The reported patient was compound heterozygous for a missense and a nonsense variant. Splice site variants have been identified in patients with Senior-Løken syndrome caused by *WDR19* variants(252, 253). However no second coding or non-coding variant could be identified, nor could any small CNVs be found, and so while this is an interesting candidate gene it was not pursued.

3.2.1.1.5 Variants in *ALMS1*

Three rare heterozygous variants were observed in *ALMS1* in BBS-018. One, a heterozygous in frame deletion p.Pro526del, is considered to be a benign polymorphism in ClinVar

(RCV00206500.2), where it was assessed as benign or likely benign by 3 submitters(254). The second, a heterozygous in-frame duplication, p.Glu28dup is also listed in ClinVar (RCV00206500.2) as a benign or likely benign variant. As both are in-frame variants and have been classified as benign, they are classified as ACMG class two or likely benign variants. The third is a rare heterozygous missense variant, p.Ile1288Val, which was not seen in any of the population databases. *ALMS1* is known to be associated with Alström syndrome, whose main features are retinitis pigmentosa, deafness, obesity and diabetes mellitus, similar to BBS. Knockdown of Alms1 resulted in defective ciliogenesis(255). However, polydactyly has not been observed in Alström syndrome, and learning difficulties tend to be mild. Missense variants have never been reported as a definite cause of Alström syndrome, with only nonsense and frameshift variants known to be causative. Such missense variants as have been identified in patients are considered variants of unknown significance at present and this variant is classified as a VUS under ACMG guidelines(256). *ALMS1* is also present on diagnostic ciliopathy panels and variants have not yet been identified as a cause for BBS. As no additional coding or likely functional non-coding variants or small CNVs were identified, variants in *ALMS1* were not pursued as potentially causative in this case, although they cannot be entirely ruled out.

3.2.1.1.6 Variants in *PKHD1*

Three heterozygous coding variants were seen in *PKHD1.* The first, p.Gly48Asp, is rare, as is the second, p.Ser2704Thr. Both are ACMG class three VUSs. The third, p.Met2841Val, was seen at a frequency of 0.819% in 1000Genomes, with 3 homozygotes recorded in ExAC and 12 in gnomAD. The variant is in dbSNP (rs113562492). It is in ClinVar (RCV000169052.3) where it is described by two submitters as benign or likely benign. It is also listed in HGMD (CM052345) where it is described as of uncertain significance(257, 258). This gives it an ACMG classification of class two, likely benign. *PKHD1* is known to be associated with autosomal recessive polycystic kidney disease (ARPKD) and also with hepatic cysts and intracranial aneurysms and generally manifests prenatally or in infancy(259). The product of *PKHD1* interacts with the product of *PKD2* and together they are involved in cell-cell signalling in the kidney. *PKHD1* is an unlikely cause of BBS as it has never been associated with polydactyly, learning disability or rod-cone dystrophy. It is also included on diagnostic ciliopathy panels such as the one offered by the North East Thames Regional Genetics Service and has never been implicated in BBS or any ciliopathy other than ARPKD. Therefore it was considered an unlikely causative gene candidate.

3.2.1.1.7 Variants in primary ciliary dyskinesia genes- *DNAAF5, DNAH5, DNAH9, CCDC39*

Single heterozygous missense variants were identified in *DNAAF5, DNAH5* and *DNAH9*. The variant in *DNAAF5* was p.Val261Ile, which was seen in a heterozygotes in the GnomAD database. However *in silico* predictions suggest that it is damaging, resulting in an ACMG class three classification. The variant in *DNAH5* was p.Ser732Phe, like the p.Pro3181Leu variant in *DNAH9*, was not seen in population databases. Both variants were predicted to be possibly or probably damaging by *in silico* tools. Both are classified as class three under ACMG guidelines.

A heterozygous variant predicted to result in splice site loss, c.2406+6G>A was identified in *CCDC39.* According to Ingenuity Variant Analysis™ (IVA) software, this is likely to result in splicing at an alternative splice site on the 3' side of the intron/exon boundary. This is predicted to result in an insertion leading to a frameshift. A frameshift is considered very strong evidence of pathogenicity, but other than this there is only 1 moderate evidence criterion, low frequency in controls, resulting in an ACMG classification of three. All of these genes, *DNAAF5, DNAH5, DNAH9 and CCDC39,* are causative genes for primary ciliary dyskinesia (PCD) and have functions in the formation and motility of the dynein arms of motile cilia. Genes causing PCD have not been reported to cause non-motile ciliopathies, and although these variants were rare, no additional coding or likely functional non-coding variants or small CNVs could be identified, so they were not considered further.

3.2.1.1.8 Variants in *PCM1*

A single heterozygous missense variant was identified, p.Val1514Met, at a highly conserved nucleotide. It is predicted to be damaging by PolyPhen2 and has a high CADD score (32-predictive of the most damaging 0.01 to 0.1% of variants) consistent with pathogenicity. It is rare, with a frequency of 0.002% in ExAC. It is classified as class three under ACMG guidelines. When non-coding variants were considered, an intronic variant, c.3584+8T>C variant was identified. The residue is conserved but although it is close to an intron-exon boundary it was not predicted to be part of a splice site. It is also classified as class three under ACMG guidelines.

PCM1 is a ciliary protein which is known to interact with *CEP290*. It is a component of centriolar satellites(260). Both *PCM1* and *CEP290* are required for ciliogenesis and localisation of *CEP290* to centriolar satellites has been found to be *PCM1* dependent, with ciliary formation being affected by impaired retrograde trafficking of PCM1 in mice(151, 261, 262). *PCM1* has never been implicated in a ciliopathy, but is a possible candidate gene in view of its interactions with known BBS-causing proteins. It is possible that homozygous or compound heterozygous variants affecting *PCM1* are lethal and this may be the reason they have not been observed. STRING suggests associations with *BBIP1, BBS1, BBS2, BBS4, MKS1, SDCCAGA1* and *TTC8* in addition to C*EP290* (Figures 3.4 and 3.5). No second coding variant could be identified and no small CNVs were found. For this reason, *PCM1* was not thought to be a primary candidate gene, but it is possible that it could act as a modifier of other disease-causing mutations (section 3.4.1.6).

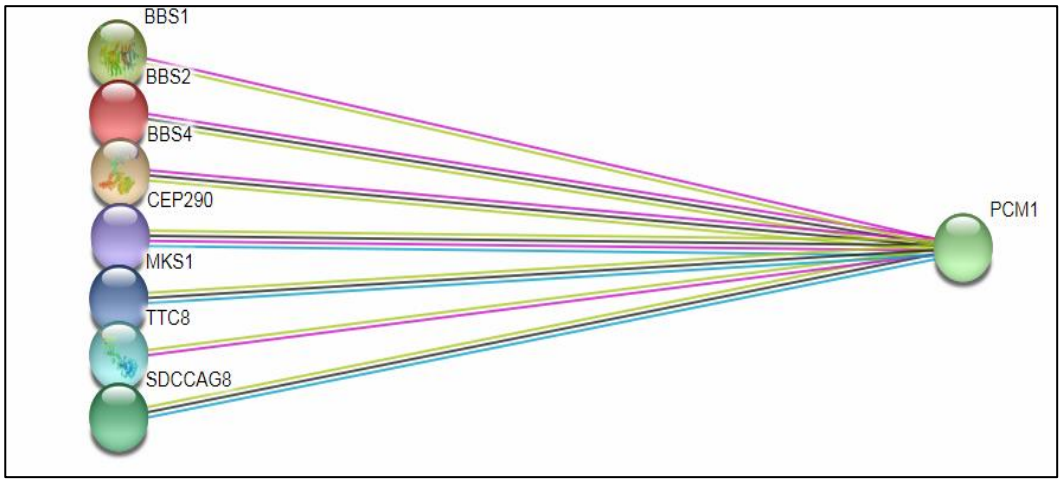

Figure 3.4 Associations between PCM1 and known BBS genes. Drawn by www.string-db.org. Pink lines represent experimental evidence of connection, black represent co-expression data, blue represent curated databases and green represent text mining data

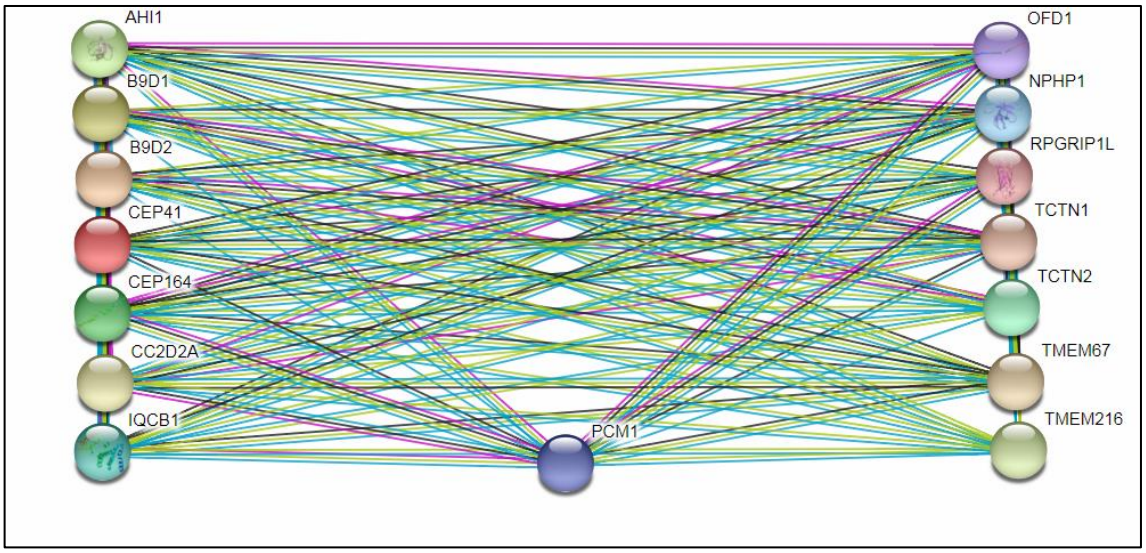

Figure 3.5 Associations between PCM1 and known ciliopathy genes. Drawn by www.string-db.org. Pink lines represent experimental evidence of connection, black represent co-expression data, blue represent curated databases and green represent text mining data

3.2.1.1.9 Variants in *TTC8*

A single, non-coding, heterozygous variant was identified in *TTC8*, another gene reported to cause BBS, although only splice site variants and deletions have been implicated(263, 264). This variant, c.-1004_-998delATTATTA, thought to be in the promoter region, is a seven base pair deletion. While there are no *in silico* predictions, IVA software predicts that it will result in promoter loss. It is considered a class three VUS under ACMG guidelines. *TTC8* is required for ciliogenesis and interacts with *BBS4* and *PCM1*. This variant was seen at a frequency of 0.043% in gnomAD, so is relatively rare. However, no other rare coding or non-coding variants or deletions could be identified in this gene. It was excluded therefore from further consideration.

3.2.1.1.10 Variants in *DNAH12*

Two heterozygous coding variants were seen in *DNAH12.* The first, p.Ile457Val, was seen at relatively high frequencies in population databases. The second, p.Met771Val is rare. Neither had a high CADD score and no predictions were made by PolyPhen2 or SIFT so computational evidence did not point towards pathogenicity. Although both variants were at conserved nucleotides, they are classified as class three under ACMG guidelines. No additional variants or CNVs were identified. *DNAH12* has been identified as a gene coding for a ciliary protein and has homology to axonemal dyneins but has never been identified as causing a ciliopathy(265). Other axonemal dynein heavy chain genes, such as *DNAH1, DNAH5* and *DNAH11* are known to cause PCD. Analysis using the STRING database found no associations between known BBS genes and *DNAH12*, but did identify associations with *DNAH5*, which is known to cause PCD and *DYNC2H1*, which causes short rib thoracic dysplasia with or without polydactyly(86). While *DNAH12* is a possible candidate gene, it is was not pursued as a primary candidate.

3.2.1.2 Variants in non-ciliopathy disease genes in patient BBS-018

3.2.1.2.1 Variants in *LAMB1*

Two heterozygous missense variants were seen in patient BBS-018, p.Ser972Leu and p.Val1157Ala. Both were rare variants not seen in homozygotes in ExAC or GnomAD, and were predicted to be benign by computational methods, although the CADD score of p.Val1157Ala was 23. This residue was highly conserved. The other, p.Ser972Leu, was less highly conserved and in fact, leucine is seen in this position in the rhesus monkey, suggesting that this variant is well tolerated. Both variants would be classified as class three VUS under ACMG guidelines.

LAMB1 codes for the protein Laminin Beta-1, which along with alpha and gamma subunits forms Laminin, an extracellular matrix protein. Both loss of function and missense variants in *LAMB1* have been found to cause lissencephaly and variants in *LAMB2*, a homologue, cause Pierson syndrome and nephrotic syndrome with structural eye abnormalities(266-268). Reported patients have presented with developmental delay, neurocognitive impairment, seizures and encephaloceles, but none have had features typical of BBS. Interestingly, a study by Hochgreb-Haeglele et al. found that zebrafish with variants in lamb1a, the homologous gene, had defects in gut organ laterality and had reduced cilia length in Kuppfer's vesicle(269). Review of the patient's

MRI scan revealed no lissencephaly. There was not enough evidence to pursue *LAMB1* as a candidate gene, but in view of the patient's situs inversus it remains interesting.

3.2.1.2.2 Variants in *TCOF1*

A total of 4 coding variants were seen in *TCOF1.* These were three rare heterozygous missense variants, p.Ser272Gly, p.Glu956Val and p.Ala1004Thr, and p.Lys322_Thr330del, an in-frame 9 amino acid deletion. All would be classified as VUS under ACMG criteria. *TCOF1* is thought to be involved in the transcription of ribosomal DNA, and is vital for craniofacial development in the embryo. Variants in *TCOF1* cause Treacher Collins syndrome, a rare, autosomal dominant craniofacial malformation syndrome whose chief features are facial bone hypoplasia, structural ear and eye abnormalities and deafness(270). Most variants associated with Treacher Collins syndrome generally result in a truncated protein but rarely, causative missense variants have been identified(271-273). None of the variants seen would cause a truncated protein, although the in-frame deletion variant would lead to a slightly shortened protein. *TCOF1* has never been implicated in a ciliopathy and analysis with STRING showed no association with any known ciliopathy gene. The patient's clinical features do not fit with those of a *TCOF1*-related phenotype so these variants were excluded from further analysis.

3.2.1.2.3 Variants in *TTC37*

Two heterozygote missense variants were seen in *TTC37*, p.Gln179Arg and p.Phe208Cys. Both were rare and would be classified as class three VUSs under ACMG guidelines. *TTC37* is a widely expressed protein of unknown function. Variants are known to cause trichohepatoenteric syndrome (THE) whose main features include a neonatal enteropathy resulting in intractable diarrhoea, abnormal brittle hair, short stature, liver disease and café-au-lait patches. Other features such as immunodeficiency and learning difficulties are sometimes seen. The condition is rare and causative variants are usually premature termination codons or frameshifts. Analysis with STRING showed no link to a known ciliopathy gene. While patient BBS-018 has a diagnosis of Crohn's disease, this was not of neonatal or very early onset, and there is no hair abnormality. Both missense variants are at moderately conserved residues, but the phenotype inconsistent with the patient's and so they were not considered further.

3.2.2 Sanger sequencing

3.2.2.1 Sanger sequencing of *CEP290* in the proband, BBS-018 and parent

Sanger sequencing confirmed the presence of both heterozygous variants, c.5167A>G, p.Met1723Val and c.-1003T>C in the proband (Figures 3.6 and 3.7). Neither was seen in a control. DNA could not be obtained from the proband's father. Sanger sequencing of the mother showed that she carried neither variant (Figures 3.6 and 3.7). Therefore biallelic inheritance could not be confirmed.

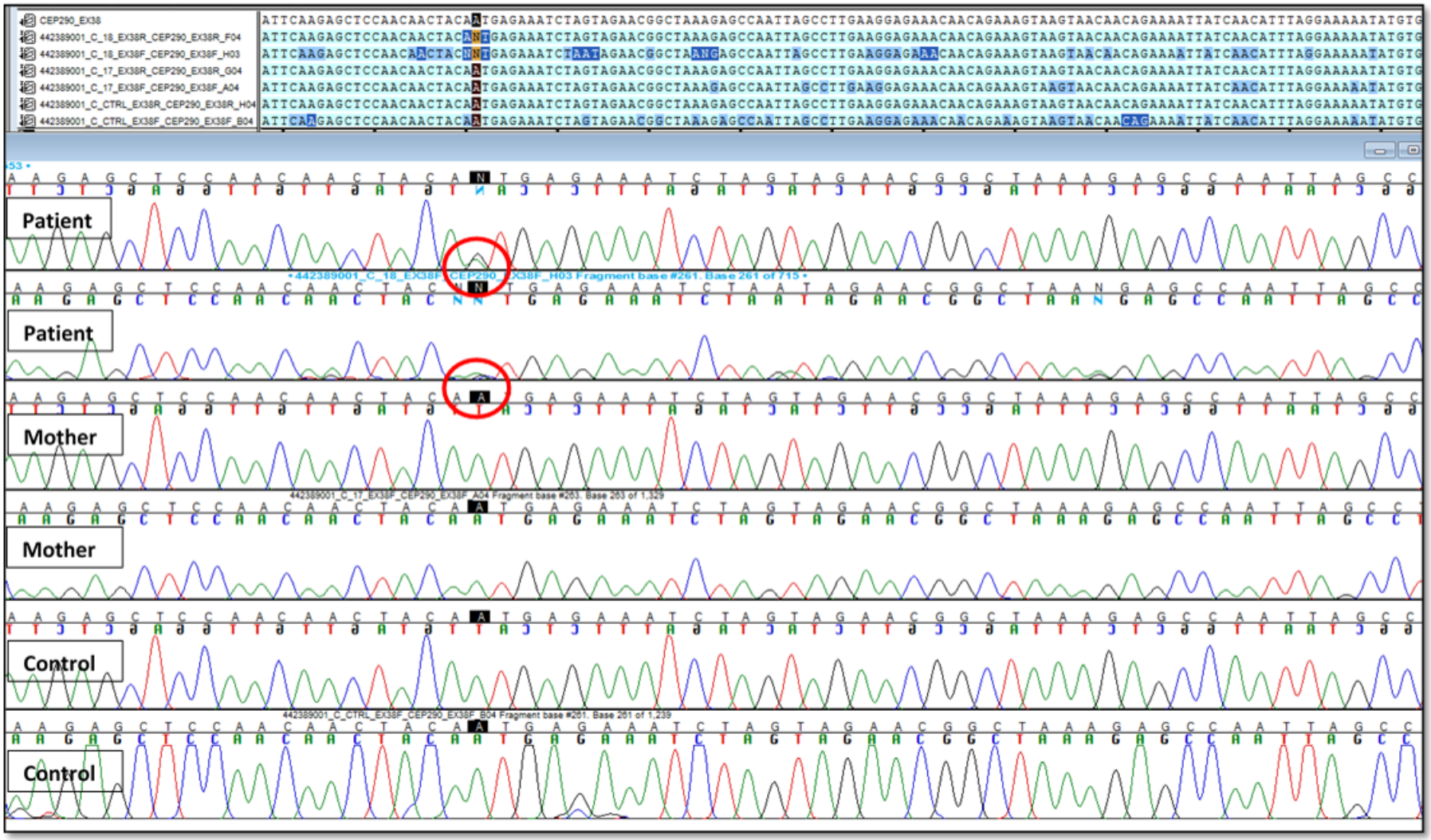

Figure 3.6 Sanger sequencing of c.5167A>G, p.Met1723Val in BBS-018 (top two lines), mother of BBS-018 (middle two lines) and healthy control (bottom)

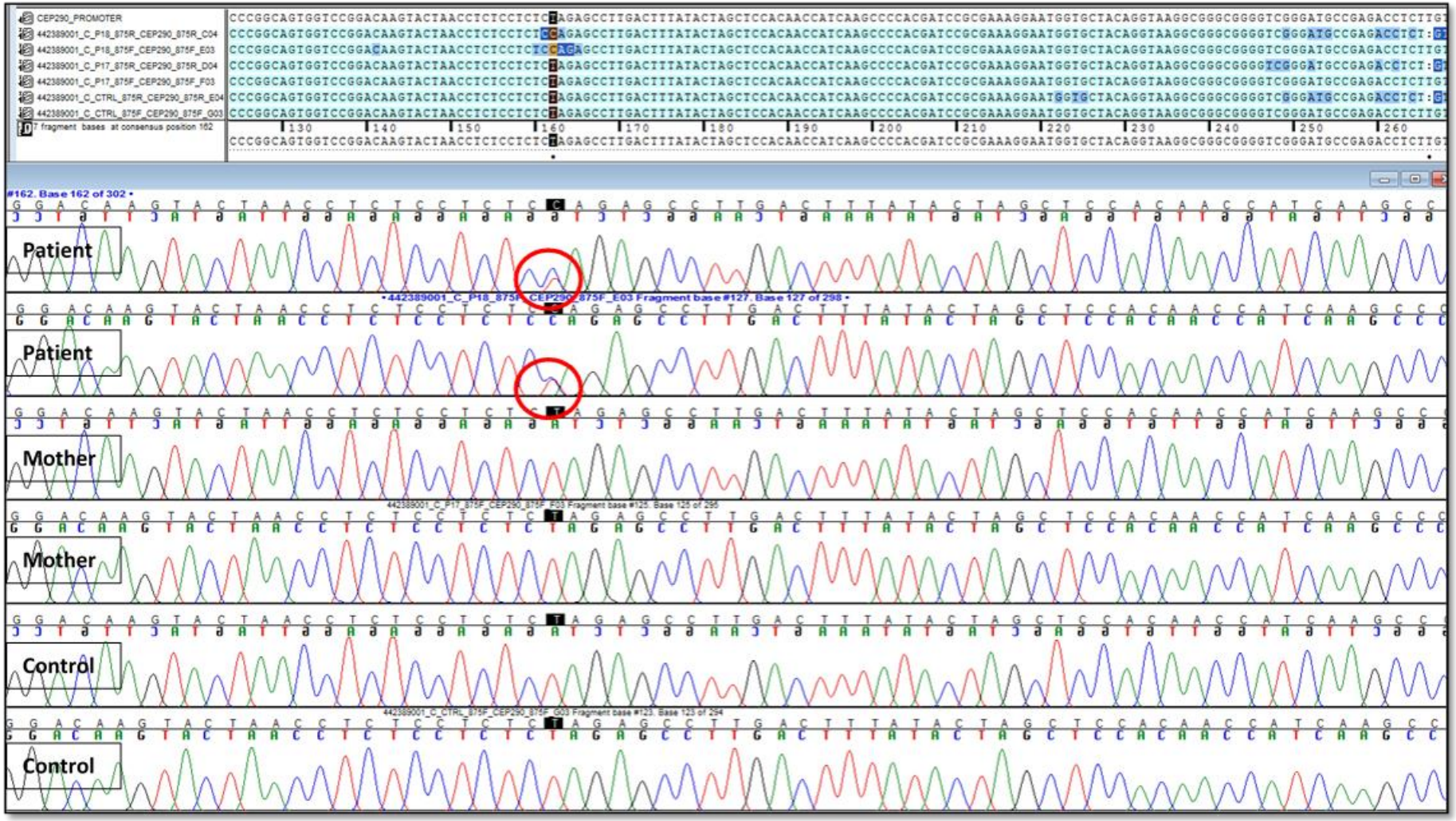

Figure 3.7 Sanger sequencing of promoter variant c.-1003T>C in BBS-018 (top two lines), mother of BBS-018 (middle two lines) and healthy control (bottom two lines)

3.3 Results for patient BBS-016 and BBS-017

3.3.1 Next generation sequencing results

In patients BBS-016 and BBS-017, WGS was performed as described in Chapter 2, sections 2.1 to 2.3. Data were analysed as described in Chapter 2, section 2.4. Common and low-confidence variants were excluded and remaining variants filtered. As patients BBS-016 and BBS-017 are monozygotic twins, all discussed variants were seen in both patients. As patients were male, hemizygous variants in X chromosome genes were considered. All remaining coding variants in genes causing ciliopathies and known ciliary genes in these patients are listed in Table 3.8. Non-coding variants with possible functional effect in genes causing BBS, other ciliopathies and known ciliary genes are listed in Table 3.9. Variants in other genes of interest are listed in Table 3.10.

3.3.1.1 Variants in known BBS, ciliopathy and ciliary genes in BBS-016 and BBS-017

3.3.1.1.1 Variants in *ABCA4*

Two heterozygous coding variants were identified in *ABCA4*, the only non-motile ciliopathy gene for which this was the case. The first heterozygous variant, p.Arg2040Ter, is rare and has a CADD score of 50. This variant has previously been reported as pathogenic in two patients with Stargardt disease and results in a premature termination codon(274). The mutation type, frequency and the fact that it has been reported as pathogenic put it in ACMG class five. It is listed in HGMD (CM032170). Variants in *ABCA4* are known to cause recessive RP and Stargardt disease, as well as non-juvenile macular degeneration. Most causative variants appear to be missense, though null variants have been reported as pathogenic previously, including two that flank the variant in question, one at position 2030 and one at position 2565(275, 276). However, the study by Baum et al. only looked for variants in 15 of 51 exons, and did not report whether a second variant was found in either patient, nor did it give any clinical details apart from the diagnosis.

The second heterozygous variant is p.Asn1868Ile, which has also been reported as pathogenic, although it has a relatively high population frequency with an allele frequency of 2-4% in the databases considered(277-280). Zernant et al. describe it as causing late-onset macular degeneration. It is seen five times more frequently in those patients with one other disease-causing variant with disease compared to controls, and has been seen in the homozygous state in patients. It is listed in HGMD (CM015091). One strong and 2 supporting criteria would result in an ACMG classification of class four, likely pathogenic, but in the context of BBS, the population frequency is too high, reducing it to an ACMG class three variant. No additional sequence or copy number variants could be identified.

Gene	Amino acid change	Protein change	Position	Read depth	Allele fraction %	Panel	Hom/ Het	Disease	Polyphen	SIFT	1000G %	EXAC %	Gnom AD %	NHLBI %	CADD	ACMG/ HGMD
ABCA4	c.5603A>T	p.Asn1868Ile	1:94010911	36	49.5	CC	Het	RP, Stargardt	Possibly damaging	Damaging	2.077	4.456	4.206	4.775	26.2	HGMD path (ACMG 3)
ABCA4	c.6118C>T	p.Arg2040Ter	1:94005470	31	50.7	CC	Het	RP, Stargardt	None	None	0	0.002	0.001	0	50	HGMD Path ACMG 5
ATP8A2	c.183C>A	p.Asn61Lys	13:25469083	31	49.5	CC	Het	CAMRQ	None	Damaging	0	0.001	0.001	0.006	33	ACMG 3
CEP164	c.41574C>T	p.Arg1392Trp	11:117411805	36	47.1	C	Het	NPH, (JBTS, MKS)	Probably damaging	Damaging	0	0.001	0.002	0	35	ACMG 3
DNAH14	c.2621A>G	p.Gln874Arg	1:225079403	30	57	CC	Het	None	Benign	Tolerated	0	0	0	0	<10	ACMG 3
HYDIN	c.2144C>T	p.Pro715Leu	16:71064772	39	40.6	S	Het	PCD	Probably damaging	None	0	0.002	0.002	0	28.1	ACMG 3
INPP5E	c.1132C>A	p.Arg378Ser	9:136433182	26	44.52	S	Het	JBTS, MORM	Probably damaging	Damaging	0	0	0	0	27	ACMG 3
NME7	c.1106-1109dup ACTT	p.Phe370fs*8	1:169132806	37	49.6	S	Het	None	None	None	0	0.036	0.031	0.016	35	ACMG 3
PKHD1	c.9925A>G	p.Ile3309Val	6:51746794	27	47.5	C	Het	ARPKD	Possibly damaging	Damaging	0	0.011*	0.02*	0.023	15	ACMG 3
PKHD1	c.9866G>T	p.Ser3289Ile	6:51746794	25	58.7	C	Het	ARPKD	Probably damaging	Damaging	0	0.331*	0.316*	0.315	25.3	HGMD path (ACMG 3)

ARPKD- autosomal recessive polycystic kidney disease, CAMRQ- cerebellar ataxia, mental retardation, and disequilibrium syndrome, JBTS- Joubert syndrome, MORM- mental retardation, truncal obesity, retinal dystrophy, and micropenis, MKS- Meckel syndrome, NPH- nephronophthisis, PCD- primary ciliary dyskinesia, RP- retinitis pigmentosa. Panels: B-BBS, C- Ciliopathy, CC- CiliaCarta, S-Syscilia (see Appendix 2). *= homozygotes seen. CADD- Combined Annotation Dependent Depletion, ExAC- Exome Aggregation Consortium Browser, GnomAD- Genome Aggregation Consortium Browser

Table 3.8 Coding variants in ciliopathy and ciliary genes in patients BBS-016 and BBS-017

Gene	Amino acid change	Position	Predicted functional effect	Read depth	Allele fraction %	Panel	Hom/ Het	Disease	Polyphen	SIFT	1000G %	EXAC %	Gnom AD %	NHLBI %	CADD	Notes
ARL13B	c.-109+18 G>C	3:93980500	None	26	50.4	C	Het	JBTS	None	None	0	0	0	0	<10	ACMG 3
DNAH14	c.5146-10 T>C	1:225153740	Splice site loss	32	47.2	CC	Het	None	None	None	0.08	0.212	0.262	0	15.7	ACMG 3
Diseases: JBTS- Joubert syndrome. Panels: B-BBS, C- Ciliopathy, CC- CiliaCarta, S-Syscilia (see Appendix 2). *= homozygotes seen. CADD- Combined Annotation Dependent Depletion, ExAC- Exome Aggregation Consortium Browser, GnomAD- Genome Aggregation Consortium Browser																

Table 3.9 Non-coding variants at promoter, splice or transcription factor binding sites in BBS, ciliopathy and ciliary genes in patients BBS-016 and BBS-017

Gene	Amino acid change	Protein change	Position	Read depth	Allele fraction %	Panel	Hom/ Het/ Hemi	Disease	Polyphen	SIFT	1000G %	EXAC %	Gnom AD %	NHLBI %	CADD	Notes
ARID1B	c.2846G>T	p.Gly949Val	3:93980500	31	52.6	OMIM	Het	Coffin-Siris	Probably damaging	Tolerated	0	0	0	0	25.6	ACMG 3
CLDN34	c.140delC	p.Pro47fs*48	X:9967492	15	100	OMIM	Hemi	None	None	None	0	0	0	0	13.9	ACMG 3
FRMPD4	c.2243A>G	p.Arg748Gly	X:12716702	16	100	OMIM	Hemi	XLMR	Damaging	Possibly damaging	0	0.009	0.009	0.009	22.1	ACMG 3
GDF5	c.826G>T	p.Arg276Ser	20:35434589	19	51.5	OMIM	Het	BDC GC	None	Tolerated	0	0	0	0	35	ACMG 3
TRIP12	c.4883G>A	p.Arg1628Gln	2:229785842	42	51.5	OMIM	Het	NSMR	Probably damaging	Damaging	0	0	0	0	35	HGMD path
Diseases: BDC- Brachydactyly type C, GC- Grebe chondrodysplasia, NSMR- Non-syndromic mental retardation, XLMR- X-linked mental retardation. *= homozygotes seen. CADD- Combined Annotation Dependent Depletion, ExAC- Exome Aggregation Consortium Browser, GnomAD- Genome Aggregation Consortium Browser																

Table 3.10 Coding variants in OMIM morbid genes seen in patients BBS-016 and BBS-017

Neither of these variants has been associated with a retinitis pigmentosa phenotype, nor have any variants in *ABCA4* ever been described as causing a BBS-like phenotype. It is therefore unlikely that these variants are relevant to the diagnosis, though they do have implications in that patients BBS-016 and BBS-017, while having no ophthalmological problems at present, may have a significantly increased risk of macular degeneration in future. In view of the current clinical diagnosis, they have already been warned of the risk of developing retinitis pigmentosa and are having annual ophthalmology follow-up as a result.

3.3.1.1.2 Variants in *INPP5E*

A single heterozygous variant, p.Arg378Ser, was identified in *INPP5E*, which is known to cause the rare ciliopathy MORM (mental retardation, truncal obesity, retinal dystrophy, and micropenis) syndrome, but also the somewhat less rare Joubert syndrome(281, 282). It is known that genes such as *MKS1* and *CEP290* cause both BBS and Joubert syndrome (JS), so *INPP5E* is a good candidate. *INPP5E* is targeted to the cilium and interacts with other ciliary proteins including *CEP164* and *ARL13B*(283). The variant in the only family reported to have MORM syndrome was a homozygous premature termination variant, while missense variants have been reported in JS.

The variant in *INPP5E* in patients BBS-016 and BBS-017 is very rare and the amino acid is highly conserved. The Grantham distance between arginine and serine is 110. Arginine has a basic side chain and is hydrophilic while serine has a neutral side chain and is a smaller amino acid. The variant is in the inositol polyphosphate phosphatase domain, and a pathogenic variant, where the arginine is converted to a cysteine, has been reported at the same position(284). Under ACMG guidelines, this is classified as a VUS, although there are two moderate and one supporting pieces of evidence for pathogenicity (different pathogenic variant at same site, low population frequency and supportive computational evidence). While it could be argued that *INPP5E* has a low rate of benign missense variants, as according to ExAC, fewer missense variants are seen than are expected (194 instead of 230 (84%), z score 1.17), this rate does not reach statistical significance according to the Association for Clinical Genomic Science (ACGS) best practice guidelines for variant interpretation (www.acgs.uk.com).

No second coding variant or non-coding variant with likely functional effect could be identified. Lumpy data did not highlight any likely CNVs. When reviewing the data manually, a reduction in coverage from greater than 30 to less than 20 was seen in the in the latter part of exon 7 in patient BBS-016, suggesting a possible partial exon deletion. As it is at the end of an exon, this raises the possibility of exon skipping or other mechanisms of disease causation. The drop in coverage in patient BBS-017 was less marked (Figure 3.8). When the intron next to exon 7 was viewed, it appeared that there might be a larger deletion but in fact it was more likely to be a region of poor coverage (Figure 3.9). When other samples in the cohort were reviewed, a similar if less marked drop in coverage was seen, making it very unlikely to be pathogenic. *INPP5E* remains an interesting candidate.

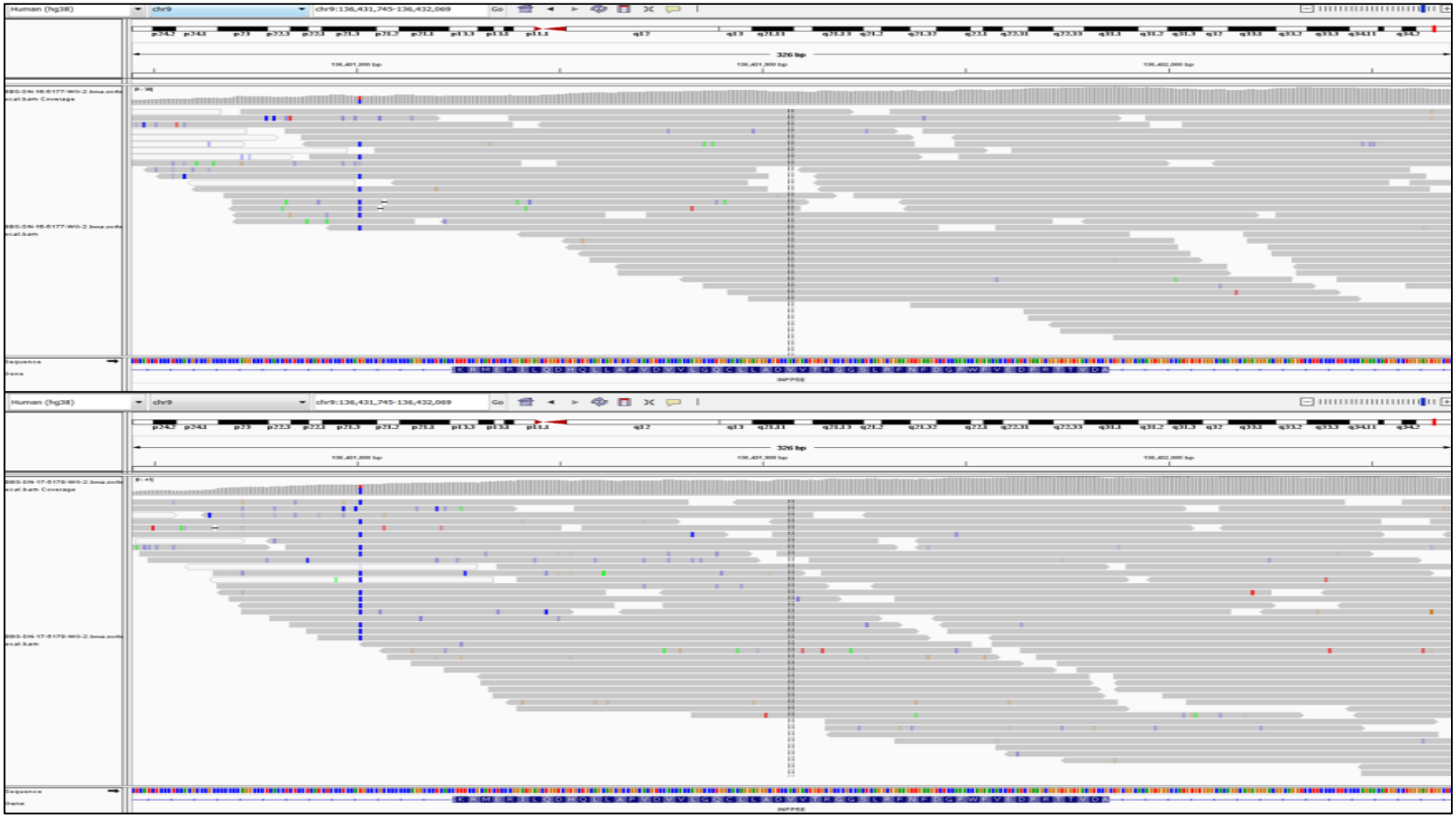

Figure 3.8 Coverage of exon 7 of INPP5E in patients BBS-016 (top) and BBS-017 (bottom)

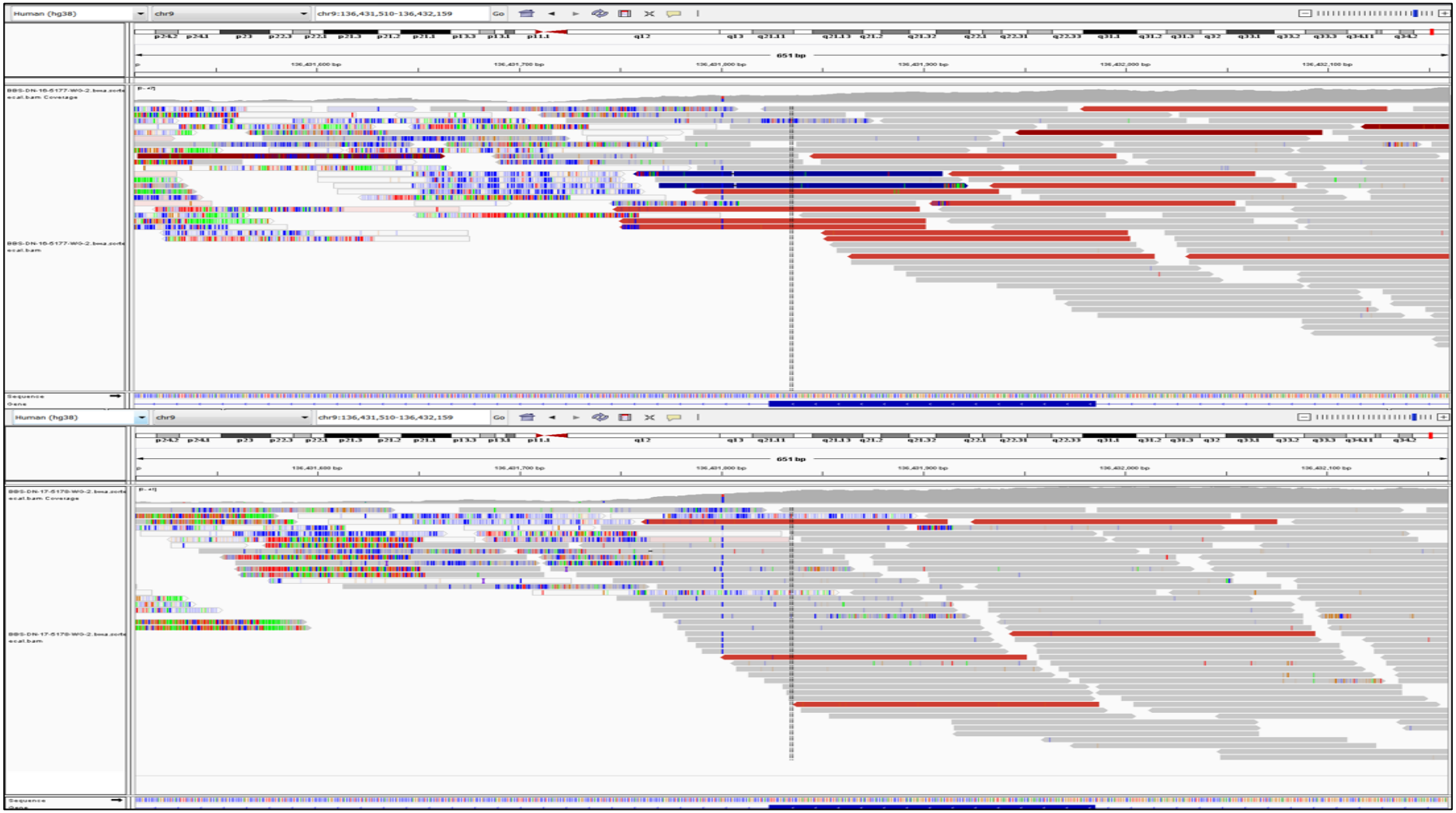

Figure 3.9 Coverage of introns adjacent to exon seven of INPP5E in BBS-016 (top) and BBS-017 (bottom) with soft clipped bases and insert sizes larger than expected (red)

3.3.1.1.3 Variants in *CEP164*

A single rare heterozygous coding variant was found in *CEP164,* p.Arg1392Trp. It was predicted to be deleterious by *in silico* tools. Under ACMG guidelines it is a class three variant of unknown significance. *CEP164* is known to cause nephronophthisis, and one study suggested that while missense variants cause milder disease including RP and Stargardt-type presentations, null variants may cause more severe disease, akin to Meckel or Joubert syndrome(219). *CEP164* is required for the normal formation of the primary cilium and also appears to have a role in DNA damage repair(285, 286). It is also important for targeting INPP5E to primary cilia(283). Although this was a good *a priori* candidate gene, no second coding or non-coding variant with likely functional effect could be identified. No CNV was identified using Lumpy data. Manual scanning of the *CEP164* gene in Integrative Genome Viewer (IGV) showed good coverage throughout. There was one exception, which was a drop in coverage in the middle of exon 29, suggesting a possible partial exon deletion (Figures 3.10 and 3.11). The coverage went from more than 40 reads to fewer than 30 in patient BBS-017 and from approximately 30 to approximately 15 in patient BBS-017. There were no heterozygous SNPs in this region, meaning a partial exon deletion cannot be excluded. The exact size, and therefore effect, of this putative deletion was difficult to determine, as while the drop in coverage was clearly delineated at the 5' end, it gradually increased towards the 3' end. If this could be confirmed it would make *CEP164* a strong candidate gene.

3.3.1.1.4 Variants in ATP8A2

A single heterozygous missense variant, p.Asn61Lys, was identified in *ATP8A2,* one of a number of genes reported to cause cerebellar ataxia, mental retardation and disequilibrium syndrome (CAMRQ). Variants in *ATP8A2* appear to be a rare cause, with only a single family reported. The reported consanguineous Turkish kindred had homozygous p.Ile376Met variants(287). A single case of a child with hypotonia, mental retardation and abnormal movements and a de novo t(10;13) (p12.1;q12.13) balanced translocation with breakpoints disrupting the *ATP8A2* gene has also been reported, with the authors hypothesising that haploinsufficiency of *ATP8A2* accounted for the condition(288). More recently, two cases of encephalopathy, intellectual disability, severe hypotonia, chorea and optic atrophy were found to have *ATP8A2* variants on WES(289). Variants in this gene have not been described as causing ciliopathies such as BBS or Joubert syndrome. The gene itself codes for a P-type ATPase, which is involved in aminophospholipid transport and creating membrane phospholipid asymmetry and which is important for vesicle trafficking across membranes(290). It is expressed in embryonal and adult brain, and also retina and testes as well as elsewhere and appears to have a role in brain development and photoreceptor survival(287, 290). This variant is classified as a variant of unknown significance under ACMG guidelines. No second coding or non-coding variant with likely functional effect or CNV could be identified so *ATP8A2* was not pursued as a possible candidate gene.

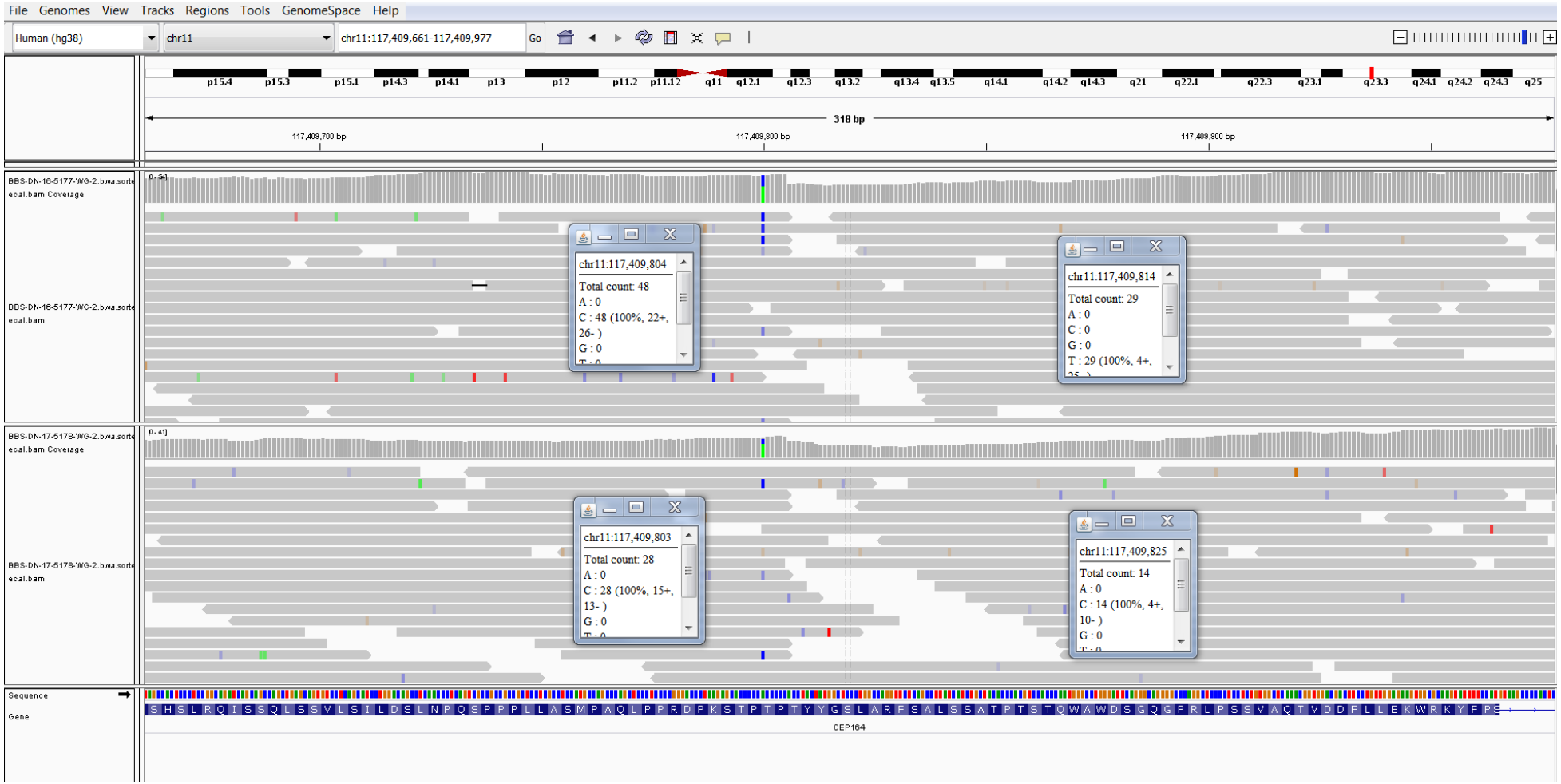

Figure 3.10 Drop in coverage in exon 29 of CEP164 in patients BBS-016 (top) and BBS-017 (bottom). Image from IGV

Figure 3.11 Drop in coverage in exon 29 of CEP164 in patient BBS-016 sorted by base

3.3.1.1.5 Variants in *PKHD1*

Two heterozygote missense variants were identified in *PKHD1*. One of the variants seen, p.Ser3289Ile, was previously reported as pathogenic and is listed in HGMD as pathogenic (CM051184). However, with a population frequency of 0.33, it is commoner than the *BBS1* p.Met390Arg mutation, which is the commonest known causative BBS mutation. In addition, homozygotes were seen in ExAC. In the context of BBS, this variant would be classified as an ACMG class one benign variant. However, as it has already been classed as pathogenic by a reputable source and is observed in a higher frequency in cases than controls, there is contradictory evidence, and so it is an ACMG class three variant of unknown significance.

The second heterozygote missense variant was p.Ile13309Val. This is somewhat rarer, but homozygotes were seen in ExAC. It is listed in ClinVar but no assertions had been made as to pathogenicity. This is an ACMG class three variant of unknown significance. No additional variants or CNVs were identified. As discussed in section 3.2.1.1.6 , *PKHD1* is an unlikely cause of BBS as it has never been associated with polydactyly, learning disability or rod-cone dystrophy, but is in fact associated with hepatic and renal cysts, neither of which the patients exhibit(259). It is also included on diagnostic ciliopathy panels such as the one offered by the North East Thames Regional Genetics Service and has never been implicated in BBS or any ciliopathy other than ARPKD. Therefore it was considered an unlikely candidate causative gene for BBS.

3.3.1.1.6 Variants in *HYDIN*

A single heterozygous missense variant, p.Pro715Leu, was identified in *HYDIN*, a highly polymorphic gene known known to cause PCD and which has never been associated with a non-motile ciliopathy. The variant is rare, and predicted to be deleterious. It is a class three VUS under ACMG guidelines, and as no second coding or non-coding variant with likely functional effect or CNVs could be identified it was not pursued as a candidate gene. It was considered unlikely to be causative and was not pursued further.

3.3.1.1.7 Variants in *NME7*

A single frameshift variant, p.Phe370fs*8, was identified in *NME7*, resulting in premature termination codon. *NME7* is involved in IFT signalling and transport(238, 291). Frameshifts are a known disease-causing mechanism in many ciliopathies, but as no ciliopathy has been reported as being caused by variants in *NME7*, it cannot be used as very strong evidence under ACMG guidelines. The variant is rare and predicted to deleterious and under ACMG guidelines would be classified as class three variant. No second coding, non-coding with likely functional effect or CNV could be identified so it was not pursued as a candidate gene.

3.3.1.1.8 Variants in *DNAH14*

A single rare heterozygous missense variant, p.Gln874Arg, was identified in *DNAH14. DNAH14* is not known to be associated with a ciliopathy, although variants in other axonemal dyneins have been implicated in PCD. The variant is a VUS under ACMG guidelines. A second non-coding variant, c.5146-10T>C, predicted to affect splicing, was identified. It is also an ACMG VUS variant. PCD genes are not associated with non-motile ciliopathies and have functions in the formation

and motility of the dynein arms of motile cilia. The population frequency in GnomAD was higher than the commonest known BBS variant p.Met390Arg in BBS1. For these reasons, *DNAH14* was not considered further as a candidate gene.

3.3.1.1.9 Variants in *ARL13B*

A single rare heterozygous non-coding variant, c.-109+18 was identified in *ARL13B. ARL13B* is expressed in cilia and is important for the targeting of *INPP5E* to cilia(282). Missense variants are known to cause Joubert syndrome and owing to the overlap between Joubert and Bardet-Biedl syndromes, it is a good candidate gene(283, 292). The variant is rare, but it is non-coding and it is not predicted to affect splicing. It is an ACMG class three variant. No second coding, non-coding with likely functional effect or CNV could be identified, and so it was considered unlikely to be causative in this case.

3.3.1.2 Variants in non-ciliopathy disease genes in patients BBS-016 and BBS-017

3.3.1.2.1 Variants in *TRIP12*

A rare heterozygous missense variant, p.Arg1628Gln, was identified in *TRIP12*. *TRIP12* has been implicated in autosomal dominant intellectual disability and p.Arg1628Gln was one of 11 variants reported in this paper, although the same individual had previously been reported by O'Roak et al.(293, 294). It is described as being a causative variant and it is listed in HGMD (CM170867). The individual described had moderate intellectual disability and spoke in phrases, unlike patients BBS-016 and BBS-017 who are more mildly affected. He was described as autistic in the original paper. Interestingly, two of the 11 cases in the Bramswig study, including the patient with the same variant were overweight. A further paper reported additional cases, four out of seven of whom had obesity(295). Unusual facial features including upslanting palpebral fissures and a downturned mouth have been noted. Following correspondence with the authors and curators of the *TRIP12* database, it was confirmed that polydactyly has not been seen in any of the patients. *TRIP12* is a ubiquitin ligase, and the ubiquitin-proteasome system has been implicated in the formation of primary cilia. *TRIM32*, a known BBS gene, is a ubiquitin ligase(296-298). STRING suggests a close association between *TRIP12* and *TRIM32* (Figure 3.12). *TRIP12* also interacts with *UBR5*, a ubiquitin ligase that is believed to have a role in ciliogenesis(299, 300). No other variants, coding or non-coding with likely functional effect, or CNVs could be identified. While dominant ciliopathies, such as ADPKD, are known, BBS is recessive in all known cases. As neither parent has intellectual disability, this would need to be a *de novo* variant. Unfortunately, parental samples have not been obtainable to confirm this.

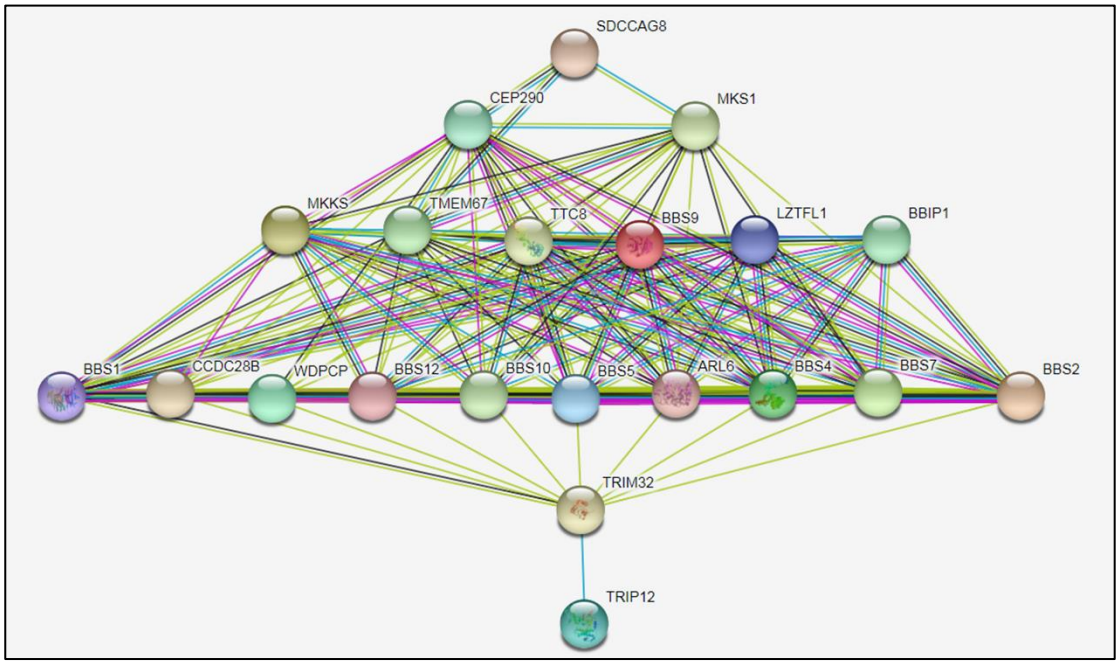

Figure 3.12 Associations between TRIP12 and known BBS genes. Drawn by www.string-db.org. Pink lines represent experimental evidence of connection, black represent co-expression data, blue represent curated databases and green represent text mining data

3.3.1.2.2 Variants in *ARID1B*

A rare heterozygous missense variant, p.Gly949Val, was seen in *ARID1B*. Null and frameshift variants in *ARID1B* are known to cause Coffin-Siris syndrome, an autosomal dominant disorder whose features include dysmorphic features, developmental delay and intellectual disability, hypoplasia of 5th fingers and toes and a range of congenital malformations including brain, cardiac and gastrointestinal anomalies(301). Haploinsufficient variants have also been reported in patients with non-syndromic intellectual disability, while missense variants have been reported in association with non-syndromic short stature(302, 303). The variant seen is rare, but *in silico* tools show conflicting evidence for pathogenicity. It is a VUS when considered under ACMG guidelines. The patients do not have Coffin-Siris syndrome clinically, and although a missense variant could contribute to intellectual disability, there is insufficient evidence to further consider this variant.

3.3.1.2.3 Variants in *GDF5*

A rare heterozygous missense variant, p.Arg267Ser, was seen in *GDF5,* also known as *CDMP1*, a gene known to cause brachydactyly types A and C and other conditions including Grebe chondrodysplasia and proximal symphalangism. Polydactyly has been reported in individuals with Grebe syndrome and occasionally in their unaffected relatives(304-306). A recent case report tells of a father and paternal uncle of a child with a diagnosis of Brachydactyly type C caused by a nonsense variant in *GDF5* who both had postaxial polydactyly(307). The variant seen in this study is a rare missense variant that has not previously been reported as pathogenic. Most of the previously reported missense variants have been between amino acids 370 and 400, with a single

variant reported at amino acid 173 (OMIM *601146). It remains as an ACMG class three VUS. However, given that these patients have polydactyly, it remains of interest, especially if obesity and learning difficulties were attributable to the variant seen in *TRIP12*.

3.3.1.2.4 Variants in *FRMPD4*

A hemizygous missense variant, p.Arg748Gly was seen *in FRMPD4,* which is located on the X chromosome. *FRMPD4* has a role in dendritic spine morphogenesis(308). A truncating variant was previously identified as being causative for x-linked mental retardation and the variant segregated with disease, but no functional studies have been done(309). A patient with a *de novo* missense variant, p.Cys553Arg, also with mental retardation, was identified in the same study, but minimal clinical details are given. None of the patients in the study had polydactyly. Recent, unpublished work from UCL Great Ormond Street Institute of Child Health suggests that *Bbs4* null mice have reduced numbers of dendritic spines(310). One proposed reason for this is that cytoskeletal modelling is impaired in BBS syndrome(311). *FRMPD4* could not be linked with any other ciliopathy genes using STRING or datamining. The variant in question is a VUS under ACMG criteria and was not pursued as a possible causative gene.

3.3.1.2.5 Variants in *CLDN34*

A rare hemizygous frameshift variant, p.Pro47fs*48, was identified in *CLDN34*. Claudins are involved in in tight junction-specific obliteration of the intercellular space, and many have been identified(312, 313). Claudins have been implicated in cancer, but more interestingly, at least one, the product of *CLDN2*, has been found to co-localise with cilia(314, 315). Another, *CLDN19,* has been implicated in a condition known as hypomagnesemia 5, renal, with ocular involvement(316). In the original kindred, neither the renal nor the ocular phenotype are similar to BBS. However a later case was reported where the child had retinitis pigmentosa(317). Reduced levels of CLDN19 have been identified in polycystic kidney disease(316). Variants in *CLDN14* cause sensorineural hearing loss, another ciliopathy feature(318, 319). Expression data suggest that *CLDN34* is widely expressed including in brain, kidney, eye and genitalia, and is localised to the plasma membrane(320). The variant in question is a frameshift, a well-known mechanism of pathogenesis and is not seen in the population databases. As *CLDN34* is x-linked and patients BBS-016 and BBS-017 are male, it is possible that this variant may have a detrimental effect. However, no disease has been reported in associated with this gene and it remains a VUS under ACMG guidelines.

3.4 Discussion

3.4.1 Patient BBS-018

3.4.1.1 Clinical features of patient BBS-018

The patient, who is now 10 years old, has a clear clinical diagnosis of BBS-018. She had been reviewed by several clinicians with an interest in BBS, including a world expert, and all were in agreement about the diagnosis. Her facial features were considered typical. She has bilateral postaxial polydactyly of the hands and unilateral postaxial polydactyly of the feet. She has rod-cone dystrophy and is registered blind. Her weight is now normal, but she had previously been overweight. She has cystic disease of the kidney and had required renal transplantation. She is not known to have genital abnormalities and as she is pre-pubertal, a diagnosis of hypogonadism has not yet been made. Her learning is now considered to be at the lower end of the normal range, though there were more concerns when she was younger, and she had significant difficulties with speech. As postaxial polydactyly, rod-cone dystrophy and renal abnormalities are all considered major criteria, she meets the criteria for a clinical diagnosis. In addition, she has several minor features, such as facial dysmorphism, liver disease, speech delay, poor coordination and brachydactyly. She does not have additional features suggestive of other ciliopathies, such as Joubert syndrome, and in particular there is no evidence of a molar tooth sign or lissencephaly. She does have additional features such as Crohn's disease and situs inversus. Situs inversus has been reported in BBS, including in a case with homozygous null variants in *BBS8* and in a case with *LZTFL1* variants(264, 321). This meant that variants in known BBS genes were the primary candidate genes for this patient. Her parents are non-consanguineous.

3.4.1.2 Choice of filtering criteria

The most common known variant causing BBS is the p.Met390Arg variant (rs113624356) in *BBS1*, causing approximately 80% of BBS1-related cases and 30% of cases overall(227). The minor allele frequency in ExAC for this variant is 0.15% and 0.1% in 1000Genomes. Variants with a frequency of >0.5% in any of 1000Genomes, ExAC, GnomAD or NHLBI-ESP were excluded from analysis unless they were known to be pathogenic, as they would be a common cause of disease if they were found this frequently. Currently, a molecular cause for BBS cannot be identified in only approximately 20% of BBS cases, meaning that the population frequency of any previously unknown pathogenic variant should be less than that of the p.Met390Arg variant(18). Initially, coding variants were considered as the majority of pathogenic variants in BBS genes are coding. However, splice site variants and small and large deletions have been identified as causative(18). Promoter variants have not been reported as a cause of BBS, though they are known to cause other Mendelian diseases(322-324). Therefore, non-coding variants were considered after the analysis of coding variants. Details of genetic filters applied can be found in Chapter 2 and Appendix 2, Tables A2.1-A2.4.

Following analysis, many variants were excluded from further consideration. Three interesting candidate genes remained: *CEP290*, a known BBS gene in which two variants of interest were seen, *NPHP4*, where a known pathogenic mutation was identified, and *CHDR1*, where a likely

pathogenic variant was seen. Also of interest was *PCM1*, where coding and a non-coding variants were seen.

3.4.1.3 *CEP290* as a candidate gene

CEP290 is one of the most interesting ciliopathy genes as it has been associated with such a wide range of ciliopathy phenotypes- Bardet-Biedl, Joubert, Leber congenital amaurosis, Meckel and Senior-Løken syndromes(176, 178-180). The gene, located at chromosome 12q21.32, consists of 55 exons and codes for a 2479 amino acid, 290KD protein that is widely expressed(180, 325).

CEP290 appears to play an important role in the assembly of the primary cilium(261, 326). It is recruited to the centriole by PCM-1, and *CEP290* knockout disrupts normal localisation of PCM-1. PCM-1 is known to interact with the product of *BBS4*, which forms part of the BBSome. Both CEP290 and PCM-1 are required for correct localisation of Rab8, which is essential for the formation of the ciliary membrane(149, 216). Once the primary cilium has been formed, CEP290 has a role in mediating IFT, by recruiting proteins such as CC2D2A and NPHP5, although this process is not yet well understood(327, 328). When ciliogenesis is not required during cell division, CEP290 is supressed by an inhibitor, CP110(326).

Sayer et al. identified a number of functional domains in *CEP290* when analysing the amino acid sequence, including 13 coiled-coil domains; an ATP/GTP binding site; areas homologous to structural maintenance of chromosomes (SMC) chromosome segregation ATPases; a domain homologous to tropomyosin; several kinase-inducible domains; and a nuclear localisation signal region(180). The function of various parts of the CEP290 protein has been further elucidated(329).

CEP290 appears to be a very rare cause of BBS, with only a single report in the literature(176). Of note, this patient had homozygous null variants in *CEP290*, p.Glu1903Ter but also a complex *TMEM67* allele, suggesting that the phenotype might be dependent on the presence of modifiers(176, 330). A single patient with compound heterozygous nonsense *CEP290* mutations (c.322C>T p.(Arg108Ter) and c.5668G>T p.(Gly1890Ter)) is known to the national BBS clinic(331). Pathogenic variants in *CEP290* are generally frameshift or null variants, with missense variants only rarely reported(217, 241). However, a recurrent intronic variant has been identified in Leber congenital amaurosis (LCA) that creates a splice donor site and results in a premature termination codon following the insertion of a cryptic exon into the mRNA(181, 243). Genotype phenotype correlation in *CEP290*-related disease is difficult but a recent study has proposed that it may depend on the total amount of protein produced(332). Genes coding for proteins that interact with CEP290 including *CC2D2A*, *NPHP5*, *RPGR* and *TMEM67,* are known ciliopathy genes(333-335).

A cellular phenotype has been identified in *CEP290*-related LCA, where fibroblast cells from patients had fewer cilia than controls and the cilia were found to be shorter(336). Interestingly, gene therapy with full length *CEP290* delivered in a lentiviral vector rescued the ciliary defect. Another study by Shimada et al. showed that while cilia appeared normal in *CEP290*-related LCA

patient-derived fibroblasts, the photoreceptor cilia were abnormal. In *CEP290*-related JBTS patient-derived fibroblasts however, there were fewer cilia. Unlike in the previous study by Burnight et al., these cilia were longer than in controls(337). There is conflicting evidence as to whether structurally abnormal cilia are found in BBS. One study found no structural evidence of structural abnormality in Bbs4 knockout mice, while another found a reduced ciliation percentage and shorter cilia in *Bbs4-/-* mouse renal cells(204, 311). However, no abnormal ciliary phenotype has been observed in cultured patient fibroblasts from patients with BBS1 or BBS10(338).

As discussed in section 3.2.1.1.1, a coding and a possible promotor variant were identified. Sanger sequencing was done to attempt to clarify whether variants were in *cis* or in *trans*, but while both variants were detected in the patient, neither was found in the mother. Paternal samples are unavailable, so it remains possible that the variants are *in cis* and can be discounted as the cause. The other possibilities are that one variant is present in mosaic form in the mother or that it arose *de novo* in the patient. This is not possible to clarify at present.

In order to further elucidate pathogenicity there are further steps that could be taken. To determine variant phase, maternal saliva or skin samples could be taken or deep resequencing done to try to prove mosaicism. However, even if the variants were determined to be in *trans*, this would not clarify pathogenicity. A next step would be to undertake functional studies. This might include RTPCR to look at RNA levels, antibody staining of CEP290 to see if staining appears normal and possibly making a cell or animal model using a technology such as CRISPR-Cas9 gene editing(339, 340).

3.4.1.4 *NPHP4* as a candidate gene

As a known pathogenic mutation was identified in *NPHP4*, a ciliopathy gene implicated in Senior-Løken syndrome (SLS) and nephronophthisis, it was a strong candidate gene. The patient has both retinitis pigmentosa and cystic renal disease and other genes causing SLS, *CEP290 and SDCCAG8,* have been implicated in BBS. NPHP4 protein localises to cilia at the basal body as well as to the centrosomes of actively dividing cells(341). The products of *NPHP4* and *NPHP1* have been found to interact and are thought to form part of a common signalling pathway(342).

In general, the pathogenic mutations observed in NPHP4 have been nonsense or frameshift variants, with few missense variants reported. The paper that reports this variant as pathogenic shows that it is in *trans* with a frameshift variant. The only additional coding or non-coding variant seen in BBS-018 was a known variant, thought to be a benign polymorphism, seen at relatively high frequency in the population, and therefore very unlikely to be causative. However, it is possible that a small CNV may be being missed. Coverage of *NPHP4* is good and so it is unlikely that a coding variant has been missed. However, a deep intronic variant which has not been predicted to affect the promoter, a transcription binding site or a slice site may be responsible.

Again, functional work would be required to see if *NPHP4* is responsible for BBS in this patient. RTPCR and cell staining would be a good initial approach. Sanger sequencing of the entire gene might also pick up a variant that has been missed by NGS.

3.4.1.5 *CDHR1* as a candidate gene

CDHR1 was an interesting candidate because the variant seen was a frameshift, a common mechanism of disease causation in ciliopathies. *CDHR1* has been implicated in isolated rod-cone dystrophy, but not in syndromic ciliopathies(250). The mutations found previously are one base-pair deletions or duplications resulting in frameshifts like the one seen in BBS-018(250, 343).

However, *CDHR1* appears to have expression restricted to the retina, making it an unlikely candidate for a multisystem ciliopathy(249). In view of this, and the fact that no second variant could be identified, it seems that it is unlikely to be worth pursuing. However, it is known that modifiers are important in ciliopathies, and variants such as this may be important in modifying the phenotype.

3.4.1.6 *PCM1* as a possible modifier

As discussed in section 3.2.1.1.8, knockout of *PCM1* is likely to be lethal and mutations in *PCM1* have never been identified as a cause of a ciliopathy. PCM1 is known to interact with CEP290, and it is possible that if *CEP290* were the causative gene, the phenotype severity could be modified by the presence of a *PCM1* variant. It is thought that the presence of a *TMEM67* variant may have modified the phenotype of the case of BBS that presented with *CEP290* mutations(176). However, as discussed in section 3.1.5.3.4, proving the effect of modifiers is challenging, even in the presence of known pathogenic mutations.

3.4.1.7 Digenic inheritance

Another possible explanation to consider is digenic inheritance, where a disease phenotype is present only when heterozygous variants in two genes are present, but a single heterozygous variant is not enough to cause disease(344). There are examples of this, including in non-syndromic hearing loss, neural tube defects and Alport syndrome(345-347). It has also been reported in Joubert syndrome and retinal dystrophy caused by ciliopathy gene mutations(348, 349). The main candidate variants for digenic inheritance in BBS-018 are the heterozygous coding variants in *PCM1*, *CEP290* and the promoter variant in *TTC8* (Figure 3.13). PCM1 and CEP290 proteins are believed to interact. The product of *TTC8*, in which a possible promoter variant was seen, is also thought to interact with the product of *PCM1*. It is possible that a combination of two or all of these variants could cause the observed phenotype. Proving this would be difficult as it is likely to require an animal model with the specific mutations introduced.

3.4.1.8 Summary of findings in BBS-018

No definitely causative variants were identified, but some candidates were identified for possible future work. Reasons why the cause may not have been identified are discussed in section 3.4.3.1.

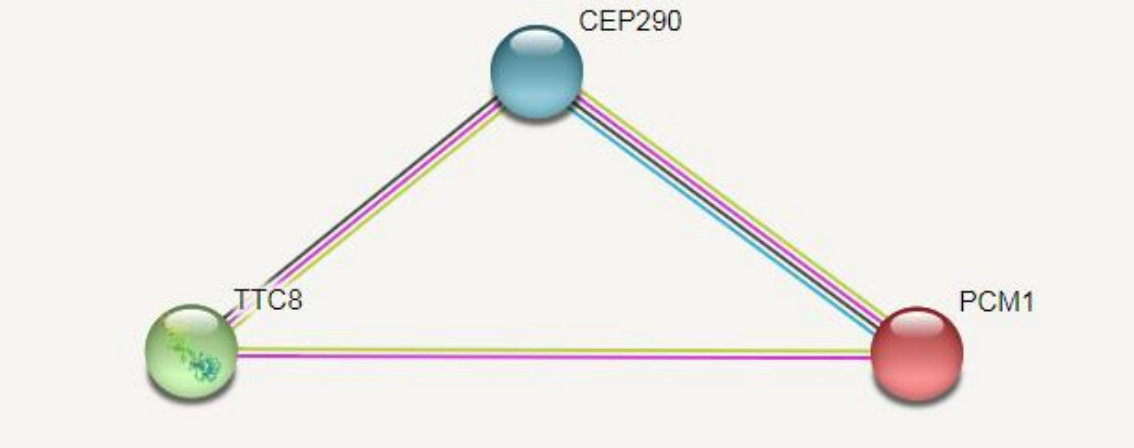

Figure 3.13 Associations between CEP290, PCM1 and TTC8. Drawn by www.string-db.org. Pink lines represent experimental evidence of connection, black represent co-expression data, blue represent curated databases and green represent text mining data

3.4.2 Patients BBS-016 and BBS-017

3.4.2.1 Clinical features of patients BBS-016 and BBS-017

The monozygotic twins, BBS-016 and BBS-017, born of unrelated parents and who are now aged 19, have had a longstanding clinical diagnosis of BBS, and have been seen by several clinicians with expertise in BBS, including a world-expert. They are both considered to have typical facial features, both have obesity and hypogonadism. Both have postaxial polydactyly of the hands, though not of the feet. Both have mild learning difficulties and had developmental and speech delay when younger. Of interest, neither displays any evidence of rod cone dystrophy, clinically or on electroretinogram. Neither has any evidence of renal failure, structural renal or liver abnormalities or endocrine features. Both are normally grown, with heights above the 75th centile. They have three major and two minor features, meeting the criteria for a clinical diagnosis of BBS. This meant that variants in known BBS genes were the primary candidate genes for these patients. In addition, only variants seen in both were considered, with the exception of variants in known BBS genes in either, which were then manually checked in IGV to see if they had been missed or if coverage was poor.

3.4.2.2 Choice of filtering criteria

The choice of filtering criteria was the same as for patient BBS-018 and are discussed in Chapter 2. In addition, only variants seen in both were considered, with the exception of variants in known BBS genes. If a BBS gene variant was seen in one but not the other, they were manually checked in IGV to see if they had been missed or if the coverage was poor in the other. Only deep intronic variants in BBS genes were identified in one but not the other and were not seen in the second patient when checked in IGV.

3.4.2.3 *ABCA4* as a candidate gene

Although two variants previously determined to be pathogenic were identified in *ABCA4*, this gene has never been associated with a multisystem ciliopathy, but only with eye phenotypes including

autosomal recessive retinitis pigmentosa, Stargardt disease and adult-onset macular degeneration. In addition, the gene appears to be expressed only in retinal tissue, making it a poor candidate gene for BBS syndrome. The patients in this case do not have any evidence of retinal degeneration on ophthalmological testing, so this gene can probably be excluded as causative. However, it does raise the interesting question of secondary findings, which are discussed further in section 3.4.3.5.7. In research studies such as this, secondary findings were not traditionally returned, and only findings related to diagnosis would be verified in a diagnostic laboratory and communicated to patients. However, this practice is changing, and consideration needs to be given to what will be returned and how consent will be obtained. This is likely to become even more of an issue as more treatments for genetic disease are developed. In the case of these patients, they are already having regular ophthalmology examination, and so it is likely that any macular degeneration would be picked up early. However, there are lifestyle modifications that can probably reduce the risk of developing macular degeneration, such as avoidance of smoking, so there might be utility to knowing about this result even now.

3.4.2.4 CEP164 as a candidate gene

CEP164 is known to cause nephronophthisis but with reported phenotypes ranging from Stargardt disease to much more severe Meckel and Joubert syndrome phenotypes it is a good candidate gene. It is widely expressed, being found in almost all tissue types(350). CEP164 is believed to have a role in UV-mediated DNA damage repair and a role in targeting INPP5E to cilia(283, 286).

Only one rare coding variant could be identified. While non-coding variants were present most were common, none of them were near and intron-exon boundary, and none were predicted to affect splicing. However, a possible small deletion was seen in exon 29. Further work would be required to look at whether *CEP164* is the causative gene. This might include RTPCR to look at RNA expression levels or staining to look at the presence of protein. Proving the presence of the deletion is difficult as it is small. Parental samples might help, but they have not been available so far. As discussed in section 4.3.3, long-read genome sequencing will help to elucidate such variants.

3.4.2.5 *INPP5E* as a candidate gene

INPP5E is implicated in MORM and Joubert syndromes. *INPP5E* is targeted to the cilium and interacts with other ciliary proteins including *CEP164* and *ARL13B*(283). In this case a single variant of unknown significance was seen, but no second variant could be identified when the putative deletion was discounted. Interestingly, as variants have been seen in both *INPP5E* and *CEP164* whose products are known to interact, it is possible that variants in one could modify a phenotype caused by the other.

3.4.2.6 *TRIP12* as a candidate gene

As discussed in section 3.3.1.2.1, a variant previously reported as pathogenic was identified in *TRIP12. TRIP12* is not a known ciliopathy gene but it is a ubiquitin ligase. Other members of the family are known to play roles in ciliogenesis and *TRIM32,* a known BBS gene, also encodes a

ubiquitin ligase. The patient in the Bramswig study with the same variant as patients BBS-016 and BBS-017 have moderate learning difficulties and obesity(293). While 90% of BBS patients will develop rod cone dystrophy at some point, patients BBS-016 and BBS-017, now in their late teens, have no evidence of it yet, either clinically or electrophysiologically. A *TRIP12* variant would not account for polydactyly or hypogonadism.

One of the first steps in investigating variants causing dominant disease is to determine whether a variant is inherited or *de novo*. If it were not *de novo* it would raise questions about the pathogenicity of this variant in the context of parents with normal learning. Unfortunately, DNA from parents is not yet available, so this analysis has not been performed. Also, if the hypothesis were to be that *TRIP12* is a ciliopathy gene causing a BBS phenotype, it would be the first reported case of an autosomal dominant form of BBS. Another possibility is that *TRIP12* is a recessive BBS gene but a second variant or CNV has been missed. Lumpy data do not confirm this, nor are there any areas of low coverage that would be likely to result in variants being missed.

Another possibility is that the *TRIP12* variant accounts for the patients' developmental delay and obesity, but not for the polydactyly or hypogonadism. As the hypogonadism is a clinical rather than a biochemical diagnosis, with both patients achieving puberty without assistance and biochemical hormone tests within the normal range, this diagnosis may possibly be due to external genitalia looking small because of surrounding adipose tissue. As discussed in section 3.3.1.2.3, a single variant in *GDF5*, which has been implicated in polydactyly was identified, but it is not one previously identified and cannot be said to be pathogenic

3.4.2.7 Summary of findings in BBS-016 and BBS-017

No definitively causative variants were identified in patients BBS-016 and BBS-017. However, interesting variants in several genes were identified, including a variant previously reported as pathogenic in a known autosomal dominant intellectual disability gene. Reasons why cause of disease may not have been identified are discussed further in section 3.4.3.1.

3.4.3 Utilisation of next–generation sequencing for diagnosis

3.4.3.1 Reasons for lack of diagnostic success using whole genome sequencing

Three patients (one singleton and a pair of monozygotic twins) with a clinical diagnosis of Bardet-Biedl syndrome had whole genome sequencing performed in an attempt to achieve molecular confirmation of the diagnosis. Following a lengthy period of analysis, the molecular diagnosis could not be confirmed in either the singleton (BBS-018) or the twin pair (BBS-016 and BBS-017), and no candidate gene was identified with more than one coding variant. There are various reasons why this might be the case.

- The causative variants may have been identified e.g. *CEP290* for BBS-018 and *CEP164* for BBS-016 and -017 but there is insufficient evidence to confirm pathogenicity.
- The gene or genes may be a very rare cause of BBS and one that has not yet been associated with any disease so it may not have been picked out during analysis.

However, all genes currently known to be associated with cilia were examined, as were any genes with a stop gain or frameshift variant. All known morbid OMIM genes were looked at and rare coding variants in them considered also.

- The causative gene may be one of those identified as having a single coding variant but the second variant may be a deep intronic variant. This is increasingly being recognised as a mechanism of disease, including in ciliopathies(351-353). However, the number of variants obtained in whole genome sequencing, and the current lack of knowledge in how to identify pathogenic deep intronic variants means that these are very difficult to identify, especially in small rare disease cohorts.
- Certain genes or gene regions may be covered poorly or not at all, leading to variants being missed.
- Filtering strategies and quality control filtering may lead to the removal of variants which would then not be considered during analysis. For example, variants covered at a read depth of less than 10 were filtered out during analysis. While this is a recognised strategy to avoid identification of spurious variants, it does increase the risk of missing true variants that are poorly covered.
- CNVs can be picked up using whole genome sequence but it may not be very sensitive and it is difficult to pick up small insertions and deletions.
- Somatic variation may be missed by whole genome sequencing, and while this is unlikely in the case of monozygotic twins, it is possible for a singleton patient.
- Certain genetic variants such as short tandem repeats (STR) may be poorly called by short read sequencing technologies. While STRs have never been identified as a mechanism of causing ciliopathies, and is unlikely to be a cause, this might be an issue in other conditions.
- The mode of inheritance is not autosomal recessive. While dominant inheritance has not been reported for ciliopathies such as BBS, digenic inheritance has been seen in ciliopathies such as Joubert syndrome.

3.4.3.2 Advantages of WGS over WES and other methods

While the advantages of next-generation sequencing approaches over older methods such as Sanger sequencing, such as the ability to sequence genes in parallel rather than in series, have been accepted, it is becoming clear that WGS has some advantages over other next-generation sequencing methods such as WES or panel-based sequencing.

3.4.3.2.1 Detection of intronic variants

As mentioned in section 3.4.3.1, deep intronic variants are increasingly being recognised as a disease-causing mechanism(351-353). WGS gives much better coverage of intronic regions than WES, which may only cover the first few base pairs after an intron-exon boundary. As yet, not much is known about deep intronic variants and how to identify pathogenic ones, but this is changing with our increased understanding of splicing and the annotation of the human genome with branchpoint and splicing information(74, 354-356). Computational modelling of splicing is challenging, but this is also likely to improve in time. Tools such as Genomizer may also assist

with this(357). WGS means that the sequences are already available as knowledge increases, and tests do not need to be repeated in individuals with negative genetic test results. Reanalysis of the results of this study will be important as knowledge increases.

3.4.3.2.2 Improved coverage and detection of coding variants

There is now evidence that WGS provides better coverage than WES at equivalent read depths, and that WES requires a read depth 2 to 3 times greater to give equivalent coverage(118). Coverage is one of the most important factors determining the ability to correctly identify variants. The capture methods employed by WES also introduce bias, especially in difficult to sequence regions such as those with high GC content(118, 358). Overall, WGS allows for better detection of coding variants(359).

3.4.3.2.3 Improved coverage and detection of CNVs

WGS also allows the detection of CNVs, and is better than WES at delineating their exact boundaries and at picking up small CNVs(117, 360, 361). Utilisation of more than one method can increase accuracy(362). Long-read sequencing technologies are likely to further improve the accuracy of this(363, 364).

3.4.3.2.4 Flexibility of approach

WGS can be used to look at all coding and non-coding variants if desired. However, it is possible to analyse only the coding regions and not consider any of the non-coding variants, or even to look at a virtual gene panel instead, while still retaining the ability to expand the analysis to include other genes of interest and non-coding regions. This allows a stepwise approach during initial analysis or the ability to revisit results as understanding of the non-coding regions of the genome increases.

3.4.3.2.5 Added value data

In addition to diagnostics, other information can be gleaned from sequencing data. A good example of this is pharmacogenomic data, discussed further in Chapters 5 and 6. A number of important pharmacogenomic SNPs are intronic, and so WGS is the superior technology for this.

3.4.3.3 Disadvantages of WGS over WES and other methods

3.4.3.3.1 Large numbers of variants

The immediately obvious disadvantage of WGS is the very large numbers of variants that can be identified in a single genome, particularly if it is not done as part of a trio or with controls run on the same platform. For example, analysis of the coding regions of the genome of BBS-018 resulted in 4,725 variants that passed quality control filtering. Analysis of the entire genome resulted in 4,878,324 variants. While this number can be significantly reduced by using filtering strategies such as exclusion of common variants, it still represents a significantly increased workload in terms of numbers of variants to be considered. Many of these remain as variants of unknown significance, despite best efforts to classify them as benign or pathogenic.

3.4.3.3.2 Increased requirements for expertise

Non-coding variants with clinical significance are more difficult to identify as discussed in section 3.4.3.2.1. Clinicians and scientists who may feel relatively confident about classifying coding variants, particularly in familiar genes, may not have the expertise to classify non-coding variants in the same way. Building up of expertise, along with data collection and publication, is essential to maximise WGS as a diagnostic and research tool.

3.4.3.3.3 Increased cost, equipment and computational requirements

While the cost is decreasing all the time, WGS is still more expensive than WES, especially for long-read technology. In addition, while many smaller and diagnostic laboratories are equipped to perform WES, few are doing WGS in-house, instead outsourcing to major companies. This may come with a spending or contract-duration commitment. There is also a significantly increased requirement in terms of computational resources required for storage and analysis, as well as additional staff costs if analysis is more time-consuming than for WES or panel-based diagnostics. In addition, the turnaround time for WGS may be longer than for WES, but this is changing with rapid sequencing and analysis methods. The time and resources required to consent properly for testing and return complex results must also be taken into account.

3.4.3.3.4 Validation requirements

While this is likely to change, currently the gold standard is for next-generation sequencing results to be validated in a diagnostic laboratory using Sanger sequencing. This adds expense and time to the diagnostic process. Evaluation of this process has shown that it may not be necessary, with the vast majority of variants from high-quality next-generation sequencing data being confirmed(365, 366). Some countries, such as the Netherlands, no longer validate NGS variants.

3.4.3.4 Increasing the efficiency of WGS diagnostics

Although WGS does have disadvantages compared to WES, these need to be offset against the advantages, particularly in terms of future-proofing. As more is known about intronic variants and our ability to detect structural variants increases, the greater the utility of WGS sequencing, even retrospectively. WGS is increasingly being used for diagnostics, although this is not without difficulties. However, there are some strategies which may make this easier and more efficient

3.4.3.4.1 Recording and sharing of analysed variants

One of the most effective ways to increase understanding of variants is to share data on variants that have been seen before and the phenotype they have been seen with. Currently while such databases exist, they are often difficult and time consuming to add variants to and frequently include little phenotypic data, limiting their usefulness. Large studies such as Deciphering Developmental Diseases (DDD) have developed databases that share both variants and clinical data. The 100,000 Genomes project has plans to do the same. However, many smaller research groups and diagnostic laboratories are not routinely sharing data, and until this happens there will be inevitable duplication of effort.

3.4.3.4.2 Improved population databases

While there are many population databases, the more population data that are available, the more accurate filtering strategies for variants can be. This is particularly true for intronic variants about which there is less data. This is something that will improve as more WGS is done. An additional issue is that some populations and ethnic groups are less well covered than others, making interpretation of variants in these groups difficult. Something that may be extremely rare in Caucasians may be quite common in other populations, but may appear to be of significance if a patient from this population is having variants filtered based on inaccurate population data(367). While less of an issue in rare disease, another problem is that not everyone in these databases is disease free. This is particularly true for late onset diseases such as dementia and common, multifactorial conditions such as diabetes and hypertension(368).

3.4.3.4.3 Analysis and comparison of variant calling and filtering algorithms and platforms

Most laboratories develop their own bioinformatics pipelines. The validation of these with highly sequenced samples, such as Genome in a Bottle, is vital as is the comparison of various sequencing platforms, and the publication of these data(369).

3.4.3.4.4 Introduction of new technologies

Novel WGS technologies such as long-read sequencing will increase the ability to sequence repetitive regions of the genome, to identify and characterise structural and CNVs, increase variant detection and analyse variant phase(370-373). Currently this technology is expensive, but as with all sequencing technologies, costs will drop in time. Increased understanding and integration of multiomics data may also help in the understanding of WGS data, and may be useful in the validation of variants of unknown significance.

3.4.3.4.5 Utilisation of trio genomes

While sequencing trios is expensive, currently it is a useful strategy for reducing the number of variants requiring analysis, especially in the setting of recessive or *de novo* dominant disease. This has been part of the strategy used by DDD and 100,000 Genomes studies, and may be a reasonable strategy in some circumstances, costs permitting

3.4.3.4.6 Revisiting previously analysed results and recontacting patients

As previously discussed, one of the main benefits of WGS is the ability to revisit and reanalyse data, particularly as population databases are added to, more variants are published and more understanding of non-coding variants is gained. It is best practice to have a procedure in place to do this, rather than reanalysing on an *ad hoc* basis, as that increases the risks that certain patients, such as those that miss follow up, may not have their results reconsidered. This is an issue with many historical studies, even where single genes were sequenced, as historically, some variants were called pathogenic where current data indicate that they are not, and other variants were dismissed because they were common while now they are known to be pathogenic(277). A case that recently went to the Supreme Court in the US, *Williams v Quest/Athena,* dealt with the death of a child with an *SCN1A* variant which was classified as a VUS at the time of testing. The report was later reissued without update despite several published

papers that suggested pathogenicity of the variant. However, the case did not clarify what the duty to reanalyse and recontact is because the laboratory was determined to be acting as a healthcare provider and the statute of limitations for wrongful death suits had run out(374). However, the plaintiff has the option to sue under medical negligence laws and so clarification may come from a later case.

3.4.3.4.7 Recording and publication of accurate phenotyping data

This is discussed in detail in Chapter 4, but high-quality phenotyping data can aid interpretation of WGS data. HPO terms are particularly helpful for comparison of patients and computational analysis of data.

3.4.3.5 Additional issues with Next-Generation sequencing data

There are some additional issues that affect all next-generating sequencing studies and diagnostics

3.4.3.5.1 Obtaining informed consent

Obtaining informed consent is a long-established principle in genetic testing. However, as the amount of sequencing data generated increases, so do the things that can be inferred from it. In addition, genomic data are a hugely useful resource and so consent may need to take account of future uses of the data, as well as unexpected findings of clinical importance. All of this increases the complexity of obtaining informed consent as well as the time taken to do it(375). The time required for obtaining truly informed consent for WGS has been estimated at up to eight hours(376)

3.4.3.5.2 Variable quality of published variants

Once a variant has been published as benign or pathogenic, it is highly likely that further instances of that variant will be classified the same way for the same disease. The ACMG guidelines state that previously published pathogenic variants can be recognised as disease-causing(85). Even for novel variants, one of the ACMG guidelines for determining pathogenicity of variants refers to a variant having the same amino acid change as a previously established pathogenic variant. Incorrectly classified variants in the literature are then reinforced by later published data. When using previously established pathogenicity as a criterion for determining pathogenicity, it is important to review the reasons that variant was called pathogenic previously, especially for variants reported only once or a small number of times, and to consider whether this was based on functional studies or just its presence in an affected individual and not in population data. HGMD is full of examples of variants that were initially called pathogenic and later downgraded to VUS and then probably benign or benign(111). A 2011 study looking at severe, recessive, childhood-onset disease determined that up to 27% of variants in the literature were misannotated(377). Taking a diagnosis away from a patient is incredibly difficult and may undermine trust in clinicians. In addition, wrongly ascribing pathogenicity to a variant has major consequences in terms of recurrence risks, and increasingly, in access to treatments, and it also removes the opportunity to correctly identify the actual pathogenic variant.

3.4.3.5.3 Problems with functional prediction tools

One of the supporting criteria in the ACMG guidelines is computational evidence of pathogenicity. There are various tools for this including PolyPhen2 and SIFT as discussed above and CADD scores. Prediction tools often do not work for intronic variants, and even for coding variants are frequently contradictory and may perform differently depending on the genes being studied(378). Many well-established pathogenic variants will be classified as benign by these tools. For this reason, they were not chosen as a filtering strategy in this study. They are likely to improve with time, especially if good datasets of pathogenic variants are collated and used as training sets, but it is important not to use them to add too much weight to determining pathogenicity.

3.4.3.5.4 Changes to the reference genome

The move to GRCh38 from GRCh37 increases the accuracy and completeness of the reference genome. However, it does mean that variants may be renamed in various iterations of the genome, making it difficult to compare previously published variants to variants of interest. This is partially resolved by resources such as dbSNP, but it still increases the risk that variants may be misannotated and can make it difficult to identify prior reports of the variant in the literature.

3.4.3.5.5 ACMG guidelines for variant annotation

While they are widely used and very useful for classification and annotation of variants, there are some issues with the ACMG guidelines(3, 4). One of the main reasons for this is that some of the standards are subjective and open to differing interpretations(4, 379). The guidelines are likely to be refined over time and there are various computational methods being developed to overcome some of the difficulties(3). Because of the risks of misannotated variants, the ACMG guidelines make it difficult to move a variant from VUS to benign or pathogenic which can lead to the accumulation of VUS on a patient's record.

3.4.3.5.6 Returning results to patients

Clinicians have been returning pathogenic results to patients for many years. However, the question remains about how much patients should be told about VUS and plans to review results, and how awareness and understanding of VUS may impact patients' lives(380). This is something that is probably best managed by local policy and may depend on the level of a patient's interest and involvement as well as factors such as cost and time. Many patients do not find living with a VUS easy(381). A clear local protocol delineating how these patients should be managed is important for consistency, for example, protocols for screening and management of patients with VUS in cancer predisposition genes.

3.4.3.5.7 Returning variants of clinical significance not related to diagnosis

WGS provides an opportunity to identify variants of clinical significance unrelated to a patient's diagnosis, often referred to as incidental or secondary findings. Obvious examples include pathogenic variants in cancer predisposition genes, in genes for late-onset neurological disease such as Huntington disease or genetic dementia or heterozygous variants for recessive disease (carrier status). When considering whether or not variants are reported back to patients, consideration should be given to whether interventions are available to prevent or mitigate

disease and whether knowing this information will allow patients to make decisions which would change outcomes. The ACMG recommends the return of secondary findings(5). Currently the 100,000 Genomes project offers participants the opportunity to find out about variants in cancer predisposition genes (*MLH1, MSH2, MSH6, MUTYH*, ***APC**, BRCA1, BRCA2,* ***RET**, **VHL**, **MEN1***)* and familial hypercholesterolaemia genes (***LDLR**, **APOB** and **PCSK9***). In addition they offer screening for cystic fibrosis carrier status (*CFTR*) (https://www.genomicsengland.co.uk/taking-part/results/)(382). This is an opt-in part of the study, and patients are warned that the genes screened for variants may change over time with results being returned many years after enrolment. Some results are returned to adults participants only, and others (in bold) to adults and children. The 100,000 genomes study does not currently return results from all the genes recommended for consideration by the ACMG.

There are significant implications when feeding back secondary results such as the above, not least in terms of time and cost. Studies have estimated that if ACMG guidelines were followed, rates of incidental findings would be identified in up to 3.3% of individuals(4, 383). In addition, consideration needs to be given to reviewing the list of genes from which results will be fed back, and whether or not existing results need to be reviewed when this list changes, as discussed above for primary findings. Late onset neurological conditions such as Huntington disease, which are currently not screened for as they are incurable, may be included on the list should a treatment become available, especially if treatment is recommended during a pre-symptomatic or prodromal phase. Patients need to be warned that incidental findings are both a risk and benefit of next-generation sequencing and understand that the list of conditions is far from comprehensive. This adds to the complexity of obtaining informed consent in the diagnostic setting, where there may be significant time pressures(375). However, as well as the cost of feeding back results, the potential benefits of preventing these conditions, both economic and non-economic, must be considered(384). A further thing to consider is the patient's right not to know, and considering this should form a part of the consent process

3.4.3.6 Use of WGS data for NHS diagnostics

NHS England plans to introduce limited diagnostic WGS from 2019, though the exact details of which conditions will be eligible for WGS are not yet known. It is known that other specialties besides clinical genetics will be able to request tests related to their specialty, and will therefore be consenting patients for WGS, receiving the results and returning them to patients, a challenge in the current time-pressured climate of the NHS. Another thing to be considered is that while geneticists are used to considering implications to other family members, other specialists may not routinely consider this, which may be of particular importance when it comes to incidental findings, which may be different to the ones being tested for when patients were consented, and in many cases, entirely unexpected. It is not clear either how incidental findings will be fed back to patients or by whom. As with 100,000Genomes data, it is likely that genomes will be analysed initially using a virtual gene panel, but results and incidental findings will still need to be revisited. It is an opportunity to collate high-quality data about phenotype and variants and to generate a national genomics resource. However, in the current NHS funding climate, the increased cost of

genome sequencing and the health economic arguments for and against WGS must be considered. Sanger sequencing of variants adds to the cost and it remains to be seen whether this will be done when WGS moves into the diagnostic setting. In addition, access to education and training in genomics for all clinicians will be essential, both for the understanding and conveying of results, but also for understanding and dealing with the familial implications and the ability to properly consent patients.

3.5 Conclusions and future directions

The variants causing disease were not definitively identified for BBS-018 or BBS-016 and BBS-017. However, candidate variants were identified. Functional work will now be required to determine whether these variants are pathogenic or not. The identification of VUS can be frustrating for both clinicians and patients and a decision has to be made on which, if any, should be fed back to patients. In the case of patient BBS-018, the variants in *CEP290* have been discussed with an international expert, and he is unconvinced that they are causative but suggests that functional work might be appropriate. In addition, it is worth considering additional sequencing of *CEP290* and *NPHP4* in case variants have been missed. Of interest, since this analysis was carried out, the *CEP290* variant has been added to ClinVar, where it is assessed as a variant of unknown significance, having been reported in an individual with Joubert or Meckel syndrome features (www.ncbi.nlm.nih.gov/clinvar/variation/530914/). It is also now listed in two individuals in GnomAD, giving it an allele frequency of 0.0008%, still rare enough to be causative. In the case of BBS-016 and BBS-017, both the variants in *CEP164* and *INPP5E* merit further investigation and possibly Sanger sequencing. The *INPP5E* variant has now been reported in ExAC with a population frequency of 0.003%. Additional directions might include an array to identify possible deletions and, in the case of identified candidate genes, RT-PCR or antibody staining of proteins to look at expression. In addition, should pathogenic variants be identified, they would require confirmation in a certified diagnostic laboratory. In both cases, the patients and their parents would like a genetic result. In both cases, it has been important to highlight that only rarely does a genetic diagnosis lead to a treatment, and as there are no real treatments for ciliopathies at present, having a genetic diagnosis is unlikely to change anything except advice about recurrence risk both for the parents and for the patients when they come to start their own families.

Also to be considered are the duties to publish results and revisit WGS results to take account of future knowledge and developments. These patients are on a list of unsolved cases, and a departmental decision about revisiting results is required. This might be done in a number of ways such as

- Revisiting all unsolved cases at a fixed interval, such as three yearly, using the variants called during this analysis.
- Reanalysing all outstanding WGS cases whenever a variant calling pipeline is updated.

In this study a decision was made not to look for secondary variants of clinical significance. This is something that could be revisited and the patients reconsented for at a later date if appropriate. It is important to have a departmental policy for this, so that it can be discussed during the consent process. As all patients are analysed on a research basis, it is currently policy not to return secondary variants of clinical significance, but again, this is something that should be reconsidered regularly.

In terms of publishing results, neither case is at a position where it could be published as a case report and there is no departmental or institute policy for sharing variants found during WGS sequencing. One option is to submit variants to gene matcher, but this is something that can only be done with variants that are likely to be causative. This is something that requires further discussion so a consensus can be reached. A national database of genomic variants arising from the NHS genomics testing service, anonymised but accessible to clinicians and possibly researchers, would be very useful in reducing duplication of effort within an NHS setting.

The ACMG guidelines provide a good framework for determining pathogenicity, but result in many variants being called as VUS. As it is more likely that harm will be caused by calling a variant pathogenic when it is not, for example if prenatal testing was offered, this is a safe method. The UK Association for Clinical Genomic Science (ACGS) best practice guidelines for variant interpretation (www.acgs.uk.com) very helpful in clarifying the meaning of the terminology used, for example around missense constraint scores. However, it is likely that the ACMG guidelines will be updated as more is learned about variant interpretation.

Whole genome sequencing has both benefits and drawbacks in the diagnostic setting. The cost in terms of the test and the human and computational resources required should be offset against the benefit of having data that can be revisited as more is learned about the human genome, especially its non-coding regions, and having data that have utility outside of the diagnostic setting.

Chapter 4 Phenotyping in the HIGH-5 project

4.1 Introduction

The term phenotyping was first used by a Danish botanist, Willem Johannsen, who defined it as the way in which the hereditary dispositions of organisms or genotypes result in observable physical features or phenotypes(385). This definition described a long-standing concept in biology, going back to the time of Gregor Mendel, when physical characteristics of pea plants were used to understand the concept of heredity and modes of genetic inheritance. Phenotyping has been used throughout the history of medicine to stratify diseases and to learn their characteristics including causes, incubation times, treatments and outcomes. In this respect all doctors are phenotypers, though it is considered a more important skill in some specialties than in others. The delineation of dysmorphic features in clinical genetics is a vital diagnostic tool, while in cardiology, echocardiography has all but replaced clinical assessment of murmurs in the diagnosis of cardiac disease.

The definition of phenotype has broadened in recent years from a description of external physical features to encompass all measurable and observable ways in which an organism deviates from what is considered normal. This may include the measurement of biomarkers including gene expression, protein levels and function and more(386).

Accurate and consistent phenotyping is a major challenge. It is essential for diagnostics, where genotype-phenotype correlation is vital for identifying candidate genes from the many variants seen in exome or genome sequences where both pre- and post-test phenotyping may help to filter variants and identify pathogenic mutations, but also for the stratification of patients and the further delineation of the spectrum of genetic disease(387-390). Improved understanding of genotype-phenotype correlations facilitates better advice about risk and prognosis(391-393). A clinical service is important, not just for feeding back results, but also in determining their significance(388, 394, 395). Phenotyping is also essential in multiomics studies and trials of therapies, where it is important to be aware of the similarities and differences of groups of patients(396, 397). Genomics and multiomics are becoming widely if not routinely used, but there are ongoing issues with the interpretation of the data being generated. There are many reasons for this including a lack of control data, which will improve over time as multiomics studies become widespread. Multiomics is expensive and the cost of testing large control groups is prohibitive(398, 399). Another reason is that some samples or patients are difficult to obtain controls for, examples being paediatric patients, surgical samples and minority ethnic groups. A further issue is that as some of the technologies are newer, the degree of reproducibility between methods is still uncertain, an ongoing issue with much published research(400). All these make

the interpretation of omics data difficult. This is especially true in rare paediatric disease cohorts, more so when patients are recruited from a population with an international component. The cohorts are small, patients may have different ethnic backgrounds, samples may be unobtainable from controls and disease phenotype may be very variable(401). This gives rise to the “small sample size, large variability” problem(399, 402).

4.1.1 Deep Phenotyping

Deep phenotyping is the cornerstone of precision medicine and enables the use of small cohorts and datasets in research(403, 404). It involves going beyond the normal detail that is recorded about a patient and gathering together in an accessible form all the physical and clinical descriptors that are used to stratify patients. Peter Robinson describes it as “the precise and comprehensive analysis of phenotypic abnormalities in which the individual components of the phenotype are observed and described”(405). As a strategy, it has huge potential to increase our understanding of disease(406, 407). Patient stratification is also of enormous importance in understanding responsiveness and reactions of patients to novel and existing therapies(408, 409). When looking for biomarkers that predict severe disease or lack of responsiveness to therapy, failure is all but inevitable without accurate phenotypic data from test subjects. One of the current difficulties is that as yet there is no certainty about which data types will be most useful or how they are best recorded. There are many challenges in obtaining and recording these data, beyond the usual ones of consent and the difficulties of tracing medical records.

Firstly, a vast amount of clinical data are recorded about every patient, especially those with rare or severe diseases which may affect many organs and lead to the patient seeing many different specialists. They are recorded in many forms and places. As yet, the NHS does not have a robust system of universal electronic health records (EHR), also known as electronic medical records. Even in hospitals with EHR, clinicians often continue to handwrite notes, which are later scanned into the system, rendering electronic searches or data mining of records very difficult. EHR systems between hospital and other healthcare settings vary, and often cannot communicate with one another, leading to siloed, inaccessible information. Even with EHR, data mining is challenging(410). Deciding which data are relevant to a disease is difficult and may be beyond the scope of machine learning in its current state. Diagnoses may change as more information comes to light and not all diagnoses recorded in the notes will be relevant(411). Another issue is that data relevant to a single clinical episode or admission may be recorded in multiple locations. Blood results may be recorded in one database and may not be added to notes. Imaging reports are saved on another. Pharmacy records may be kept only on paper drug charts or e-prescribing systems and drugs that are not continued after discharge may never be recorded on letters sent to the GP or in the hospital notes. Even significant adverse drug effects may be inaccurately or unrecorded(412). Genetics notes and test results frequently are kept separate from hospital notes, both to enable family records to be kept together and to protect patient confidentiality(413). Again, this may result in important clinical information being unavailable or overlooked. The use of clinical personnel to extract data from records is expensive and very labour intensive, but at present may be the only way of ensuring that sufficient and accurate data are recorded for patients

in studies where deep phenotyping is important. In the longer term, however, the integration and extraction of phenotyping data with and from the EHR is likely to be more efficient and will allow patient phenotyping to be a dynamic, rather than a static “snapshot” process and allow it to be scalable for large projects(414-417).

4.1.2 Standardisation of phenotyping using ontologies and medical languages

Phenotyping of physical features is subjective and two clinicians may describe a patient differently, even at the same time point. In addition, phenotypic information may change over time, as additional symptoms develop or become more obvious, such as mild developmental delay that can later be described more accurately as a specific motor delay secondary to a syndrome such as muscular dystrophy. However, standardisation of data is important in the description and recording of phenotypes. For example, a hand malformation may describe anything from a vestigial additional digit to ectrodactyly or the “mitten hand” of Apert’s syndrome, and the use of such vague phenotypic terms results in both the loss of important clinical data that might be useful and the risk that patients may on paper look more or less similar to one another than they actually are. Without standards, such data risks being meaningless. This may be a reason that phenotypic data are often very scant in papers.

The Oxford English Dictionary defines ontology as being “the science or study of being; that branch of metaphysics concerned with the nature or essence of being or existence”. A phenotype ontology aims to link disease causes with syndromes through description of the features observed. Ontologies also enable cross-species comparison, which allows the identification of model organisms, vital in developing understanding of, and therapies for, human disease. The first biomedical ontology was the Gene Ontology (GO) and since then, various initiatives have attempted to aid with the standardisation of phenotypic data. Bodenreider et al. set out 7 qualities desired of an ontology in 2009(418). However, these were principles for medical ontologies in general, rather than those specifically for phenotyping, but are useful when considering the utility of a given ontology. They were as follows:

- No intellectual property
- Standardised, user-friendly format
- Existence of mapping to clinical terminology
- Harmonisation with other biological ontologies
- Regular maintenance
- Exhaustive coverage of diseases
- Support for automatic reasoning

4.1.2.1 The Human Phenotype Ontology

The Human Phenotype Ontology (HPO) came from an initiative to classify and link physical descriptors to aid accurate phenotyping and it is considered one of the most reliable ways of recording such data(419, 420). The HPO website defines it as “computational representation of a domain of knowledge based upon a controlled, standardised vocabulary for describing entities

and the semantic relationships between them" and the Human Phenotype Ontology as aiming "to provide a standardised vocabulary of phenotypic abnormalities encountered in human disease"(421). It fully or partially meets 6 of the 7 desirable qualities described above, and fails to score more highly because it restricts itself to phenotypes(418).

There are now over 13,000 HPO terms, including terms relevant to both rare and common disease. All the terms relate to abnormalities and describe deviation from an expected reference or normal phenotype. It is updated every three to six months and it is being translated into multiple languages(421). The terms are extracted from medical databases and literature and individuals can apply for additional terms to be included. Each term is given a unique number. In addition to phenotypic descriptors, terms describing inheritance, onset and course of disease are included(420). HPO terms are vital to the computational analysis of phenotype data, allowing clinical features to be grouped or separated. Without an ontology, a computer has no idea whether or how terms are related to one another.

The HPO has a branching structure of classes and subclasses related by the statement "y is a subclass of x"(422). At the root are general terms such as neurological abnormality (HP:0000707). Branching from this are subclasses which in the case of neurological abnormality would include abnormality of nervous system morphology (HP:0012639), abnormality of nervous system physiology (HP:0012638) and abnormality of the peripheral nervous system (HP:0410008). Each subclass has further subclasses until a phenotypic abnormality is classified to its fullest extent, for example subarachnoid haemorrhage (HP:0002138). One strength of the directed acyclic graph system utilised by the HPO system is that the subclasses (child terms) can be related to more than one class further up in the hierarchy (parent terms), which maximises the possibility of identifying the most accurate child term. The HPO numbers are linked to ensure that related terms are identified. For example unilateral post-axial polydactyly of the hand (HP:0001162) and ectrodactyly (HP:0001171) will both link to abnormality of the hand (HP:0001155). An example of HPO terms seen in the Bardet Biedl syndrome cohort can be seen in Figure 4.1.

The distance between HPO terms can be calculated computationally(423). The HPO covers seven different classes of phenotypic abnormality: morphological abnormalities, abnormal organ processes, abnormal cellular processes, abnormal behaviours, abnormal laboratory findings, abnormal electrophysiological findings and abnormal imaging findings(420). Terms are qualitative rather than quantitative, for example hypoglycaemia rather than a specific low blood glucose measurement, though sometimes qualifiers can be used to express degree of affectedness, such as mild, moderate, severe or profound learning difficulties.

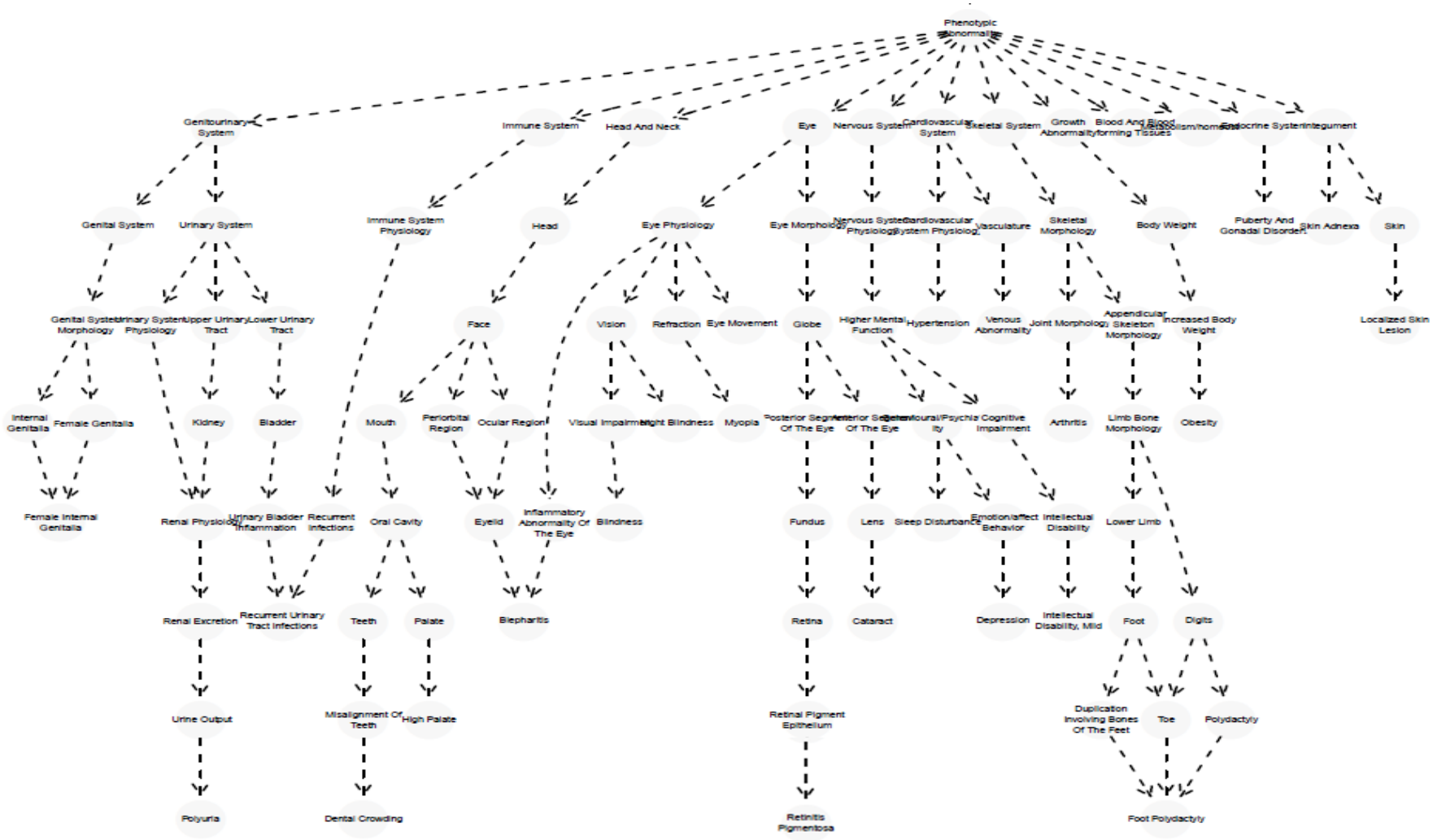

Figure 4.1 Directed acyclic graph of HPO terms appearing in BBS cohort in at least three patients. Diagram courtesy of Dr Tomi Peltola, Aalto University

Categories of age of onset can also be recorded. HPO terms can be used as a standalone tool, incorporated into software or databases or used in a pre-existing software such as PhenoTips®(424).

HPO terms are used by major research initiatives such as Orphanet, DECIPER and the 100,000 Genomes Project and widely used databases such as ClinVar(420, 425). Some journals, such as Cold Spring Harbour Molecular Case Reports are beginning to require authors to describe phenotypes using HPO terms which will facilitate searching medical literature for relevant cases(426). Currently the absence of HPO terms and the use of umbrella phenotypic or descriptive terms make it difficult to ascertain which studies are relevant to which patients and restrict the use of tools such as Matchmaker Exchange or PhenomeCentral(427, 428). The adoption of this standard by journals or its routine use in clinical practice will facilitate the use of phenotypic data(422).

4.1.2.2 SNOMED-CT terms

Systemized Nomenclature of Medicine (SNOMED-CT) terms arose from Systemized Nomenclature of Pathology (SNOP) terms, initially devised by the American College of Pathologists and later developed by a multinational collaboration of 27 countries including the UK, the International Health Terminology Standards Development Organisation (IHTSDO). They were developed to provide a comprehensive bank of standardised descriptors for clinical incidents and outcomes(429). For this reason, it covers a much wider range of topics than HPO terms, including medical interventions such as surgery and medicines, tissue types and procedures.

As of January 2018 it contained 341,105 terms, and is available in various languages. It is widely used in the NHS for clinical coding and its strength is that it is almost universal among EHRs, opening up the potential for communication between systems. Like HPO terms, its aim is to reduce variability in the way clinical incidents and outcomes are reported. It is free to use within member countries, but licences are required for non-members. It is described as having a logical multi-axial subtype hierarchy and is considered to be the most comprehensive group of clinical descriptors overall(430). It includes many clinical descriptors absent from the HPO, for example those related to infectious disease or cancer. However, analysis by the National Library of Medicine (NLM) showed that only approximately 30% of HPO classes were contained in SNOMED-CT, suggesting that its phenotypic descriptors are less detailed(431, 432). As with the HPO, it fully or partially meets 6 of the desirable qualities described by Bodenreider and Burgur.

4.1.2.3 Additional ontologies and medical languages

Various other ontologies exist, including the International Statistical Classification of Diseases and Related Health Problems (ICD)(433). However, none is as comprehensive as HPO for phenotyping.

4.1.3 HPO-based phenotyping and phenotype analysis tools

Currently much of phenotype annotation in both humans and animal models is manual and does not use formal ontologies. A way of overcoming this is text mining, which is difficult(434, 435). Using ontologies and storing data in a searchable and matchable format is more efficient and much more useful. Various phenotyping tools are available to enable this. They utilise ontologies such as the HPO and can be used to sort and sometimes store patient data, and identify syndromic conditions that may match the patient's phenotype. Another benefit of this and the codes used in ontologies is the ability to integrate phenotypic data with other data sets, such as genomics, transcriptomics and proteomics. Clinical interpretation is till of vital importance in interpreting the output of these programs(436).

4.1.3.1 PhenoTips®

PhenoTips® is a free, open source software developed in collaboration with the HPO group(424). Phenotypic features in the HPO can be selected from a list or searched for using free text. The software prompts for selection of the most specific annotation possible. It also prompts the user to consider terms that they have not selected and there is the option to note these, or any other terms, as not present. There is also the option to enter free-text non-HPO terms if a feature of note is not yet supported by HPO terms. An annotation sufficiency meter developed by the Monarch initiative shows how detailed and specific any given entry is(437). Features such as height, weight and head circumference are automatically plotted against the normal range for the patient's age, and if appropriate, terms such as microcephaly will be added to the list of phenotypic descriptors. The inputted HPO terms are then compared with the HPO terms associated with diagnoses, and a list of suggested diagnoses for the patient is generated. Data can be outputted in various formats for integration with other datasets or computational use. It can also be used as a simple database for storage of phenotypic information and there is space to record pedigrees, the outcome of tests done and whether a diagnosis has been identified. PhenoTips® is now being incorporated into other resources(438).

4.1.4 Data recording and safety

In addition to the HPO-based phenotyping terms, a wide range of clinical information may be desired for deep phenotyping. These may include results of laboratory or imaging tests, medications that the patient has been prescribed, information about family history and more. Different projects may require different information and formats, and a priority is being able to output these data for amalgamation with other data sets, for example omics data. Many of the available clinical databases, such as the well-known and widely used OpenClinica are very expensive and only partially customisable. One of the most flexible solutions is the use of a custom database, for example utilising the widely available Microsoft Access (MS Access) database and designing it to suit an individual project's needs.

4.1.4.1 MS Access database

Data was stored in an MS Access database as described in section 2.7.2.2. In MS Access, tables can be designed to restrict the types of data that can be inputted and users can be given varying

levels of access, for example, some can have read-only access, others can enter data or run queries and administrators can do all of the above and modify the database structure. Queries allow data from different tables to be aggregated and outputted in a form chosen by the user, allowing data to be restricted to specific cohorts, for example, patients with BBS, or subtypes of data, such as drugs prescribed. Databases can be outputted in several forms including text files, which are useful for integration of clinical data with other datasets.

4.1.4.2 University College London (UCL) Data Safe Haven (IDHS)

When phenotyping patients as part of a clinical project, sensitive patient data need to be recorded. At the time of the project, this was covered by the Data Protection Act (1998), but has now been replaced with the Data Protection Act (2018) in response to a new EU directive which came into force in May 2018, known as the General Data Protection Regulation (GDPR)(439, 440). These acts define such data as

- "the racial or ethnic origin of the data subject
- his political opinions
- his religious beliefs or other beliefs of a similar nature
- whether he is a member of a trade union
- his physical or mental health or condition
- his sexual life
- the commission or alleged commission by him of any offence
- any proceedings for any offence committed or alleged to have been committed by him"

The acts specify how these data must be processed, including the requirement for the data to be protected against "unauthorised or unlawful processing of personal data and against accidental loss, destruction of, or damage to data". The 2018 Act covers pseudanonymised data, unlike the 1998 Act.

Clinical projects that include phenotyping may include names, dates of birth and death, ethnicity and clinical data and so fall within the remit of the act. While data can be anonymised or pseudanonymised they still need to be stored securely where it would not be accessible to anyone except clinicians working on the project and would not be liable to loss or destruction. If pseudanonymised data are used, there must be no way in which an individual could be identified from their pseudonym. Alternatively, data may be recorded in a non-anonymised form and stored securely, and later anonymised if transfer of data is required for analysis. Anonymisation steps include the removal of names, dates of birth and unique identifiers. In extremely rare diseases it should be noted that patients may be identifiable from anonymised clinical data.

UCL provides access to the Safe Haven (IDHS) for researchers to securely store sensitive information as described in section 2.7.2(441).

4.1.5 Aims

This chapter aims to look at the requirements of a deep phenotyping database. It describes how a flexible deep phenotyping database was built, maintained and populated, and how data were outputted and utilised. In addition it looks at data protection requirements and how these were met.

4.2 Results

Two functional versions of the MS Access phenotyping database are provided in Supplementary Information S1.1 and S1.2 (CD-ROM). One is empty and one contains fully anonymised clinical data. Table 4.1 shows the types of patient data collected as part of the HIGH-5 deep Phenotyping project. Figure 4.2 is a schematic representation of the tables contained within the database that relate to patient investigations and their links to one another. The relationships between the tables can be seen in the relationships tab in the anonymised version of the database in the supplementary information. Anonymised summary HPO terms outputted from PhenoTips® are available in Supplementary Information S1.3 (CD-ROM).

4.2.1 MS Access phenotyping database

4.2.1.1 Location of MS Access phenotyping database

MS Access was available in the IDHS and so no installation or permission was required to use it. Only the researcher inputting data and the principal investigator (PI), both of whom were clinicians with honorary contracts at Great Ormond Street NHS Foundation Trust (GOSH), could access the data. The researcher had direct access, while the PI had the ability to request access in case of emergency.

4.2.1.2 Tables in MS Access phenotyping database

The MS Access phenotyping database consisted of static and dynamic tables, linked to one another. The links between the various tables are known as relationships and a selection of these is illustrated in Figure 4.2. Full details can be found in the anonymised database in supplementary information. There were 21 lookup tables in the database (Table 4.2). These were static tables containing the information required to fill the dynamic tables. For example, the lookup table “Site” contained all body parts where a test might be conducted or symptom might be experienced, such as brainstem, aorta, hand, while the lookup table “Units” contained all the units that a measurement or test result might be reported in, such as mmol/L or mg/kg. Other tables then had columns which could be filled by selecting from the list of entries in the lookup table. For example, the “Units” table is linked to the “Blood Markers” table so that each blood test type is linked to a specific unit, for example haemoglobin is linked to the unit g/L and alkaline phosphatase is linked to IU/L. These linked entities are then used in other tables such as “Blood Tests”, one of the dynamic tables, of which there are 15 (Table 4.3).

Data description	Data content	Number of entries
Patient information	Demographic data including name, date of birth, date of death, ethnicity, referrer	58
Diagnosis	Diagnosis, age and date of onset, family history	58
Timeline	Age of onset of each symptom (BBS only)	101
Hospital visits	Date of visit and whether the visit was associated with any tests, investigations or patient data	229
Blood test results	Date and type of test, result, whether result was high, low or normal and what the reference range was	6294
Radiology results	Date, site and type of test, result, whether normal or abnormal	43
Investigation results	Date and type of test, result, whether normal or abnormal for all investigations other than blood tests or radiology	49
Histology results	Date, site and results of histology tests and whether result was normal or abnormal	754
Treatments	Type and name of medical or surgical treatment, start and end date of treatment, dose and administration schedule	366
JDM-specific information	Antibodies/ scores in disability and quality of life scales/ symptoms experienced along with associated dates	17/ 172/ 5797
HPO terms	Physical descriptors of patient features	705

Table 4.1 Types and amount of phenotypic data

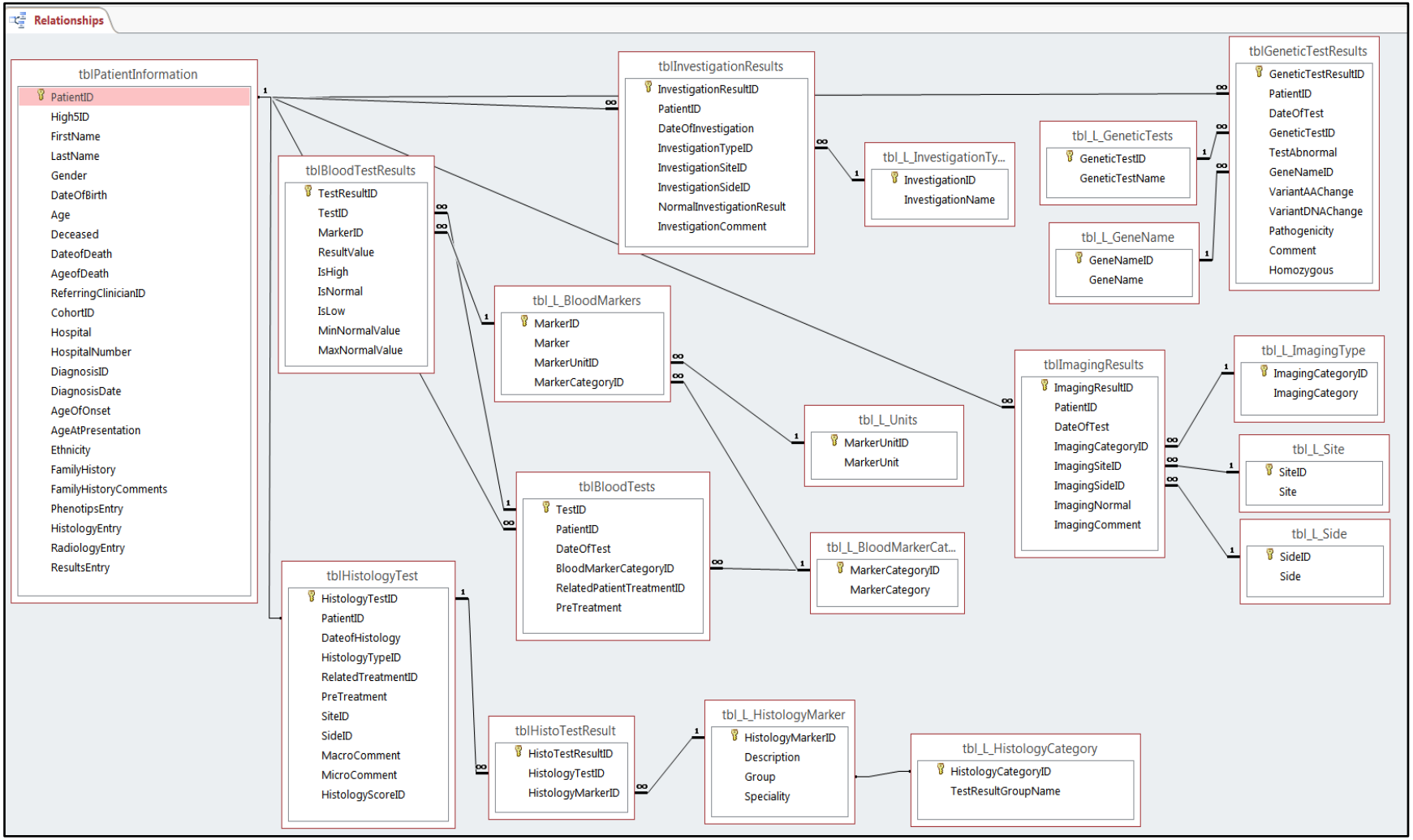

Figure 4.2 Investigation-related tables in MS Access phenotyping database and their relationships

Lookup Tables	Contents
Blood decision	Looks up formula for whether blood result is high, low or normal
Blood marker categories	Which category of test, e.g., immunology, haematology
Blood markers	Which exact test, e.g., CD8+ t cell count, haemoglobin
Cohort	Which of 7 disease cohorts the patient is in, e.g., BBS, IBD
Diagnosis	Patient's diagnosis, e.g., Usher syndrome type I
Gene name	Name of gene in diagnostic test, e.g., *BBS1, IL10RB*
Gene test	Type of gene test, e.g., single gene test, exome
Histology category	Macroscopic or microscopic
Histology marker	Result of pathology test, e.g., cryptitis, lymphoid hyperplasia
Histology specialty	Which organ system, e.g., gastrointestinal, brain
Histology scores	Specific histology scoring systems, e.g., grade of malignancy
Imaging type	Type of radiological investigation, e.g., CT, MRI
Investigation type	Type of non-radiological investigation, e.g., echocardiogram, endoscopy
HPO terms	HPO terms by number (JDM cohort only)
Referring clinician	Name of clinician in charge of patient care
Side	Left, right, left and right, unknown, not applicable
Site	Location, e.g., kidney, spine
Timeline of events	Dates that disease features developed (BBS cohort only)
Treatment name	Treatments by name, e.g., atenolol, hemicolectomy
Treatment type	Treatments by type, e.g., antihypertensives, surgical
Units	Standard units of measurement, e.g., mg/kg, mmHg

Table 4.2 Lookup (static) tables in the HIGH-5 deep phenotyping project

Each line in the table "Blood Tests" contains information about which test was performed on whom and on which date, and is illustrated in Figure 4.3. The information from this table is utilised in the table "Blood Test Results". This table, illustrated in Figure 4.4, contains the patient's name in column two (hidden in figure). This column is a link to a specific line in the "Blood Tests" table and indicates which test which patient had on which date. Column three contains the name of the test, column four the result, which is linked to the table "Blood Markers" which also includes data about the units for that result. Columns five, six and seven are optional tick boxes. These do not need to be filled if the reference ranges are known. Columns eight and nine are the lower and upper limits of the reference range for a patient of that age and gender. Column ten indicates whether the result is normal, low or high (section 2.7.2.2).

Data Tables	Contents
Blood Pressure Results	Blood pressure results by date
Blood Tests	Date and type of test, e.g., haemoglobin
Blood Test Results	Results of blood tests listed in Blood Tests table
Genetic Test Results	Genetic test results by date
Histology Tests	Date and type of test, e.g., gastrointestinal microscopic
Histology Test Results	Results of histology tests listed in Histology Tests table
Imaging Results	Radiology results by date
Investigation Results	Non-radiological investigation results by date
JDM Autoantibody Results	Which autoantibodies patients from JDM cohort had tested positive for i.e. which subtype of disease they had
Scores (Disease-specific)	JDM-specific scoring system results
Symptoms (Disease-specific)	JDM-specific symptoms present by date
Patient Information	Demographics of patient, e.g., name, date of birth, ethnicity
Treatments Received	Treatments received by date and with dose if applicable
Vision	Results of optometry assessments
Visits	Appointments or admissions associated with test results

Table 4.3 Data (dynamic) tables in the HIGH-5 deep phenotyping project

tblBloodTests

TestID	DateOfTest	Blood Marke	PreTreatmen
1	28/10/2013	Haematology	
2	28/10/2013	Coagulation	
3	28/10/2013	Inflammation	
4	28/10/2013	Biochemistry	
5	28/10/2013	Renal	
6	28/10/2013	Liver	
7	28/10/2013	Vitamins	
8	29/10/2013	Other	
9	22/01/2014	Haematology	
10	22/07/2014	Biochemistry	
11	22/07/2014	Renal	
12	22/07/2014	Liver	
13	22/07/2014	Inflammation	

Figure 4.3 Sample from Blood Tests table. The patient's name in the second column has been hidden. The date and type of test are listed in columns three and four

tblBloodTestResults

Test Result ID	Marker ID	Result Value	High Result	Normal Resu	Low Result	Lower Limit of No	Upper Limit of	BloodDecisic
1	Haemoglobin	109	☐	☑	☐	100	135	1
2	Red Blood Cell Count	3.88	☐	☑	☐	3.7	5.3	1
3	Haematocrit	0.31	☐	☐	☑	0.33	0.39	2
4	Mean Cell Volume	81.2	☐	☑	☐	70	86	1
5	Mean cell Heamoglobin	28.1	☐	☑	☐	23	31	1
6	Mean Corpuscular Haem	346	☐	☑	☐	315	355	1
7	Red Cell Distribution Wid	14.5	☐	☑	☐	11	16	1
8	Platelets	272	☐	☑	☐	150	400	1
9	White Cell Count	9.11	☐	☑	☐	5	15	1
10	Neutrophils	6.17	☐	☑	☐	1	8.5	1
11	Lymphocytes	1.84	☐	☐	☑	3	13.5	2
12	Monocytes	1	☐	☑	☐	0.3	1.5	1
13	Eosinophils	0.08	☐	☐	☑	0.1	0.3	2
14	Basophils	0.02	☐	☑	☐	0	0.2	1
15	Erythrocyte Sedimentatic	95	☑	☐	☐	3	13	3
16	C-reactive Protein (CRP)	29	☑	☐	☐	0	5	3
17	Sodium	143	☐	☑	☐	133	146	1
18	Potassium	3.5	☐	☑	☐	3.5	5.5	1
19	Urea	3.9	☐	☑	☐	2.5	6	1
20	Creatinine	23	☐	☑	☐	20	43	1

Figure 4.4 Sample from Blood Test Results table. The patient name is normally visible in the second column (patient ID) but has been hidden in this example. The patient ID column links to the Blood Tests table which contains details of what test was done on which patient on what date. The blood decision score column gives a result of 1 for normal, 2 for a result below and 3 for above the reference range

4.2.1.3 Data recording in MS Access phenotyping database

All data were recorded as a number, even if they were displayed as text. Each entry in a table is automatically assigned a number. For example in the Blood Marker table which lists all the different things that can be measured by a blood test, such as haemoglobin or IgE levels, each item was allocated a unique number. Therefore when the test word haemoglobin was entered into the database, it was in fact recorded as the number 5, meaning that when the data were outputted for integration with other datasets, it was useable both by humans and computers. This was the case for all unique entries, such as cohort, patient, imaging investigation etc. For ease of use, the numerical values were hidden when entering information into the database or when using the database to look up patient information. Data entry was made simpler by the creation of forms which laid out the field for completion in an intuitive, convenient manner and prepopulated any sections that could be prepopulated to minimise duplication of effort.

4.2.1.4 Queries in MS Access phenotyping database

Subsets of data can be combined from multiple tables and extracted from the database by running queries. Queries can be customised to show any combination of table entries, for example haemoglobin levels for all patients, all blood tests done on a particular date or all data gathered on a single patient. It is a useful way to anonymise data also, in that results can be outputted using another identifier. An example of selected anonymised data is shown in Figure 4.5. Queries can be saved as can their results, meaning that they can be run repeatedly as data are entered or adjusted to look at different metrics.

4.2.1.5 Exporting data from MS Access phenotyping database

Fully anonymised data were exported from the HIGH-5 phenotyping database at various points during the study, for example for researchers looking at various omics datasets such as RNAseq, genomics and proteomics. This was done in various forms such as Excel files and tab and comma-delimitated text files. Drug prescription information was used for pharmacogenomics analysis (Chapter 5). The structure of the database was used to form the basis of the clinical data part of Sapientia™ for Omics, a product being developed by Congenica UK (www.congenica.co.uk) as part of an Innovate UK grant, and a mock dataset entered into the phenotyping database was used to test anonymisation and data extraction by the program.

4.2.2 PhenoTips® database

A summary of entries and the patients they applied to can be found in Supplementary Information S1.3 (CD-ROM). There were 705 entries with 272 unique terms. Most were HPO terms, but occasionally free text entry was used.

x_BBSBloodResults

High5ID	DateOfTest	MarkerCategory	Marker	MarkerUnit	ResultValue	IsHigh	MinNormalValue	MaxNormalValue
BBS-014	07/05/2002	Haematology	Neutrophils	x10*9/L	8.2	-1	1.5	7
BBS-014	07/05/2002	Renal	Urate	µmol/L	0.55	-1	0.18	0.35
BBS-014	07/05/2002	Renal	Urea	mmol/L	10.4	-1	2.5	7.5
BBS-014	07/05/2002	Renal	Creatinine	µmol/L	141	-1	65	101
BBS-014	05/11/2002	Renal	Urea	mmol/L	9.5	-1	2.5	7.5
BBS-014	05/11/2002	Renal	Creatinine	µmol/L	145	-1	65	101
BBS-014	05/11/2002	Liver	Alkaline Phosphatase	IU/L	120	-1	31	116
BBS-014	02/05/2008	Haematology	White Cell Count	x10*9/L	11.2	-1	4	11
BBS-014	02/05/2008	Haematology	Neutrophils	x10*9/L	8.1	-1	1.5	7
BBS-014	02/05/2008	Renal	Urate	µmol/L	0.44	-1	0.14	0.34
BBS-014	02/05/2008	Renal	Urea	mmol/L	10.3	-1	1.7	8.3
BBS-014	02/05/2008	Renal	Creatinine	µmol/L	156	-1	45	84
BBS-014	02/05/2008	Lipids	Cholesterol	mmol/L	5.4	-1	0	4
BBS-014	11/05/2010	Biochemistry	White Cell Count	x10*9/L	11.2	-1	4	11
BBS-014	11/05/2010	Renal	Urea	mmol/L	12.3	-1	2.5	7.5
BBS-014	11/05/2010	Renal	Creatinine	µmol/L	194	-1	65	101
BBS-014	11/05/2010	Lipids	Cholesterol	mmol/L	5.1	-1	0	4
BBS-014	24/10/2012	Haematology	Mean cell Heamoglobin	p/g	33.5	-1	27	32
BBS-014	24/10/2012	Haematology	Neutrophils	x10*9/L	7.8	-1	1.5	7
BBS-014	24/10/2012	Renal	Creatinine	µmol/L	385	-1	45	84
BBS-014	24/10/2012	Liver	Alkaline Phosphatase	IU/L	130	-1	35	129
BBS-014	24/10/2012	Lipids	Cholesterol	mmol/L	5	-1	0	4
BBS-014	24/10/2012	Lipids	LDL Cholesterol	mmol/L	3.27	-1	0	3
BBS-014	24/10/2012	Lipids	Triglycerides	mmol/L	1.83	-1	0	1.7
BBS-014	24/10/2012	Inflammation	C-reactive Protein (CRP)	mg/L	19	-1	0	5
BBS-014	24/10/2012	Endocrine	Cortisol	nmol/L	654	-1		
BBS-014	24/10/2012	Endocrine	Testosterone	nmol/L	2.2	-1	0	1.8
BBS-014	24/10/2012	Urine	urine albumin:creatinine ra		16.6	-1	0	2.4
BBS-014	25/06/2014	Biochemistry	Creatinine	µmol/L	128	-1	45	84

Figure 4.5 Output of query- Blood results of patient BBS-014 with values above the upper limit of normal, some columns hidden for clarity

4.2.2.1 Location of PhenoTips® database

PhenoTips® was installed in the IDHS so that demographic information could be safely stored. Arrangements were made to update the list of HPO terms as automatic updates cannot take place in the IDHS.

4.2.2.2 Storage of data in PhenoTips® database

Each patient had an individual entry in PhenoTips® which included demographic data, metrics such as weight and height, pedigrees if relevant family history was present, a clinical diagnosis and a list of HPO terms.

4.2.2.3 Output of data from PhenoTips®

Fully anonymised PhenoTips® data were outputted at various points as both XML and JSON files for researchers working on multiomics datasets (section 4.2.1.4).

4.3 Discussion

One of the aims of the High-5 multiomics project was the integration of deep phenotyping data with the other multiomics datasets including genomics, transcriptomics and proteomics with a view to using phenotype to inform analysis of the multiomics datasets. It was clear at the outset that clinical data should be available to inform analyses, but less clear which data would be useful or how it should be collected and stored.

4.3.1 Choice of databases

4.3.1.1 MS Access database

The main requirements for the database were determined to be the following:

- The ability to accept all types of data relevant to the study
- The ability to modify the database to accept data types that had not been anticipated when the database was built or set up
- The ability to output data anonymously
- The ability to output data in a manner that would make it easy for computational integration with other datasets
- Low cost
- Be accepted for addition to the IDHS

When the options for databases were considered, many of the commercial solutions, such as OpenClinica, were not set up to record all data types required. In addition they could not be modified by researchers, but instead would require a request to the developer to amend them. Furthermore, they would have to be loaded onto the IDHS and would not be able to communicate externally. The cost of these commercially available solutions was very high. OpenClinica quoted US $29,000 for a two year study with ongoing costs of $12,000 per year.

The only database software available on IDHS was MS Access. The decision to use this meant that the database would not have any costs in addition to researcher's time and would be fully customisable. The database would record data both numerically and as text, which was important for when data were being outputted for computational integration with other data sets. It also met the criteria listed above. Three researchers were trained in the building and maintenance of the database, so that if the primary researcher left the organisation the database could be maintained and modified by someone else. An individual with expertise within the organisation was also identified in case problems were encountered. The database was mapped out on paper in the first instance to plan the relationships, but it was continually modified during building and use, to ensure it could capture all data types.

One issue that was encountered during the planning stages was the extremely large number of HPO terms. This would have made for unwieldly dropdown lists and also would have made it hard to search for the most accurate HPO term. A decision was made that the HPO terms for phenotyping should be kept separately in a specific phenotyping database to maximise phenotyping accuracy.

4.3.1.2 PhenoTips® database

When the options for specific phenotyping databases were considered, the PhenoTips® software was the preferred choice. It was suitable for installation in IDHS, it was free and required no training to use. Data could be outputted as a text file. Although the HPO terms were recorded as text, each was linked to a unique HPO number, which was then suitable for computational integration with other datasets. In addition, PhenoTips® encourages users to use the most accurate term by suggesting HPO terms and it allows the marking of HPO terms as "not present". There is a difference between "not present" and "not looked for or commented upon" when phenotyping, so the ability to confirm the absence of a specific feature is helpful. It prompts users to consider other HPO terms that may be relevant, which may or may not be considered advantageous.

A disadvantage of using PhenoTips® was the fact that HPO terms had to be installed separately in the IDHS rather than the software looking them up from a central, regularly updated list. This meant that updates to the list of HPO terms would had to be installed manually. An agreement was reached that PhenoTips® updates would be installed at 6-monthly intervals by IDHS staff. It was also agreed that any researcher wishing to use PhenoTips® would have a standalone version associated with their IDHS login, so that data could not be seen except by the researcher themselves, the PI who could access the data if they needed to, and any authorised individual the researcher chose to share the data with.

4.3.2 Data types for collection

In the early stages of the project it became clear that there was little or no consensus about what data types would be useful at the analysis stage, except that HPO terms for physical characteristics would be essential in order to be able to stratify patients. As there were multiple

cohorts, some very heterogeneous, there was also the difficulty that different types of data would be available for each cohort. One strategy was to collect all the data available for each patient. This was quickly ruled out, as all data had to be extracted from the medical record manually and manually re-entered into the phenotyping database, and a single hospital admission could result in hundreds, or even thousands, of blood tests. Following discussion with the PIs who had oversight of each cohort and the scientists who would analyse the data, a strategy was defined.

4.3.2.1 Demographic and cohort data

A minimum set of clinical and demographic data were recorded for all patients. This included name, date of birth, hospital number, date of death if applicable, ethnicity, cohort, diagnosis, age at and date of onset and diagnosis, treating clinician, treating hospital and whether or not there was a family history. These data were felt to be important as some, such as ethnicity and age, would be necessary for stratification and omics results interpretation. Having names and dates of birth reduced the possibility of data being recorded for the wrong patient as each patient had a name, date of birth and hospital number as well as a unique HIGH-5 study number to reference. The unique HIGH-5 number consisted of the 3 letter cohort code (BBS, IBD, JDM, SRS, USH) and then the number order in which they were enrolled, e.g., BBS-001, BBS-002, etc.

4.3.2.2 Physical descriptors as HPO terms

All medical letters were read to extract physical descriptions of patients. These were then entered into PhenoTips® using the most appropriate HPO term. This was determined by following the branching structure of HPO terms to the most descriptive one that could be applied to the patient. If further physical descriptors were identified, for example from imaging or other investigation results, these were entered as they were encountered. If a physical descriptor later resolved, a note was made of this, but the HPO term was not removed. However, if a physical descriptor was later determined to be incorrect it was removed or amended.

4.3.2.3 Blood results

Initially all available blood test results were recorded. However, it quickly became clear that this was both enormously labour intensive, but that also the data were very difficult to use if they could not be linked either to a multiomics sample or other clinical report or outcome. Instead, blood results which could be linked to other clinical data were recorded, such as those taken on the same date as a clinic visit or an investigation. In order to determine whether values were within normal limits, reference ranges were recorded also, as these varied, especially among the paediatric cohort or between two hospitals. For example, many of the haematology indices have variable reference ranges, especially under the age of one. It was considered whether this could be done automatically by the database, but owing to blood tests having been carried out at a number of sites, and on occasion the reference range changing at a single site due to a new testing method, this was not possible. It could certainly be done if all patients had blood tests at a single site. The reference ranges could be stored as a lookup table, and added according to age at time of test. This would reduce some of the effort involved in recording blood results. Abnormal results were flagged in a series of tick boxes as described in section 2.7.2.2. Having

both the numerical and tick box options was important, the tick boxes because they made the database easy to use to scan for abnormal results by eye, and the numbers because they could be outputted and combined with other data sets by computer. Blood results were linked to other investigation results both by date and by the “Visits” table in cases where bloods or other results were dated differently from that of an investigation or clinic visit. This was of particular importance in the case of histopathology results which would often be dated several days after the investigation date. In addition, blood samples which were related to both the investigation and histopathology results were often done on the day before the investigation.

4.3.2.4 Other results

All histopathology results were recorded in the database, along with any associated bloods and clinical information. Along with the date and site of each test, there was a tick box to record whether the histopathology was taken pre- or post-treatment. Histopathology results were divided into macroscopic, i.e. appearance of the specimen to the naked eye, and microscopic, i.e. appearance post-sectioning, staining and microscopy. If the specimen was normal it was recorded as “normal microscopic” or “normal macroscopic”. If it was abnormal, the specific features, such as cryptitis or necrosis, were recorded.

All imaging and investigation results were also recorded and linked to blood results where possible. Imaging was described as normal or abnormal. Abnormal results were entered as HPO terms or as free text in the PhenoTips® database.

4.3.2.5 Additional clinical information

Treatments that the patients were receiving were recorded, along with dates started and stopped, dose, route and reason if available. Other cohort-specific data sets were also recorded, such as a timeline for the onset of symptoms for BBS patients or symptom severity scores for JDM patients. Treatments proved important for pharmacogenomic studies (see Chapter 5).

4.3.3 Sources and format of data before input

The data for input into the two phenotyping databases were in a variety of locations and formats. Many of the data were stored at GOSH in various databases that did not communicate with one another and had no option for outputting data apart from printing it onto paper. Pathology results which included blood and histopathology results were on one database, genetics results were on a second, imaging results were on a third, and so on. Clinical letters were stored in paper folders or scanned and stored on an electronic database. In some cases, such as the patients in the BBS and USH cohorts, paper notes were stored at an entirely unconnected hospital with pathology and imaging results being stored on the databases specific to that hospital. For some cohorts, little or no clinical data were available, such as the SRS cohort, where the patients were from many years previously and the majority of the cohort consisted of their parents, for whom no information whatsoever was available. In some cases such as the JDM and IBD cohorts, some data had been collected by researchers and was in the form of MS Excel spreadsheets. This was supplemented with information from GOSH databases and electronic health records (EHR). As

none of the data sources were linked with one another or the IDHS, there was no part of this process that could be automated, meaning that all data required manual inputting.

4.3.4 Standardisation of data

As the data were held in many formats and locations it was important to standardise them. During the database design, it was determined which bits of data were important to hold, and this was built into the structure of the database. Then all results were recorded in the same format. If specific pieces of information were unavailable, then the field was left blank. For example, for every blood test the following was recorded: name of patient (linked to all other demographic data), date of test, category of test e.g. haematology, biochemistry, specific test, for example, haemoglobin or albumin, result, units, lower and upper limits of normal for age/gender/hospital and whether the result was above, below or within the normal range. These pieces of information were stored in six different tables; Patient Information, Blood Marker Category, Blood Marker, Blood Tests, Blood Test Results and Blood Decision. All of this information was available for the majority of tests on the majority of patients, but tests for JDM patients were recorded without reference ranges in an Excel spreadsheet, so in these cases the result was recorded as high, low or normal.

4.3.5 Backup of data

Initially there was no plan for backing up the data outside of the IDHS. This was for several reasons. Firstly, the IDHS is the only approved way for storing non-anonymised clinical data in UCL. Secondly, transfer of data outside the IDHS is a complicated, multistep process and therefore time-consuming when only small adjustments were being made. Thirdly, it is not considered good practice to have multiple copies of a database, as it is easy to amend one copy and not the others. Fourthly, when the project was started, the IDHS backup was considered extremely secure. There were onsite and offsite servers on which duplicate copies of IDHS were stored in case of fire or other damage. All the files were backed up with hourly versions being saved. This meant that theoretically, if data were lost from an iteration of the database, it would be possible to view the backup file from an hour before the loss, minimising the chance of permanent data loss. These would not be overwritten for a period of 90 days. Therefore it would be possible to go back to a much earlier version if necessary.

However, there was an unexpected network outage of one IDHS server when the database was being used. No explanation was found, but when it was restored, the database was corrupted and could not be opened. When IDHS staff attempted to retrieve an earlier version, they discovered that the IDHS system was overwriting the files rather than saving multiple iterations and no backup file was available. This was the case for many files within IDHS and so it highlighted a major weakness of the system and meant that the ISO27001 standards had been breached. This was reported to the required authorities by UCL. This meant that the database had to be almost entirely rebuilt as only an early prototype remained. In addition, all the clinical data that had been entered had to be re-entered. In total, this amounted to several months of work.

After this incident, despite the rectifying of the backup issue in IDHS, it was determined that a second copy of all data had to be kept outside the IDHS. A UCL- and NHS- approved flash drive with military grade XTS-AES 256-bit hardware encryption was chosen. Accessing the flash drive required knowledge of a 7 to 15 digit pin, which was committed to memory. Removing the flash drive from the computer automatically locked it and it also automatically locked if it was not used for a specified amount of time. The drive was kept in a locked drawer which could only be accessed by the researcher using the database and was kept separately from the digital secure key required to access IDHS. This was agreed with UCL as in keeping with legal requirements for information governance. The PhenoTips® database, which was a piece of software installed onto IDHS could not be saved directly onto the flash drive. At no times were there any problems with the IDHS version of PhenoTips® and backups appeared to be working, the data were outputted as an Excel file and saved on the secure flash drive after any amendment. This meant that, should the software be lost or corrupted, a copy of the data would be available.

4.3.6 Output of data

This is described in section 2.7. All data were fully anonymised before transfer out of the IDHS.

4.3.7 Uses of data

The data from the phenotyping databases were used in a variety of ways to enhance the analysis of the HIGH-5 project.

4.3.7.1 Use of phenotypic information for burden analysis

Information held about the BBS and IBD cohorts was extracted from the database and used to stratify patients for analysis of omics data. The BBS patients were stratified in a number of ways, for example by clinical features for genomic burden analysis. Burden analysis looks at the overall number of variants in genes that may be related to a disease or family of diseases. There is now evidence to suggest that the total “burden” of these rare variants may contribute to phenotypic variability as well as the overall likelihood of developing a disease(442, 443). This has long been understood in cancer, but is a more recent discovery in Mendelian disease(444, 445). Oligogenic cases have been reported in BBS and even in monogenic cases it has long been realised that the phenotypic variation seen cannot be accounted for by causative mutations alone(183, 389, 446, 447). Multiple ciliopathy modifiers have been identified(446). The BBSome, which is formed by many of the proteins produced by BBS genes and the close interaction with other ciliary proteins provide a model for how this burden effect might work. The effect of reduced levels of a protein might be exacerbated or alleviated by altered interaction with other proteins(442, 447).

There are many strategies for burden testing but the simplest is to simply count the variants in genes of interest - for example in the case of BBS, all known BBS or known ciliopathy genes - and to compare the number per patient with those in controls(442, 448, 449). Various modifications have been proposed, including looking at the number of variants in each individual patient, variants in each gene of a set, weighting variants according to function or frequency. Another method is to look at extremes of phenotype, a strategy successful employed elsewhere

in the study of rare disease(401, 450). The data held in the phenotypic databases were used both to identify patients with extremes of phenotype, for example, patients with 4 limb polydactyly versus patients without, and to identify patients with early versus late onset phenotypes. The BBS burden analysis required mainly physical features described by HPO terms but also some blood results such as cholesterol and other clinical metrics such as blood pressure measurements and weights.

A similar strategy was pursued with the IBD cohort, where patients were stratified according to age of onset, treatment resistance, severity of features and more. This required access to and manipulation of many categories of clinical data including histology and blood results, for example to identify treatment response. These data were submitted in other PhD theses (Dr Rosalind Davies, Dr Jochen Kammermeier). Work is ongoing on the JDM cohort. It was not possible to do this for the SRS cohort as detailed clinical data were not available for these patients and so they were not included in the database. The patient numbers were too small and the phenotypic features too diverse in the USH cohort. These analyses would not have been possible without phenotypic data that was easy to extract and manipulate, as it would have been extremely time consuming to extract and order such data. In addition the researchers doing such analyses did not always have access to clinical notes as they were not clinicians, and therefore the anonymised outputs meant that they could access data that would otherwise have been unavailable to them.

4.3.7.2 Use of phenotypic information to explore omics results post-analysis

Initially, omics analysis was carried out without researchers being informed of any clinical information. Once this had been done, phenotypic and other clinical information were used to aid understanding of the results. For example in the IBD cohort, there was a consistent outlier during analysis. Exploration of the phenotypic information allowed understanding of this. The patient had characteristics that made them significantly different to the other patients. Firstly, they were the only cohort member to have received a cord blood transplant and one of only three to have required a total colectomy for treatment of disease. In addition, the patient appeared to be more unwell than most in the cohort from the time they were enrolled. These data allowed understanding of the omics results in this case.

In addition, clinical data were used to check clustering post analysis, to rule out other causes such as age, gender or ethnicity in causing clustering. These data have been submitted as part of a PhD thesis (Dr Jochen Kammermeier).

4.3.7.3 Use of phenotypic information to stratify patients for omics analysis

In addition to using the clinical data to explore omics results post analysis, the omics results were also analysed, with patients having been pre-stratified into various groups following discussion with the lead clinician for each cohort. For example, in the case of the IBD patients, they were stratified into infantile and non-infantile onset, into treatment responders and non-responders, into groups depending on site of histological abnormalities, into patients responding to particular

treatments and other groups. Work is ongoing in the JDM group, but the work done on the BBS and IBD cohorts has been submitted as part of other PhD theses as above.

4.3.7.4 Use of phenotypic information to identify patient characteristics for diagnostics

Several patients without genetic diagnoses were included in the BBS cohort (BBS-016, BBS-017 and BBS-018). Clinical information was used to explore possible pathogenic variants in these patients, both in determining which genes would be looked at during the variant filtering stages and later in evaluating candidate genes that were not known to cause BBS. A detailed discussion of this and clinical details of the patients can be found in section 3.4.

4.3.7.5 Use of phenotypic information for pharmacogenomics analysis

Pharmacogenomics analysis was carried out as part of the HIGH-5 project. Information about which drugs patients were taking was utilised in this analysis, along with data such as age and ethnicity as well as whether patients had already been prescribed relevant drugs. Again the ability to extract this quickly rather than looking through paper notes or electronic prescribing records was vital, as otherwise the analysis would have been too time consuming to justify. A detailed discussion of this can be found in Chapter 5.

4.3.7.6 Use of phenotypic data and databases in developing a multiomics tool

As part of an Innovate UK award, HIGH-5 researchers collaborated with Congenica UK in the development of Sapientia™ for Omics, an add-on to the Sapientia™ genome analysis tool, which was developed for the analysis of exome and genome data. It is expected that Sapientia™ for Omics will allow the analysis and integration of multiomics data including genomics, proteomics and transcriptomics. It will also attempt to integrate clinical data including HPO terms with multiomics data. Phenotyping databases were used in two ways in the development of this software.

Firstly, the Sapientia™ for Omics software aims to be able to auto-populate the clinical information from spreadsheets or databases including an anonymisation step. In order to test this, a mock version of the phenotyping database was provided, both as SQL database (converted from the original MS Access format) and a series of Excel spreadsheets. They contained clinical details for patients such as Harry Potter, Hercule Poirot and Anne Shirley. They had fictitious dates of birth and clinical information. In this way, the ability of the Sapientia™ for Omics to anonymise clinical information could be confirmed. A variety of names was used including rare and non-British names to ensure that it would cope with the range of patient names it might encounter. Secondly, subsets of both omics data and fully anonymised clinical data were provided for analysis and the results compared with existing omics analysis as part of the development process.

4.3.7.7 Use of phenotypic data for machine learning

Collaborators at the Department of Computer Science, University of Aalto, Finland utilised anonymised clinical data to see if machine learning could help with omics analysis. They were

provided with metrics such as HPO terms and other clinical data such as blood pressure. The results of this have not yet been published, though some preliminary work has included mapping HPO terms for a cohort and calculating the similarities of patients to one another, work that will be helpful in ongoing stratification of the cohorts. Figures 4.6 and 4.7 illustrate some of the preliminary work done by Dr Tomi Peltola.

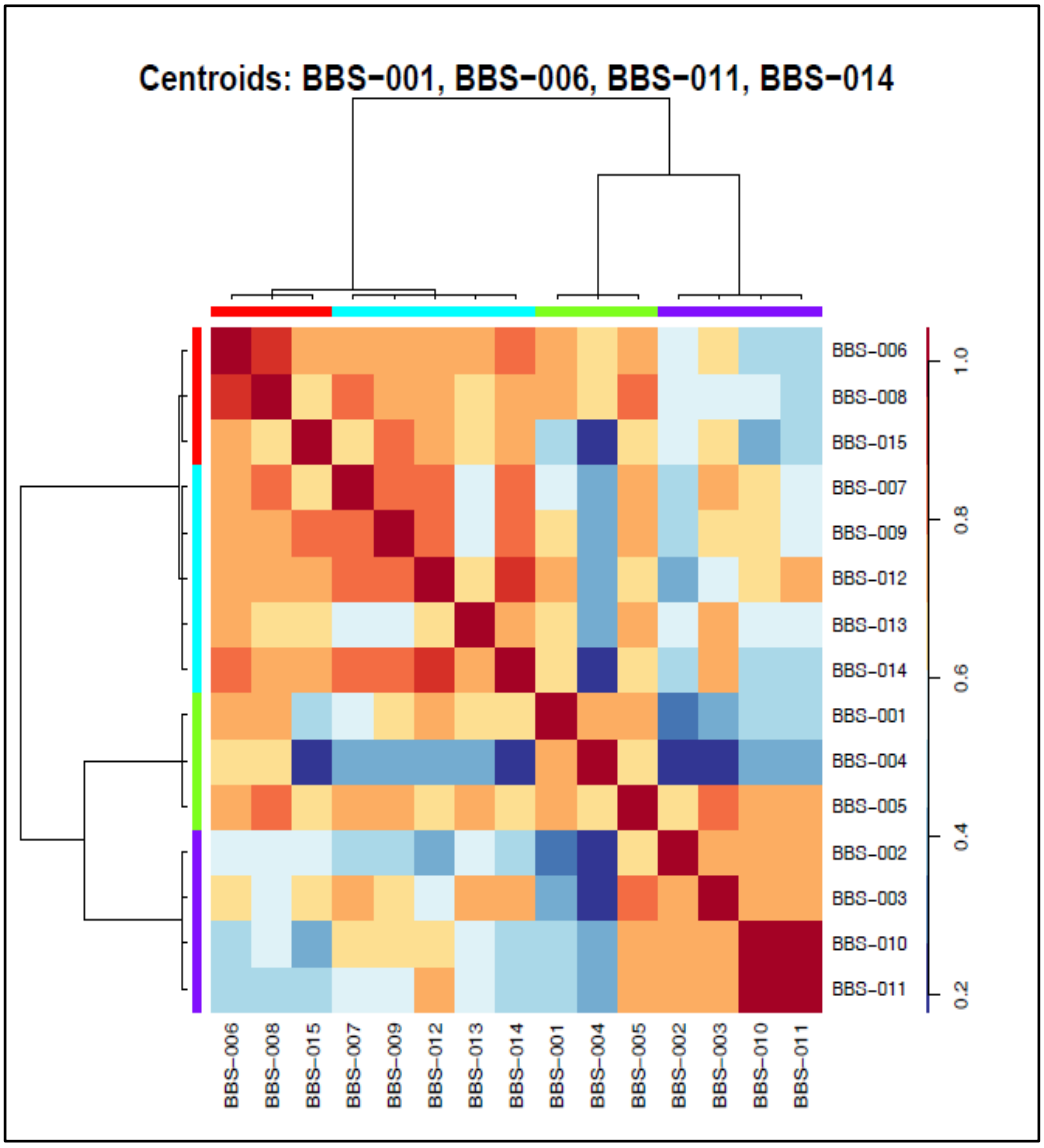

Figure 4.6 Heat map showing HPO term similarity amongst patients of BBS cohort. Courtesy of Dr Tomi Peltola, University of Aalto, Finland. Red represents similarity, blue represents dissimilarity. The darker the colour the more similar or dissimilar the patients are

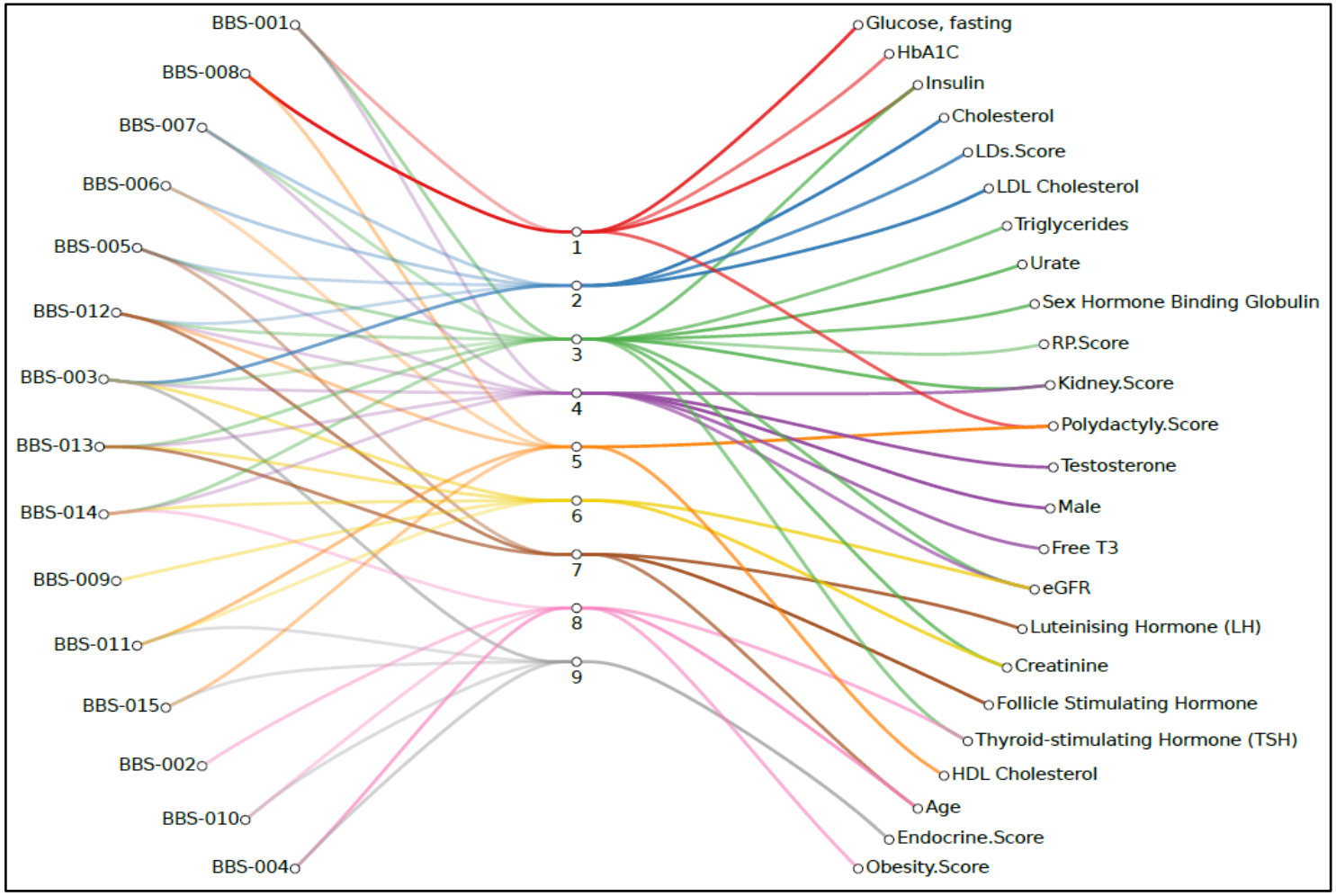

Figure 4.7 Eye diagram showing links between BBS patients and phenotypic data. Courtesy of Dr Tomi Peltola, University of Aalto, Finland

4.3.8 Utility of databases

Data collected in the databases was used and continues to be used in a variety of ways. However, researchers working on the multiomics datasets found the clinical information difficult to integrate with the omics data. This may have been due to several reasons, including lack of experience and difficulty with access to computing facilities, which was a problem in the earlier stages of analysis. Much of the published omics data has minimal integration of clinical data. Patients tend to be stratified into cohorts for omics studies and clinical data may not be considered much further. This is an area which is being explored further by the HIGH-5 project and others(8). However, the clinical data were useful in interpretation of multiomics analysis and in patient stratification for burden and other analyses. It was also important in pharmacogenomics analysis, as data such as age, ethnicity and prescribed medications were easily accessible and available to inform analysis. It was clear that the HPO terms were more useful to researchers than the more in-depth clinical information such as bloods and other metrics, but both were useful in some circumstances. Further work needs to be done to collect and integrate deep phenotypic data in clinical research.

4.3.9 Challenges of collecting phenotypic data

Several challenges were encountered in the collection of phenotypic data. The first was the variety of locations and formats in which phenotypic and clinical data were stored. Manually extracting and re-entering these data was time consuming and open to the introduction of errors. However, because of the variety of formats, the fact that many of the storing databases do not have export function (aside, in some cases, from a patient's entire medical record or single records as a printout) meant that this was unavoidable. However, it meant that the process was labour intensive and it was impossible to record all the available data. It also became clear that since data were being recorded selectively, a clinician or other healthcare professional was required to filter data and make these decisions. This has significant implications for staffing costs in a project such as the HIGH-5 project. There were additional benefits to having a clinician recording the data, including the fact that they had pre-approved access to most of the medical databases required and it was relatively easy to obtain access to others, and that they were confident in selecting the most appropriate HPO terms, being familiar with them before the start of the project.

A second challenge was the storage of data in two separate databases. There were good reasons for this that would have been very difficult to overcome, such as the extremely large number of HPO terms the MS Access database would have been required to hold. However it meant that the process of entering and outputting data was more time-consuming than it otherwise might have been, both because some demographic data had to be entered and removed more than once, but also because of the need to swap back and forth between databases. One possible solution would be to integrate PhenoTips® software into a clinical database. Another possibility would be to generate the HPO numbers in PhenoTips® but then upload them to the MS Access database. This would not solve the problems mentioned above, but it would mean that all data for a given patient would be stored together and outputted together with a single anonymisation step.

A third challenge was the issue of backups and data loss within the IDHS. While the issues described in section 4.3.6 were eventually resolved, it reduced researcher confidence in the safety of the data. An external backup, while important, also added to the time-consuming nature of the project, because the procedure to extract data from IDHS is arduous and backups were required several times per day if structural changes were being made to the database. When no structural changes were being made, data were backed up only once or twice per day, as the risk of loss of a single day's data entry was considered relatively trivial.

A fourth issue was that PhenoTips® is not designed to record phenotypic data that changes over time, but rather static phenotypic data such as polydactyly, which remains a feature of the patient's condition regardless of whether or not extra digits have been removed. However, features such as colitis, joint inflammation and deformity or skin condition may vary significantly. The MS Access database was much more useful for recording these types of data. This is because PhenoTips® was designed to record features related to genetic diseases and not for longitudinal profiling of patients with variable signs and symptoms. This is something that needs further consideration in the development of phenotyping software or databases in future, as it seems that one of the important uses of multiomics may be in the longitudinal monitoring of patients with chronic disease(451).

A fifth issue was that while HPO terms covered all of the terms required to describe the phenotypic abnormalities of patients with BBS and Usher syndrome they covered the majority, but not all, of the terms required to describe the IBD and JDM patients. However, it was possible to manually add non-HPO terms to both Phenotips and the MS Access database and this was done in the interests of phenotypic recording accuracy. Utilising SNOmed or other terms would not have provided additional benefit over HPO terms.

A sixth issue was that the cohorts in the HIGH-5 project were both small and heterogeneous. This meant that when patients were stratified into groups by reference to clinical data, the number of individuals in each group was small, making analysis very challenging. This means that larger, more homogeneous cohorts may be necessary until more is known about multiomics analysis.

A challenge that did not arise during the study, but might have become important, was that MS Access databases cannot deal with very large volumes of data and become difficult to use if too large. This might have to be considered if a very large cohort was being studied or if the study was likely to continue to collect phenotypic data longitudinally over many years.

4.3.10 Ways in which recording phenotypic information might be improved

Developing a database that had both the capabilities of both PhenoTips® and the MS Access database would be an improvement on the current situation of using two databases. Another consideration is whether a secure NHS-approved cloud based storage, such as the one being used by the nationally commissioned Bardet-Biedl Service would be a useful alternative. This has the advantage of being accessible by multiple permitted people from any site. The disadvantages would be the inability to modify a commercial product as and when it was required, which was

one of the major advantages of the MS Access database. However, as more omics studies are done and researchers become more familiar with which datasets are useful, this problem may be overcome. A solution would be the development of a national EHR where all data for a single patient would be available. Currently in the UK, even when EHR are in use, they do not contain all categories of clinical data. In addition, historical records are generally not available as part of the electronic record or, if they are, it is as scanned PDFs, which may or may not be searchable. While other countries may be ahead of the UK in this area, it is clear that national EHR will not be without its challenges(452-455). An EHR might also allow software such as Sapientia™ for Omics to harvest and anonymise relevant clinical data, significantly reducing the labour-intensity of recording phenotyping data. Utilising rare disease registries, where much clinical data are recorded, especially if the use of HPO terms becomes widespread, is another possible strategy(456).

4.3.10.1 Minimum omics dataset

A minimum data set for rare disease research was proposed in 2015(457). A 2017 paper by Kennell et al. considered which forms of data in the EHR would be most useful to researchers(458). A knowledge of what types of data would and would not be useful for multiomics would have solved many of the challenges listed in section 4.3.9, but this is not a question that has been answered yet. Even within a single project, researchers varied in the types of data they found useful or even attempted to use. This may become clear as further multiomics studies are published. Multiomics research is more complex than rare disease genetic and genomic research, due to the presence of multiple large datasets requiring integration and interpretation for every patient. It is therefore possible that a greater depth of clinical knowledge will be required for multiomics studies.

4.4 Conclusions and future directions

Tools to facilitate deep phenotyping are becoming available and need to be more widely utilised. HPO terms allow the accurate classification of patients, and should be used in medical literature and clinical notes. Comprehensive, integrated EHR will facilitate the use of deep phenotyping data in research. Studies such as HIGH-5 allow determination of which metrics are useful, how they are best recorded, and by whom. Recording of all clinical data is impossible in a non-automated system, but whether any other data types would have been useful may become clearer as analysis continues. Data such as blood results not linked to samples or phenotypic information is difficult to use, hence the recording of results linked to samples being analysed, clinic visits or changes in phenotype. A minimum data set for rare disease research has been proposed but this is unlikely to provide the depth of phenotypic information required for omics research.

While the aim of building a database and recording phenotypic information was achieved, many difficulties were encountered. These could, at least in part, have been avoided had the data collection been done prospectively rather than retrospectively. Collecting data prospectively is unrealistic except in the context of a standardised electronic health record that would hold all

patient data. Collecting retrospective data may become easier with increased use of HPO terms and the ability to mine computerised records for data.

Future directions include adding further tables to the database to allow the addition of other data types and to continue working on the integration of a clinical database with multiomics analysis software. As electronic medical records gain traction, the aim will be to pull these data directly from the medical record, which will save time and reduce the possibility or error. Further work on the multiomics datasets will allow more use of the clinical phenotype. One very positive step will be the utilisation of HPO terms in clinical phenotyping. Although currently this is not a widely used practice, the 100,000 genomes study has increased awareness of it outside the field of clinical genetics and the if the policy of insisting on HPO terms as well as clinical descriptors is taken up more widely, this may increase use also.

Chapter 5 Extracting pharmacogenomic information from whole genome sequence data

5.1 Introduction

Pharmacogenomics, previously known as pharmacogenetics, is the study of how genomic variation affects drug response. The term pharmacogenomics reflects the understanding that non-coding genetic variants are important in drug metabolism. It is a rapidly growing field as evidenced from a PubMed search, with six papers with pharmacogenetics in the title published in 1967, while since 2008 over 400 papers with either term in the title have been published each year. The term pharmacogenetics was first published by Vogel in 1959, while pharmacogenomics was introduced by Marshall in 1997(459, 460). Pharmacogenomics is applicable to every field in medicine and affects every person who will, over the course of their lifetime, take medication. The aim of pharmacogenomics is to allow the true personalisation of drug therapy to maximise benefit and minimise side effects for every patient.

5.1.1 Early pharmacogenomics

Although the first paper to use the term pharmacogenetics in the title was published in 1967, the history of pharmacogenomics is longer than that. G6PD deficiency was described in the 1950s, although Pythagoras was aware of the potential risks of eating fava beans in 500BC(461, 462). As early as the 1920s, haemolysis in response to antimalarials had been documented and was later linked to G6PD deficiency(463, 464). Other deficiencies including pseudocholinesterase and N-acetyl transferase deficiencies had been identified also(465). Motulsky summarised the state of pharmacogenomic research in 1957 and from there the field advanced rapidly(466). Subsequent research identified some of the genetic variants that caused these phenomena(467). Discovery of the cytochrome P450 genes and their genetic complexity was another huge advance in the field(468, 469). Developments in pharmacogenomics reflected the growing armoury of drugs that physicians had access to in the mid-20th century, and grew from a desire to understand why some patients benefitted while others did not, and why only some suffered side effects. Drug safety issues, such as those with thalidomide which came to light in the early 1960s, also prompted the development of a field which studied adverse drug reactions (ADRs). The study of pharmacogenomics benefitted greatly from emerging scientific techniques such as spectrophotometry and genetic sequencing(470). From the beginning, the field of

pharmacogenomics was linked to race, as variable responses to drugs had been seen in different ethnic groups. While this led to some controversy, Kalow, who was a pioneer in the field, realised that this had profound implications, as many drugs were tested on small groups of patients, often of a single ethnicity, and there was potential for missing serious ADRs in other ethnic groups(470).

5.1.2 Pharmacogenes

Pharmacogenes are genes which are involved in the absorption, distribution, metabolism and excretion of drugs, genetic variants in which affect the action and side effects of medication. Many have been identified, but there is great variety in what is known about each one. PharmGKB (www.pharmgkb.org) is a site that aims to analyse the pharmacogenomic literature, curate it and extract the relevant variants(471). Important variants are given a clinical annotation depending on amount of evidence, study sizes and degree of statistical significance (Table 5.1). PharmGKB lists 21,107 variant annotations at time of writing, with new variants being added constantly. The evidence is assessed and assigned a weight. Level 1a and 1b annotations are the most important. Genes with good evidence are assigned very important pharmacogene (VIP) status and currently 65 are listed(472-494). These are genes where significant information about the effect of genetic variants on drug action is known. In the US, many of these genes and their effects are listed on Food and Drug Administration (FDA) drug labels, divided into the following categories. *Testing required* means that the FDA suggests that this drug should not be prescribed without a test, while *testing recommended* means that it would be preferable to perform testing prior to prescription. *Actionable* means there is information available about the possible effects of genetic variants and *informative* means that there is information about a gene being involved in the pharmacodynamics or pharmacokinetics of a drug. Pharmacokinetics refers to the body's effect on a drug, such as altering absorption, distribution, metabolism or excretion, while pharmacokinetics refers to the drug's effect on the body, for example, its ability to stimulate or inhibit a particular biological pathway. The FDA currently lists 164 drugs that have pharmacogenomic guidance on their label but there may not be robust evidence available for how to vary prescribing for all of these(495). The European Medicines Agency (EMA) has also introduced pharmacogenomic labelling(496).

5.1.2.1 Pharmacogenes and alteration of prescribing

For some pharmacogenes, there is enough evidence to determine the likely effect of genetic variants on drug effect, allowing the development of pharmacogenomic prescribing guidelines. Currently, PharmGKB lists 18 genes that have prescribing guidelines associated with them (section 5.1.2.2). These are referred to as *clinically actionable*. The guidelines have been developed by a number of working groups, mainly the Clinical Pharmacogenetics Implementation Consortium (CPIC) and the Dutch Pharmacogenomics Working Group (DPWG). Some drugs have additional guidelines from groups such as the Canadian Pharmacogenomics Network for Drug Safety (CPNDS) or other groups. There are 100 different guidelines relating to 72 different drugs listed. Prescribing is altered depending on what genotype or diplotype the patient has. For many genes, patients can be divided into normal, intermediate, slow or ultra-rapid metabolisers, while for others, prescribing is based on the presence or absence of a certain SNP. Additional pharmacogenes without associated prescribing guidelines are discussed in Chapter 6. The

development of guidelines is a laborious process. It has to take into account the accuracy of testing methods, the level of evidence for phenotypic effect and the evidence for alteration of prescribing. To be worthwhile, testing and implementation of guidelines need to be useful in preventing side effects, increasing treatment efficacy or reducing costs(497). They also need to be easy to use as, in a time-pressured healthcare system, there is little scope for the addition of complex processes.

Level	Evidence requirement
Level 1A	Annotation for a variant-drug combination in a CPIC or medical society-endorsed pharmacogenomic (PGx) guideline, or implemented at a Pharmacogenomics Research Network site or in another major health system
Level 1B	Annotation for a variant-drug combination where the preponderance of evidence shows an association. The association must be replicated in more than one cohort with significant p-values, and preferably will have a strong effect size
Level 2A	Annotation for a variant-drug combination that qualifies for level 2B where the variant is within a VIP (Very Important Pharmacogene) as defined by PharmGKB. The variants in level 2A are in known pharmacogenes, so functional significance is more likely
Level 2B	Annotation for a variant-drug combination with moderate evidence of an association. The association must be replicated but there may be some studies that do not show statistical significance, and/or the effect size may be small
Level 3	Annotation for a variant-drug combination based on a single significant (not yet replicated) study or annotation for a variant-drug combination evaluated in multiple studies but lacking clear evidence of an association
Level 4	Annotation based on a case report, non-significant study or *in vitro*, molecular or functional assay evidence only

Table 5.1 Levels of evidence for clinical annotation, adapted from www.PharmGKB.org

5.1.2.1.1 Publishers of Guidelines

CPIC is an international network of volunteers with expertise in pharmacogenomics who have set out to analyse the available pharmacogenomic evidence and develop prescribing guidelines. The project began in 2009, and all guidelines assess evidence in a standardised and systematic manner, and are peer-reviewed before publication (www.cpicpgx.org). The DPWG is a multidisciplinary body, established by the Royal Dutch Pharmacists' Association in 2005. Its aims are both to develop guidelines but also to facilitate the integration of guidelines into everyday practice in the Netherlands(498, 499). The CPNDS was set up to reduce adverse drug effects in children, but their guidelines often cover both adult and paediatric patients (www.cpnds.uba.ca).

5.1.2.1.2 Single or combined guidelines

Many genes have guidelines that look only at a single gene and how it affects metabolism of a drug or drug class. However, in some cases, effects of more than one gene are considered together. For example, when prescribing phenytoin, the CPIC recommends that both *CYP2C9* and *HLA-B* genotypes are considered. Sometimes there are separate guidelines for more than one gene and a single drug, but the advice has not been combined, which may lead to conflicts in prescribing, for example in the case of tricyclic antidepressants (TCAs), where separate guidelines exist for *CYP2C19* and *CYP2D6*. Occasionally, there are conflicts between guidelines, such as which phenotypic category the diplotype should be considered under, or what the alteration in prescribing should be. This is discussed in section 5.3.5.

5.1.2.2 Clinically actionable pharmacogenes

Details of clinically actionable pharmacogenes and guideline outcomes can be found in Supplementary Information S2.1 (CD-ROM).

5.1.2.2.1 Cystic Fibrosis Transmembrane Regulator *(CFTR)*

CFTR is considered a clinically actionable pharmacogene, but only in people with a genetic diagnosis of cystic fibrosis (CF)(483, 500). Depending on which mutations a patient has, they may or may not benefit from ivacaftor, originally licensed for patients with the p.Gly551Asp mutation. This is a class III mutation and therefore affects the gating ability of the CFTR channel, resulting in loss of chloride transport(501). The commonest class III mutation, it accounts for approximately 4% of CF worldwide(502). Other class III mutations are rare. Ivacaftor has now been licensed for other CF mutations and can be prescribed in the UK for people with any of the following: p.Gly178Arg, p.Ser549Asn, p.Ser549Arg, p.Gly551Ser, p.Gly1244Glu, p.Ser1251Asn, p.Ser1255Pro or p.Gly1349Asp in *trans* with any other CF mutation(503). Ivacaftor is a CFTR channel potentiator, keeping the CFTR channel open and allowing chloride ions to pass through. It can also be prescribed with lumacaftor for p.Phe508del, the commonest CF mutation(504). *CFTR* is not usually included in pharmacogenomic panels as it is only relevant in CF, diagnosis of which routinely includes mutation identification.

5.1.2.2.2 Cytochrome P450, subfamily IIC, polypeptide 9 *(CYP2C9)*

CYP2C9 codes for one of many cytochrome P450 (CYP450) enzymes. It is mainly expressed in the liver. CYP2C9 is required for the metabolism of many substances, both endogenous and exogenous and is responsible for the metabolism of anti-coagulants, anti-epileptics and hypoglycaemics among others. It is estimated to metabolise up to 20% of drugs undergoing phase I metabolism(505, 506). There are more than 60 different alleles associated with *CYP2C9* but currently the guidelines cover only alleles *1, *2, *3 and combinations thereof(499, 507-510). Individuals with *1/*1 are considered normal metabolisers (NM), *1/*2 and *1/*3 are intermediate metabolisers (IM) and *2/*2, *2/*3 and *3/*3 are poor metabolisers (PM) according to CPIC guidelines and this is borne out by experimental evidence(511, 512). Other alleles are listed in PharmGKB, some of which have associated recommendations. *2 is the commonest non-wild type allele in Europeans, found in 10-20% of individuals, but is much rarer in other ethnicities(513-

515). The *3 allele is less common in white Europeans, but is the commonest allele in South East Asians, while alleles *5 and *8 are commonest in African Americans. At present, there are CPIC and CPNDS guidelines for *CYP2C9* and warfarin, with the CPIC guideline incorporating the genes *VKORC1* and *CYP4F2*. DPWG guidelines cover acenocoumarol, phenprocoumon and phenytoin. The DPWG considered additional drugs, including sulfonylureas, but have not yet made prescribing recommendations.

5.1.2.2.3 Cytochrome P450, subfamily IIC, polypeptide 19 *(CYP2C19)*

CYP2C19 is another CYP450 enzyme. It is also expressed in the liver and is involved in the metabolism of many common drugs, including anti-depressants, anti-psychotics, platelet inhibitors and proton pump inhibitors (PPIs)(478, 499, 516-520). PharmVar lists 35 different alleles, but again, prescribing recommendations have been made only for the best characterised, *1, *2, *3 and *17(507, 518). Four metaboliser phenotypes are known; NM, for example *1/*1, IM, for example *1/*2, *1/*3, *2/*17, PM, for example *2/*2, *2/*3 and *3/*3 and UM (ultra-rapid metaboliser), for example *17/*17(518, 521, 522). *17 accounts for approximately 16-17% of alleles in Caucasians and those of African-American origin, but is much rarer in Asians. *2 accounts for 12-15% in Caucasians and African-Americans but, at up to 35%, is much commoner in those of Asian origin. *3 is common in Asians at frequencies of up to 9% but rare in other populations(518).

5.1.2.2.4 Cytochrome P450, subfamily IID, polypeptide 6 *(CYP2D6)*

Another member of the CYP450 family, the CYP2D6 enzyme is responsible for the metabolism of up to 25% of all commonly prescribed drugs and is one of the most clinically important pharmacogenes(523). It is involved in the metabolism of antiarrhythmics, antidepressants, antipsychotics, antiemetics, antihypertensives, opioid painkillers and chemotherapeutic agents among others(499, 517, 524-527). Again, there are four phenotypic classes: NM, for example *1/*1, *2/*2, *1/*2, *1/*4, *1/*5, *1/*41, *2/*41; *41/*41, IM, for example *4/*10, *4/*41, *5/*9; PM, for example *3/*4, *4/*4, *5/*5, *5/*6; UM for example *1/*1N, *1/*2N and *2/*2N, where N is a gene duplication. 113 different alleles are listed on PharmVar, though the amount of functional information available varies, and published prescribing guidelines are currently limited to those listed above. Allele frequencies can be found at www.pharmgkb.org, but *1 accounts for between 30-40% in most populations, *2 for 10-30% and *41 for 7-10%. Due to the large number of alleles, the existence of copy number variants in the form of gene duplications and deletions, the presence of a pseudogene and a homologous, non-functioning gene in the same region of chromosome 22 and its homology to other CYP genes, *CYP2D6* is one of the most challenging genes to analyse for pharmacogenomics(83, 528).

5.1.2.2.5 Cytochrome P450, subfamily IIIA, polypeptide 5 *(CYP3A5)*

CYP3A5 is involved in the metabolism of tacrolimus, ciclosporin and other drugs and is expressed in the liver(529). PharmVar lists 11 alleles, some of which are subdivided, but other SNPs have been identified so it is likely that more will be described as functional consequences are delineated. Currently, prescribing guidelines cover alleles *1 to *9(499, 530). *CYP3A5* *3 is the

commonest allele. Variants in *3 and *6 cause aberrant splicing of *CYP3A5*, leading to nonsense mediated decay and reduced expression(531). The *3 allele frequency is up to 95% in white Europeans. It is lower in African Americans at about 33%, and seen at frequencies of above 50% in most other populations(531-533).

5.1.2.2.6 Dihydropyrimidine Dehydrogenase *(DPYD)*

DPYD is the enzyme responsible for the rate limiting step in the catabolism of the pyrimidine bases thymine and uracil, and is expressed mainly in blood(534). As such it is vital for the metabolism of 5-fluorouracil and related chemotherapeutic agents such as capecitabine and tegafur. Variants that reduce DPYD function increase risk of toxicity(535). There are currently 13 known haplotypes, and prescribing guidelines are available for all(499, 536-539). 3-5 % of the population carry at least 1 deficiency allele(534). Accounting for about 50% of deficiency alleles, *2A is the best characterised variant haplotype. It introduces a cryptic splice site in intron 14. The prevalence of the *2 allele is variable but has been estimated at around 1% in white Europeans(539). Heterozygosity results in partial DPYD deficiency. Homozygotes can suffer severe reactions including pancytopaenia and may require significant reductions in drug dose.

5.1.2.2.7 Coagulation Factor V (*F5*)

Factor V is a coagulation factor that requires conversion to its active form by thrombin, and is itself a co-factor in the conversion of prothrombin to thrombin(540). It is inactivated by activated protein C (APC). Mutations in *F5* can affect APC-mediated inactivation. The p.Arg534Gly mutation, also referred to as p.Arg506Gly, is known as factor V Leiden (FVL) and causes APC resistance, resulting in a pro-coagulant state(541). Studies have shown a population heterozygote frequency of about 2%(542). Individuals with FVL are at increased risk of deep vein thrombosis, pulmonary embolism and possibly other thromboembolic conditions(543, 544). Other variants have been identified in *F5*, but currently prescribing guidelines are only available for FVL and oral contraceptives(499).

5.1.2.2.8 Glucose 6 Phosphate Dehydrogenase (*G6PD*)

G6PD is an enzyme of the hexose monophosphate pathway, essential for the production of NAPDH in red blood cells. Deficiency can result in haemolytic disease as the cells become more susceptible to oxidative stress(545). As *G6PD* is found on the X chromosome, symptomatic G6PD deficiency is more likely in males. Female homozygotes are also affected, while the phenotypic effect is variable in heterozygotes(546). Many foods and drugs can cause haemolysis, including primaquine and rasburicase(547, 548). Many *G6PD* mutations have been reported and residual G6PD activity varies depending on the mutation(549, 550). Overall G6PD deficiency is estimated to affect almost 5% of the world's population and it is hypothesised that this very high frequency has resulted from deficiency being protective against malaria(551-553). Different alleles are found in different populations. Some of the most common are the Mediterranean and African alleles. Currently, guidelines cover only the prescription of rasburicase(548).

5.1.2.2.9 Major Histocompatibility Complex, Class I A (*HLA-A*)

HLA-A is involved in antigen presentation to T cell receptors. Presentation of non-native peptides triggers an immune response(554). The *HLA-A* *31:01 allele has been associated with Stevens Johnson syndrome (SJS) and other severe cutaneous adverse reactions (SCAR) in response to carbamazepine and similar drugs but the mechanism is unclear(555). The HLA genes are highly polymorphic with the World Health Organisation (WHO) HLA nomenclature database listing over 3000 variants in *HLA-A* (http://hla.alleles.org/nomenclature/index.html). *HLA-A* *31:01 is found at a frequency of 1-8% worldwide(555, 556). However, not every patient with *HLA-A* *31:01 who is exposed to carbamazepine will develop an ADR(557).

5.1.2.2.10 Major Histocompatibility Complex, Class I B (*HLA-B*)

HLA-B is, like HLA-A, part of the major histocompatibility complex and is also involved in antigen presentation(558). There are several *HLA-B* alleles of interest including *HLA-B* *15:02, which is important in antiepileptic-associated SCAR (carbamazepine, oxycarbazepine and phenytoin), *HLA-B* *57:01 which is important in abacavir hypersensitivity and *HLA-B* *58:01 which is implicated in allopurinol-related SCAR(499, 508, 555, 557-559). There is also an association between *HLA-B* *44 and ribavirin toxicity, but there are not yet any published guidelines. Frequency depends on the particular allele. *HLA-B* *15:02 is commonest in the East, South and Central Asia, with rates of up to 6% in some subpopulations such as the Han Chinese, while it is very rare in Africans and Europeans(560). *HLA-B* *57:01 is seen in about 6-7% percent of Europeans, in up to 20% of Southwest Asians and is much lower in other populations(559). *HLA-B* *58:01 is seen in up to 20% of people from some parts of Asia, but is rare in the UK and Western Europe(561, 562). As with *HLA- A* *31.01, not all patients with the particular allele will develop ADRs with relevant drug exposure. For abacavir, the positive predictive value of a test is only 50%(563).

5.1.2.2.11 Interferon, Lambda-3 (*IFNL3*)

IFNL3 is a cytokine receptor, expressed on hepatocytes and is involved in the activation of the JAK/STAT signalling pathway, which plays key roles in the immune response by altering gene transcription(564). It is a factor in whether or not sustained virologic response (SVR) can be obtained following treatment of hepatitis C with antivirals such as PEGylated interferon and ribavirin. SVR is defined as having undetectable levels of viral RNA between 12 and 24 weeks after completion of treatment. As late relapse rates are low, this is a marker for long-term efficacy(565). It was recognised that some ethnic groups were less likely to achieve SVR and later genome wide association studies (GWAS) identified *IFNL3* as the gene mediating this effect(566, 567). It is now known that the CC genotype at rs12979860 or the TT genotype at rs8099917, which are in strong linkage disequilibrium, are the best predictors of whether SVR will be achieved. Favourable alleles are seen at the highest rates in Asians, less frequently in Caucasians and most rarely in those of African origin, explaining the ethnic differences in treatment response previously noted(565).

5.1.2.2.12 Retinoic Acid Receptor, Gamma (*RARG*), Solute Carrier Family 28, Member 3 (*SLC28A3*) and UDP-Glycosyltransferase 1 Family, Polypeptide A6 (*UGT1A6*)

RARG encodes a ligand-dependent regulator of transcription, and along with *SLC28A3*, a nucleoside transporter and *UGT1A6*, which is involved in detoxification, is known to be a factor in anthracycline-associated toxicity (AAC). Anthracyclines are used in the treatment of adult and paediatric solid tumours and haematological malignancies. A single guideline exists for these genes, and genetic testing is only recommended in paediatric patients receiving doxorubicin or daunorubicin, as the evidence is stronger in children than in adults(568). While other genes are known to affect anthracycline toxicity, there is not yet enough evidence to alter prescribing based on them. The *SLC28A3* variant is considered to be protective against AAC, while variants in *RARG* and *UGT1A6* are risk factors. The allele frequencies in *RARG*, *SLC28A3* and *UGT1A6* vary with ethnicity but ExAC lists them as 8%, 15% and 4% respectively.

5.1.2.2.13 Solute Carrier Organic Anion Transporter Family, Member 1B1 (*SCLO1B1*)

SLCO1B1 whose product is OATP1B1, is a gene encoding a molecular transporter involved in the transport of drugs across the hepatocyte membrane and is vital for normal drug metabolism(569). A SNP in *SLCO1B1* has been associated with increased risk of myopathy with simvastatin, a commonly prescribed cholesterol-lowering agent(570). The SNP is associated with several alleles, which are described as low, intermediate or high-function alleles, which are associated with high, intermediate or low risk respectively of statin-induced myopathy. Reduced function of the transporter results in exposure to higher levels of the active form of simvastatin, increasing myopathy risk. However, *SLCO1B1* genotype is only one of a number of factors influencing myopathy risk with age, dose and gender, as well as drug interactions among the others. Homozygosity for the high-risk SNP is seen in up to 6% of individuals, while heterozygosity is seen in 11-36%(570).

5.1.2.2.14 Thiopurine S-Methyltransferase (*TPMT*)

TPMT is involved in the inactivation by S-methylation of various compounds including thiopurines such as azathioprine, 6-mercaptopurine (6MP) and thioguanine(571). Inactivation can also occur via other pathways but the compounds formed by these other routes are cytotoxic or inhibit purine synthesis(572, 573). As early as 1980, it was known that individuals had high, intermediate or low TPMT activity and that this was likely to be inherited in a Mendelian manner(574). Low levels of TPMT activity are correlated with increased risk of myelosuppression following treatment with thiopurines which are used for immunosuppression and as anti-cancer agents(575). Individuals with two wild type alleles account for 85-97% of the population. Between three and 14% of people are heterozygotes, while homozygotes for low TPMT levels are rare in all ethnic groups(499, 571). There is also evidence that *TPMT* polymorphisms may be implicated in cisplatin-induced hearing loss and guidelines for cisplatin prescribing in children have recently been published by CPNDS(576). It is thought that the majority (>90%) of TPMT deficient individuals can be identified with the testing of 3 individual SNPs, although many additional alleles are known(471, 571, 577). Currently, published prescribing guidelines cover only 6 alleles; *1, *2, *3A, *3B, *3C and *4(571).

Genotyping prior to prescribing has been borne out by prospective studies(578, 579). It is also possible to test TPMT activity directly.

5.1.2.2.15 UDP-Glycosyltransferase 1 Family, Polypeptide A1 (*UGT1A1*)

Mutations in *UGT1A1* have been implicated in Crigler-Najjar syndrome and other conditions such as familial neonatal transient hyperbilirubinaemia and Gilbert syndrome. UGT1A1 is involved in the glucuronidation of many substrates including bilirubin and some drugs(580). UGT1A1 has been implicated in the development of jaundice in patients taking atazanavir, an antiretroviral protease inhibitor, and in toxicity with irinotecan, a chemotherapeutic agent(499, 581, 582). The active form of irinotecan requires glucuronidation for effective clearance, while atazanavir inhibits UGT1A1, causing hyperbilirubinaemia and jaundice. Individuals can be divided into normal, intermediate or poor metabolisers depending on the alleles they have. A TA repeat in the promoter of *UGT1A1* is linked to activity. Six repeats is normal (*1) and the commonest allele across all populations, while five repeats is associated with increased levels (*6) and seven or eight with reduced levels (*28 and *27)(583). A single polymorphism, rs887829, is in strong linkage disequilibrium with the repeat, with C being associated with five and six repeats and T with seven and eight repeats. A further *6 allele has been identified as important in some Asian populations, but is rare in Europeans(581).

5.1.2.2.16 Vitamin K Epoxide Reductase Complex, Subunit 1 (*VKORC1*)

The vitamin K epoxide reductase complex subunit-1 protein is an important enzyme in the conversion of vitamin K epoxide to vitamin K, and is the target of warfarin(584). Warfarin is an inhibitor of VKORC1, and reduces available vitamin K, an important co-factor in the clotting cascade. *VKORC1* polymorphisms are more important in warfarin metabolism than *CYP2C9,* the other major gene involved(585). Additional factors such as age and weight are also important and dosing algorithms, such as the International Warfarin Pharmacogenetics Consortium (IWPC) algorithm (www.cpicpgx.org) have been designed to take account of these(509, 510). *VKORC1* polymorphisms can also affect the prescribing of other coumarins such as acenocoumarol and phenprocoumon(499). Both the CPIC and CPNDS guidelines look at *CYP2C9* and *VKORC1* in combination. The CPIC guideline also looks at *CYP4F2* and an additional SNP, rs12777823, which is important in individuals of African ancestry(510, 586, 587). People of African ancestry also require more extensive genotyping of *CYP2C9*. A single SNP accounts for most of the variation due to *VKORC1*, and the frequency is variable(509).

5.1.2.2.17 Other pharmacogenes of interest

Cytochrome P450 subfamily IVF, polypeptide 2 (*CYP4F2*), which codes for the enzyme Leukotriene-B(4) omega-hydroxylase 1 is included in the CPIC warfarin dosing guidelines and has a role in removing vitamin K from the clotting cycle(510). Studies have shown that including a single SNP in warfarin-dosing algorithms improve dosing accuracy for European and Asian populations(588).

The CPIC plans to release other guidelines in future. This includes adding *NUDT15* to the thiopurine prescribing guidance, considering the role of *ABCB1* in the metabolism of TCAs and

selective serotonin reuptake inhibitors (SSRIs), *SCN1A* in carbamazepine and oxycarbazepine metabolism, *RYR1* and *CACNA1S* and the effect of inhaled anaesthetics, *CYP2B6* in the prescription of methadone and efavirenz, *MTHFR* and methotrexate metabolism and *NAT2* and isoniazid toxicity. In addition, a number of other drugs, such as more statins, are being considered for inclusion in future guideline updates. Details can be found at https://cpicpgx.org/prioritization-of-cpic-guidelines/. As evidence accumulates, changes will be made to guidelines, and indeed some studies are undertaken specifically to see if changes to current guidelines are appropriate(589). The ever-changing guideline landscape is important when considering testing methods as current tests may become obsolete as guidelines change.

5.1.3 Adverse drug reactions

The main reason for performing pharmacogenomic testing is the avoidance of ADRs. The World Health Organisation defines ADRs as “a response to a drug that is noxious and unintended and occurs at doses normally used in man for the prophylaxis, diagnosis or therapy of disease, or for modification of physiological function”(590). They are very variable. In some cases, rapid metabolism results in sub-therapeutic drug levels as a drug is rapidly cleared, as is the case with *CYP2C9* rapid metabolisers and warfarin(591). In the case of drugs that are metabolised from a pro-drug to an active metabolite, rapid metabolisers are at risk of overdose, as in the case of *CYP2D6* and codeine(592). Others relate to the development of side-effects as in the case of hypersensitivity reactions to carbamazepine with certain *HLA* haplotypes(593).

ADRs can be serious and life-altering and occasionally fatal(593-595). They have enormous economic costs in terms of treatment, extended hospital admissions and time off work(596-600). However, it is difficult to isolate the cost of genetic causes of ADRs as there are many causes of drug toxicity. Also, modelling may not take account of ineffective treatment due to sub-therapeutic dosing, which can reduce response to treatment, lead to disease relapse or recurrence and lengthen hospital admission(601, 602). This is discussed further in Chapter 6.

5.1.4 Pharmacogenomic testing

Pharmacogenomic testing requires the identification of particular genetic variants in a patient, the conversion of these into a genotype or diplotype, assignation of a diplotype to a phenotypic category and application of relevant prescribing guidelines. Prescribing guidelines are usually based upon the phenotypic assignation.

5.1.4.1 Genotypes and diplotypes

PharmGKB and other sources publish lists of variants associated with the various alleles or haplotypes that have been defined for a gene. Identification of these is the first priority in pharmacogenomic testing. Following this, the variants are combined to determine which particular allele/haplotype they represent, thus identifying genotype/diplotype. Phasing is an issue in pharmacogenomic testing (section 5.3.2.2.1). Some genes have many different possible diplotypes, and prescribing guidelines often cover only the commonest of these. In addition, if

looking at more than just a small number of SNPs it may be difficult to determine which SNPs are relevant and which are not (section 5.3.1.2)

5.1.4.2 Clinical pharmacogenomic testing methods

Pharmacogenomic testing can be done on a case-by-case, drug-by-drug basis, or in an anticipatory manner using panel-based SNP genotyping or whole genome sequencing.

5.1.4.2.1 SNP genotyping

SNP genotyping is the commonest method of pharmacogenomic testing and is the basis of most current commercial tests (section 3.1.1.2). There are different laboratory techniques from providers such as Canon, Illumina and Thermofisher, but the principle is to identify whether or not a SNP is present at a particular genomic location. SNP genotyping can test for a few or many SNPs simultaneously, as in the case of Illumina's Infinium Omni5 Exome-4 array, which looks at 4.5 million SNPs. The limitations of SNP genotyping include the fact that only pre-specified SNPs are tested. This means that important SNPs may not be included, such as rarer SNPs that are important in certain ethnicities or ones whose clinical importance is not yet understood. SNP genotyping arrays usually do not provide information about phase, so that it cannot be determined which SNPs are associated with which copy of the chromosome, although biallelic SNP genotyping is now possible(603). SNP genotyping cannot look at copy number variants (CNVs), important for genes such as *CYP2D6*, and may not be effective for area of high sequence homology, including pseudogenes.

5.1.4.2.2 Whole genome sequencing (WGS)

WGS (section 3.1.1.3.3) overcomes some of the limitations of SNP genotyping in that if there is sufficient coverage, even rare SNPs can be looked at and SNPs later identified as being of importance can be examined retrospectively. It is also possible to pick up CNVs, although this can be challenging. However, as with SNP genotyping, short read WGS cannot give information about phase. Due to the high levels of homology between some pharmacogenes, there has been concern that next-generation sequencing data may not be accurate for pharmacogenomics(604). This is in addition to the usual issues with next generation sequencing such as read depth and accuracy(605-607). It is also more expensive and analysis, unless automated, is difficult and time consuming. Long-read WGS is likely to be able to detect phase, allow for improved detection of CNVs and be better at sequencing regions of homology(363, 608). Whole exome sequencing (WES) cannot look at intronic variants.

5.1.4.2.3 Phenotypic testing

Not strictly a pharmacogenomic test, enzyme activity can be measured for the products of some pharmacogenes and is often done for G6PD and TPMT(609, 610). The advantage of this approach is that it gives evidence of phenotypic effect and eliminates the possibility of misinterpretation of phenotype, for example due to the presence of a rare diplotype. Studies have shown that while genotype-phenotype correlation is generally excellent in *TPMT*, it is less good in those in the intermediate activity range, falling from near 100% to 70-90%(610, 611). It is also cheaper in a patient who is starting a single drug than a panel of pharmacogenomic tests.

Disadvantages include the fact that testing needs to be available locally, each enzyme must be tested individually, may delay the treatment and some studies have shown that it may be less effective than genotyping, for example in classification of TPMT-deficient patients(612).

5.1.4.3 Consequences of testing

Once patients have been put into a phenotypic category, a recommendation about prescribing must be made based on available prescribing guidelines. Occasionally guidelines conflict or are in disagreement about how a particular diplotype should be categorised and so a decision needs to be made about which guidelines should be used, preferably on a national level to avoid confusion. Many commercial pharmacogenomic testing and interpretations platforms use a traffic-light system to aid understanding of results, where drugs are highlighted as red, orange or green. Red drugs should be avoided, generally because the patient is an UM or a PM. Orange drugs should be used with caution, generally because the patient is an IM, often with a dose change or additional monitoring. Green drugs can be used as per standard prescribing guidance.

5.1.4.4 Pharmacogenomic testing in the UK

Genetic testing is recommended in the UK for abacavir, an anti-viral drug used in the treatment of HIV, although it is not always done(613). Testing is also recommended prior to carbamazepine prescription in the Thai and Han Chinese populations(614). A further example is Eliglustat, used in the treatment of Gaucher disease, which is prescribed at different doses depending on genotype and is not licensed for prescription without this information(615). TPMT testing is relatively common in the UK(616). There are also various trials that have been undertaken in the UK, such as genotyping for warfarin prescription, with early results having shown an improvement in time spent with a therapeutic INR, but overall pharmacogenomic testing is rare in the NHS setting(617). However, several companies are offering testing privately, and it is also available in the research setting.

5.1.5 Aims

The aims of this chapter were to use WGS data to visualise known pharmacogenomic SNPs and to determine genotypes or diplotypes for pharmacogenes of interest. This was in order to determine whether this is a possible use for WGS data and identify any genes for which this was problematic. In addition, it looked at whether Astrolabe, a bioinformatics tool, was helpful in confirming *CYP2D6* haplotypes and copy number from WGS data. Comparison of observed genotype or haplotype frequencies to published frequencies aimed to determine whether observed frequencies were similar to expected. Finally, in order to visualise what the clinical impacts of such tests might be, results were converted into prescribing guidance for individual patients and their medication history reviewed to determine whether relevant drugs were being prescribed.

5.2 Results

WGS data were analysed for 84 individuals from 5 cohorts, including 10 parent-child trios (SRS cohort) and 2 sets of monozygotic twins, (BBS-010 and BBS-011, and BBS-016 and BBS-017). Further details can be found in Chapter 2.

5.2.1 Genotypes and haplotypes

Genotypes of SNPs related to alleles associated with prescribing guidelines were extracted for the genes listed above, and for a single variant in *CYP4F2* and rs12777823, both important in the prescription of warfarin. Data were not extracted for *CFTR* (section 5.3.2.3.1). Copy number information for *CYP2D6* was obtained using the program Astrolabe (previously known as Constellation, see Chapter 2 and section 5.2.4(83). In addition, some haplotypes were changed after review of Astrolabe data. This is detailed in section 5.2.4. Full results for all individuals are listed in Supplementary Information 2.2 (CD-ROM) but examples for the BBS cohort are shown (Tables 5.2 to 5.12). Data were not obtained for *HLA-B* *44 or *HLA-B* *58:01, as these alleles are dependent on sequence rather than on SNPs. Although attempts were made to do this by sequence comparison, there is no clear data about what SNPs are or aren't allowable for an individual to be considered to have these haplotypes. SNPs present were then converted into diplotypes where appropriate. Haplotypes were also not interpretable for *DPYD* in some members of the SRS and IBD cohorts as discussed in section 5.3.2.2.1.

5.2.2 Prescribing advice

5.2.2.1 Future prescribing advice

Once the genotypes and diplotypes had been determined, they were compared with published guidelines, and prescribing advice determined for each individual. Both detailed and summary advice sheets were prepared. Examples can be seen in Table 5.13 and Figures 5.1 to 5.4. Full long-form and short-form prescribing advice for each individual can be seen in Supplementary Information 2.3 (CD-ROM). The number of genes for which results could change prescribing guidance is illustrated in Figure 5.5. Two patients had one diplotype or genotype that could lead to a change in prescribing guidance (2%), 12 patients had two (14%), 19 had three (23%), 27 had four (32%), 14 had five (17%), 8 had six (10%) and two had seven (2%). The mean number of actionable variants per patient was 3.8. That means that 100% of patients have a genetic change that could mean they require some alteration in dose or monitoring should they be prescribed a relevant drug.

Patient ID	Diplotype	rs1799853	rs1057910	rs28371686	rs9332131	rs7900194	rs28371685
		10:94942290	10:94981296	10:94981301	10:94949282	10:94942309	10:94981224
		g.94942290 C>T p.Arg144Cys	g.94981296 A>C p.Ile359Leu	g.94981301 C>G p.Asp360Glu	g.94949282 delA p.Lys273Argfs	g.94942309 G>A p.Arg150His	g.94981224 C>T p.Arg355Trp
	*1	C	A	C	A	G	C
	*2	T	A	C	A	G	C
	*3	C	C	C	A	G	C
	*5	C	A	G	A	G	C
	*6	C	A	C	del A	G	C
	*8	C	A	C	A	A	C
	*11	C	A	C	A	G	T
BBS-001	*1/*1	CC	AA	CC	AA	GG	CC
BBS-002	*1/*2	CT	AA	CC	AA	GG	CC
BBS-003	*1/*1	CC	AA	CC	AA	GG	CC
BBS-004	*1/*2	CT	AA	CC	AA	GG	CC
BBS-005	*1/*1	CC	AA	CC	AA	GG	CC
BBS-006	*1/*2	CT	AA	CC	AA	GG	CC
BBS-007	*1/*1	CC	AA	CC	AA	GG	CC
BBS-008	*1/*3	CC	AC	CC	AA	GG	CC
BBS-009	*1/*1	CC	AA	CC	AA	GG	CC
BBS-010	*1/*1	CC	AA	CC	AA	GG	CC
BBS-011	*1/*1	CC	AA	CC	AA	GG	CC
BBS-012	*1/*1	CC	AA	CC	AA	GG	CC
BBS-013	*1/*2	CT	AA	CC	AA	GG	CC
BBS-014	*1/*2	CT	AA	CC	AA	GG	CC
BBS-015	*1/*1	CC	AA	CC	AA	GG	CC
BBS-016	*1/*3	CC	AC	CC	AA	GG	CC
BBS-017	*1/*3	CC	AC	CC	AA	GG	CC
BBS-018	*1/*1	CC	AA	CC	AA	GG	CC

Table 5.2 Diplotypes for CYP2C9 in BBS cohort

Patient ID	Diplotype	rs4244285	rs4986893	rs28399504	rs56337013	rs72552267	rs72558186	rs41291556	rs12248560
		10:94781859	10:94780653	10:94762706	10:94852738	10:94775453	10:94781999	10:94775416	10:94761900
		g.94781859 G>A p.Pro227=	g.9480653 G>A p. Trp212Ter	g.94762706 A>G p.Met1Leu	g.94852738 C>T p.Arg433Trp	g.94775453 G>A p.Arg132Gln	g.94781999 T>A	g.94775416 T>C p.Trp120Arg	g.94761900 C>T
	*1	G	G	A	C	G	T	T	C
	*2	A	G	A	C	G	T	T	C
	*3	G	A	A	C	G	T	T	C
	*4A	G	G	G	C	G	T	T	C
	*4B	G	G	G	C	G	T	T	T
	*5	G	G	A	T	G	T	T	C
	*6	G	G	A	C	A	T	T	C
	*7	G	G	A	C	G	A	T	C
	*8	G	G	A	C	G	T	C	C
	*17	G	G	A	C	G	T	T	T
BBS-001	*1/*2	GA	GG	AA	CC	GG	TT	TT	CC
BBS-002	*1/*1	GG	GG	AA	CC	GG	TT	TT	CC
BBS-003	*1/*2	GA	GG	AA	CC	GG	TT	TT	CC
BBS-004	*1/*2	GA	GG	AA	CC	GG	TT	TT	CC
BBS-005	*1/*1	GG	GG	AA	CC	GG	TT	TT	CC
BBS-006	*1/*1	GG	GG	AA	CC	GG	TT	TT	CC
BBS-007	*1/*1	GG	GG	AA	CC	GG	TT	TT	CC
BBS-008	*1/*17	GG	GG	AA	CC	GG	TT	TT	CT
BBS-009	*1/*17	GG	GG	AA	CC	GG	TT	TT	CT
BBS-010	*1/*17	GG	GG	AA	CC	GG	TT	TT	CT
BBS-011	*1/*17	GG	GG	AA	CC	GG	TT	TT	CT
BBS-012	*17/*17	GG	GG	AA	CC	GG	TT	TT	TT
BBS-013	*1/*17	GG	GG	AA	CC	GG	TT	TT	CT
BBS-014	*1/*2	GA	GG	AA	CC	GG	TT	TT	CC
BBS-015	*1/*2	GA	GG	AA	CC	GG	TT	TT	CC
BBS-016	*1/*1	GG	GG	AA	CC	GG	TT	TT	CC
BBS-017	*1/*1	GG	GG	AA	CC	GG	TT	TT	CC
BBS-018	*1/*1	GG	GG	AA	CC	GG	TT	TT	CC

Table 5.3 Diplotypes for CYP2C19 in BBS cohort

Patient ID	Diplotype	rs16947	rs1135840	rs35742686	rs1065852	rs3892097	rs5030655	rs5030656	rs28371706	rs28371725	rs769258	Notes
		22:42127941	22:42126611	22:42128242	22:42130692	22:42128945	22:42129084	22:42128174-42128176	22:42129770	22:42127803	22:42130761	
		g.42127941 G>A p.Arg296Cys	g.42126611 C>G p.Ser486Thr	g.42128242 delT p.Arg259Glyfs	g.42130692 G>A p.Pro34Ser	g.42128945 C>T	g.42129084 delA p.Trp152Glyfs	g.42128174_42128176 delCTT p.Lys281del	g.42129770 G>A p.Thr107Asn	g.42127803 C>T	g.42130761 C>T p.Val11Met	
*1		G	C	T	G	C	A	CTT	G	C	C	
*1xN												
*2		A	G									
*2xN		A	G									
*3				delT								
*3xN				delT								
*4			G		A	T						
*4xN			G		A	T						
*4J/*4P					A	T						
*4K		A	G		A	T						
*4M						T						
*5		delGene	delGene	delGene	delGene	delGene	delGene	delGene	delGene	delGene	delGene	
*6							delA					
*6xN							delA					
*6C			G				delA					
*9								delCTT				
*9x2								delCTT				
*10			G		A							
*10x2			G		A							
*17		A	G						A			
*17x2		A	G						A			
*35		A	G								T	
*41		A	G							T		
*41x2		A	G							T		

Patient ID	Diplotype	rs16947	rs1135840	rs35742686	rs1065852	rs3892097	rs5030655	rs5030656	rs28371706	rs28371725	rs769258	Notes
BBS-001	*1/*4	GG	GC	TT	GA	CT	AA	CTT/CTT	GG	CC	CC	
BBS-002	*1/*41	GA	GC	TT	GG	CC	AA	CTT/CTT	GG	CT	CC	
BBS-003	*1/*41	GA	GC	TT	GG	CC	AA	CTT/CTT	GG	CT	CC	
BBS-004	*1/*1	GG	CC	TT	GG	CC	AA	CTT/CTT	GG	CC	CC	
BBS-005	*1/*1	GG	CC	TT	GG	CC	AA	CTT/CTT	GG	CC	CC	
BBS-006	*1/*9	GG	CC	TT	GG	CC	AA	CTT/delCTT	GG	CC	CC	
BBS-007	*4/*4	GG	GG	TT	AA	TT	AA	CTT/CTT	GG	CC	CC	
BBS-008	*4/*4	GG	GG	TT	AA	TT	AA	CTT/CTT	GG	CC	CC	
BBS-009	*1/*41	GA	CG	TT	GG	CC	AA	CTT/CTT	GG	CT	CC	
BBS-010	*1/*41	GA	CG	TT	GG	CC	AA	CTT/CTT	GG	CT	CC	
BBS-011	*1/*41	GA	CG	TT	GG	CC	AA	CTT/CTT	GG	CT	CC	
BBS-012	*1/*41	GA	CG	TT	GG	CC	AA	CTT/CTT	GG	CT	CC	
BBS-013	*4/*4	GG	GG	TT	AA	TT	AA	CTT/CTT	GG	CC	CC	
BBS-014	*1/*2	GA	CG	TT	GG	CC	AA	CTT/CTT	GG	CC	CC	
BBS-015	*2/*35	AA	GG	TT	GG	CC	AA	CTT/CTT	GG	CC	CT	Astrolabe
BBS-016	*2/*2	AA	GG	TT	GG	CC	AA	CTT/CTT	GG	CC	CC	
BBS-017	*2/*2	AA	GG	TT	GG	CC	AA	CTT/CTT	GG	CC	CC	
BBS-018	*1/*4	GG	CG	TT	GA	CT	AA	CTT/CTT	GG	CC	CC	

Table 5.4 Diplotypes in CYP2D6 in BBS cohort

Patient ID	Diplotype	rs28365083	rs776746	rs56411402	rs55965422	rs10264272	rs41303343	rs55817950	rs28383479
		7:99652613	7:99672916	7:99665237	7:99666950	7:99665212	7:99652770	7:99676198	7:99660516
		g.99652613 G>T p.Thr398Asn	g.99672916 T>C	g.99665237 T>C p.Gln200Arg	g.99666950 A>G	g.99665212 C>T, p.Lys208=	g.99652770_ 71 ins A p.Thr346Tyrfs	g.99676198 G>A p.Arg28Cys	g.99660516 C>T p.Ala337Thr
	*1	G	T	T	A	C	_	G	C
	*2	T	T	T	A	C	_	G	C
	*3	G	C	T	A	C	_	G	C
	*4	G	T	C	A	C	_	G	C
	*5	G	T	T	G	C	_	G	C
	*6	G	T	T	A	T	_	G	C
	*7	G	T	T	A	C	insA	G	C
	*8	G	T	T	A	C	_	A	C
	*9	G	C	T	A	C	_	G	T
BBS-001	*3/*3	GG	CC	TT	AA	CC	_/_	GG	CC
BBS-002	*3/*3	GG	CC	TT	AA	CC	_/_	GG	CC
BBS-003	*3/*3	GG	CC	TT	AA	CC	_/_	GG	CC
BBS-004	*3/*3	GG	CC	TT	AA	CC	_/_	GG	CC
BBS-005	*3/*3	GG	CC	TT	AA	CC	_/_	GG	CC
BBS-006	*3/*3	GG	CC	TT	AA	CC	_/_	GG	CC
BBS-007	*3/*3	GG	CC	TT	AA	CC	_/_	GG	CC
BBS-008	*1/*3	GG	TC	TT	AA	CC	_/_	GG	CC
BBS-009	*3/*3	GG	CC	TT	AA	CC	_/_	GG	CC
BBS-010	*3/*3	GG	CC	TT	AA	CC	_/_	GG	CC
BBS-011	*3/*3	GG	CC	TT	AA	CC	_/_	GG	CC
BBS-012	*3/*3	GG	CC	TT	AA	CC	_/_	GG	CC
BBS-013	*3/*3	GG	CC	TT	AA	CC	_/_	GG	CC
BBS-014	*1/*3	GG	TC	TT	AA	CC	_/_	GG	CC
BBS-015	*3/*3	GG	CC	TT	AA	CC	_/_	GG	CC
BBS-016	*3/*3	GG	CC	TT	AA	CC	_/_	GG	CC
BBS-017	*3/*3	GG	CC	TT	AA	CC	_/_	GG	CC
BBS-018	*3/*3	GG	CC	TT	AA	CC	_/_	GG	CC

Table 5.5 Diplotypes for CYP3A5 in BBS cohort

Patient ID		*CYP4F2*	*VKORC1*	*VKORC1*
	rs12777823	**rs2108622**	**rs9934438**	**rs9923231**
	10:94645745	19:15879621	16:31093557	16:31096368
	g.94645745 G>A	g.15879621 C>T p.Val433Met	g.31093557 G>A	g.31096368 C>T
BBS-001	GA	CT	GG	CC
BBS-002	GG	CT	GG	CC
BBS-003	GA	CC	GA	CT
BBS-004	GA	TT	GG	CC
BBS-005	GG	CC	AA	TT
BBS-006	GG	CT	GG	CC
BBS-007	GG	CT	GA	CT
BBS-008	GG	TT	GG	CC
BBS-009	GG	CT	AA	TT
BBS-010	GG	TT	GA	CT
BBS-011	GG	TT	GA	CT
BBS-012	GG	CC	GA	CT
BBS-013	GG	CT	GA	CT
BBS-014	GA	CC	GG	CC
BBS-015	GA	CC	GG	CC
BBS-016	GG	CC	GA	CT
BBS-017	GG	CC	GA	CT
BBS-018	GG	CC	GA	CT

Table 5.6 Genotypes for rs12777823, CYP4F2 and VKORC1 in BBS cohort

Patient	Diplotype	rs3918290	rs1801159	rs72549303	rs1801158	rs1801160	rs72549309	rs1801266	rs1801265
		1:97450058	1:97515839	1:97450066	1:97515865	1:97305364	1:97740415-18	1:97691776	1:97883329
		g.97450058 C>G	g.97515839 T>C p.Ile543Val	g.97450066 delG p.Pro633Glnfs	g.97515865 C>T p.Ser534Asn	g.97305364 C>T p.Val732Ile	g.97740415_ 18 delATGA p.Phe100Serfs	g.97691776 G>A p.Arg235Trp	g.97883329 A>G p.Cys29Arg
	*1	C	T	G	C	C	ATGA	G	A
	*2A	T	T	G	C	C	ATGA	G	A
	*2B	T	C	G	C	C	ATGA	G	A
	*3	C	T	_	C	C	ATGA	G	A
	*4	C	T	G	T	C	ATGA	G	A
	*5	C	C	G	C	C	ATGA	G	A
	*6	C	T	G	C	T	ATGA	G	A
	*7	C	T	G	C	C	_/_/_	G	A
	*8	C	T	G	C	C	ATGA	A	A
	*9A	C	T	G	C	C	ATGA	G	G
	*9B	C	T	G	C	C	ATGA	G	G
BBS-001	*1/*6	CC	TT	GG	CC	CT	ATGA/ATGA	GG	AA
BBS-002	*5/*9A	CC	TC	GG	CC	CC	ATGA/ATGA	GG	AG
BBS-003	*6/*9A	CC	TT	GG	CC	CT	ATGA/ATGA	GG	AG
BBS-004	*1/*5	CC	TC	GG	CC	CC	ATGA/ATGA	GG	AA
BBS-005	*1/*5	CC	TC	GG	CC	CC	ATGA/ATGA	GG	AA
BBS-006	*1/*1	CC	TT	GG	CC	CC	ATGA/ATGA	GG	AA
BBS-007	*1/*1	CC	TT	GG	CC	CC	ATGA/ATGA	GG	AA
BBS-008	*5/*9A	CC	TC	GG	CC	CC	ATGA/ATGA	GG	AG
BBS-009	*5/*9A	CC	TC	GG	CC	CC	ATGA/ATGA	GG	AG
BBS-010	*1/*1	CC	TT	GG	CC	CC	ATGA/ATGA	GG	AA
BBS-011	*1/*1	CC	TT	GG	CC	CC	ATGA/ATGA	GG	AA
BBS-012	*1/*9A	CC	TT	GG	CC	CC	ATGA/ATGA	GG	AG
BBS-013	*1/*9A	CC	TT	GG	CC	CC	ATGA/ATGA	GG	AG
BBS-014	*1/*1	CC	TT	GG	CC	CC	ATGA/ATGA	GG	AA
BBS-015	*5/*9A	CC	TC	GG	CC	CC	ATGA/ATGA	GG	AG
BBS-016	*1/*9A	CC	TT	GG	CC	CC	ATGA/ATGA	GG	AG
BBS-017	*1/*9A	CC	TT	GG	CC	CC	ATGA/ATGA	GG	AG
BBS-018	*1/*5	CC	TC	GG	CC	CC	ATGA/ATGA	GG	AA

Table 5.7 Diplotypes for DPYD in BBS cohort. rs1801267, rs1801268, rs72539306, rs80081766, rs78060119 and rs55886062 are not shown (all reference sequence)

Patient ID	rs398123546	rs1050828	rs1050829	rs72554665	rs5030868	rs137852327	rs137852339	rs5030869	n/a	n/a	rs5030870
	X:154532390	X:154536002	X:154535277	X:154532269	X:154534419	X:154533122	X:154533044	X:154532990	X:154533615	X:154532992	X:154535316
	Orissa	A(1)	A(2)	Canton	Mediterranean	Viangchan	Kerala	Chatham	Bangkok	Villeurbanne	Sao Boria
	g.154532390 G>A p.Arg484Cys	g.154536002 C>T p.Val98Met	g.154535277 T>A p.Asn156Asp	g.154532269 C>A p.Arg489Leu	g.154534419 G>A p.Ser218Phe	g.154533122 C>T p.Val321Met	g.154533044 C>T p.Glu347Lys	g.154532990 C>T p.Ala365Thr	g.154533615 C>G p.Leu305Asn	g.154532992 TGGdel p.Thr364del	g.154535316 C>T p.Asp143Asn
BBS-001	GG	CC	TT	CC	GG	CC	CC	CC	CC	TGG/TGG	CC
BBS-002	GG	CC	TT	CC	GG	CC	CC	CC	CC	TGG/TGG	CC
BBS-003	GG	CC	TT	CC	GG	CC	CC	CC	CC	TGG/TGG	CC
BBS-004	GG	CC	TT	CC	GG	CC	CC	CC	CC	TGG/TGG	CC
BBS-005	GG	CC	TT	CC	GG	CC	CC	CC	CC	TGG/TGG	CC
BBS-006	GG	CC	TT	CC	GG	CC	CC	CC	CC	TGG/TGG	CC
BBS-007	GG	CC	TT	CC	GG	CC	CC	CC	CC	TGG/TGG	CC
BBS-008	GG	CC	TT	CC	GG	CC	CC	CC	CC	TGG/TGG	CC
BBS-009	GG	CC	TT	CC	GG	CC	CC	CC	CC	TGG/TGG	CC
BBS-010	GG	CC	TT	CC	GG	CC	CC	CC	CC	TGG/TGG	CC
BBS-011	GG	CC	TT	CC	GG	CC	CC	CC	CC	TGG/TGG	CC
BBS-012	GG	CC	TT	CC	GG	CC	CC	CC	CC	TGG/TGG	CC
BBS-013	GG	CC	TT	CC	GG	CC	CC	CC	CC	TGG/TGG	CC
BBS-014	GG	CC	TT	CC	GG	CC	CC	CC	CC	TGG/TGG	CC
BBS-015	GG	CC	TT	CC	GG	CC	CC	CC	CC	TGG/TGG	CC
BBS-016	GG	CC	TT	CC	GG	CC	CC	CC	CC	TGG/TGG	CC
BBS-017	GG	CC	TT	CC	GG	CC	CC	CC	CC	TGG/TGG	CC
BBS-018	GG	CC	TT	CC	GG	CC	CC	CC	CC	TGG/TGG	CC

Table 5.8 Genotypes for G6PD in BBS cohort

Patient ID	*15:02	*15:02	*15:02	*15:02	57*01
	rs2395148	rs10484555	rs144012689	rs2524160	rs2395029
	6:32353777	6:31313688	6:31355003	6:31292077	6:31464003
	g.32353777 G>T	g.31313688 T>C	g.31355003 T>A	g.31292077 G>A	g.31464003 T>G
BBS-001	GG	TT	TT	GG	TT
BBS-002	GG	TT	TT	GG	TT
BBS-003	GG	TT	TT	GG	TT
BBS-004	GG	TT	TT	GG	TT
BBS-005	GG	TT	TT	GG	TT
BBS-006	GG	TT	TT	GG	TT
BBS-007	GG	TT	TT	GG	TT
BBS-008	GG	TT	TT	GA	TT
BBS-009	GG	TT	TT	GG	TT
BBS-010	GG	TT	TT	GG	TT
BBS-011	GG	TT	TT	GG	TT
BBS-012	GG	TT	TT	GG	TG
BBS-013	GG	TT	TT	GA	TG
BBS-014	GG	TT	TT	GG	TG
BBS-015	GG	TT	TT	GG	TT
BBS-016	GG	TT	TT	GA	TT
BBS-017	GG	TT	TT	GA	TT
BBS-018	GG	TT	TT	GG	TT

*Table 5.9 Genotypes for HLA-B *15.02 and HLA-B *57.01 in BBS cohort*

Patient ID	Diplotype	rs1800462	rs1800460	rs1142345	rs1800584	rs72552740	rs75543815	rs72552736	rs56161402	rs151149760	rs72552737
		6:18143724	6:18138997	6:18130687	6:18130781	6:18147910	6:18133845	6:18130725	6:18130762	6:18143606	6:18139027
		g.18143724 C>G p.Ala80Pro	g.18138997 C>T p.Ala154Thr	g.18130687 T>C p.Tyr240Cys	g.18130781 C>T	g.1z8147910 A>G p.Leu49Ser	g.18133845 T>A p.Tyr180Phe	g.18130725 A>C p.His227Gln	g.18130762 C>T p.Arg215His	g.18143606 T>G pLys119Thr	g.18139027 C>G p.Gly114Arg
	*1	C	C	T	C	A	T	A	C	T	C
	*2	G	C	T	C	A	T	A	C	T	C
	*3A	C	T	C	C	A	T	A	C	T	C
	*3B	C	T	T	C	A	T	A	C	T	C
	*3C	C	C	C	C	A	T	A	C	T	C
	*4	C	C	T	T	A	T	A	C	T	C
	*5	C	C	T	C	G	T	A	C	T	C
	*6	C	C	T	C	A	A	A	C	T	C
	*7	C	C	T	C	A	T	C	C	T	C
	*8	C	C	T	C	A	T	A	T	T	C
	*9	C	C	T	C	A	T	A	C	G	C
	*10	C	C	T	C	A	T	A	C	T	G
BBS-001	*1/*3A	CC	CT	TC	CC	AA	TT	AA	CC	TT	CC
BBS-002	*1/*1	CC	CC	TT	CC	AA	TT	AA	CC	TT	CC
BBS-003	*1/*1	CC	CC	TT	CC	AA	TT	AA	CC	TT	CC
BBS-004	*1/*1	CC	CC	TT	CC	AA	TT	AA	CC	TT	CC
BBS-005	*1/*1	CC	CC	TT	CC	AA	TT	AA	CC	TT	CC
BBS-006	*1/*1	CC	CC	TT	CC	AA	TT	AA	CC	TT	CC
BBS-007	*1/*1	CC	CC	TT	CC	AA	TT	AA	CC	TT	CC
BBS-008	*1/*1	CC	CC	TT	CC	AA	TT	AA	CC	TT	CC
BBS-009	*1/*1	CC	CC	TT	CC	AA	TT	AA	CC	TT	CC
BBS-010	*1/*1	CC	CC	TT	CC	AA	TT	AA	CC	TT	CC
BBS-011	*1/*1	CC	CC	TT	CC	AA	TT	AA	CC	TT	CC
BBS-012	*1/*1	CC	CC	TT	CC	AA	TT	AA	CC	TT	CC
BBS-013	*1/*1	CC	CC	TT	CC	AA	TT	AA	CC	TT	CC
BBS-014	*1/*1	CC	CC	TT	CC	AA	TT	AA	CC	TT	CC
BBS-015	*1/*3A	CC	CT	TC	CC	AA	TT	AA	CC	TT	CC
BBS-016	*1/*1	CC	CC	TT	CC	AA	TT	AA	CC	TT	CC
BBS-017	*1/*1	CC	CC	TT	CC	AA	TT	AA	CC	TT	CC
BBS-018	*1/*3A	CC	CT	TC	CC	AA	TT	AA	CC	TT	CC

*Table 5.10 Diplotypes for TMPT in BBS cohort. Results for rs72552738, rs20022021, rs72552742, rs9333569, rs144041067, 6:18149004 and rs777686348 (haplotypes *11-*18) not shown. All patients reference sequence at these SNPs*

Patient ID	Diplotype	rs8175347	rs887829	rs4148323
		2:233760235-233760236 [TA]	2:233759924	2:233760498
		Number of repeats	g.233759924 C>T	g.233760498 G>A p.Gly71Arg
	*1	[TA]6	2 reference or increased function alleles- CC	GG
	*6	[TA]6	1 reference or increased function allele with one decreased function allele- CT	GA
	*28	[TA]7	2 decreased function alleles- TT	GG
	*36	[TA]8	2 reference or increased function alleles- CC	GG
	*37	[TA]5	2 decreased function alleles- TT	GG
BBS-001	*1/*1	[TA]6/[TA]6	CC	GG
BBS-002	*1/*28	[TA]6/[TA]7	CT	GG
BBS-003	*1/*1	[TA]6/[TA]6	CC	GG
BBS-004	*1/*28	[TA]6/[TA]7	CT	GG
BBS-005	*28/*28	[TA]7/[TA]7	TT	GG
BBS-006	*1/*28	[TA]6/[TA]7	CT	GG
BBS-007	*1/*28	[TA]6/[TA]7	CT	GG
BBS-008	*1/*1	[TA]6/[TA]6	CC	GG
BBS-009	*1/*1	[TA]6/[TA]6	CC	GG
BBS-010	*28/*28	[TA]7/[TA]7	TT	GG
BBS-011	*28/*28	[TA]7/[TA]7	TT	GG
BBS-012	*1/*1	[TA]6/[TA]6	CC	GG
BBS-013	*1/*1	[TA]6/[TA]6	CC	GG
BBS-014	*1/*1	[TA]6/[TA]6	CC	GG
BBS-015	*1/*1	[TA]6/[TA]6	CC	GG
BBS-016	*1/*28	[TA]6/[TA]7	CT	GG
BBS-017	*1/*28	[TA]6/[TA]7	CT	GG
BBS-018	*28/*28	[TA]7/[TA]7	TT	GG

Table 5.11 Diplotypes for UGT1A1 in BBS cohort

Patient ID	*F5*	*HLA-A *31:01*	*IFNL3*	*RARG*	*SLC28A3*	*SLCO1B1*	*UGT1A6*
	rs6025	rs1061235	rs12979860	rs2229774	rs7853758	rs4149056	rs17863783
	1:169549811	6:29945521	19:39248147	12:53211761	9:84286011	12:21178615	2:233693631
	g.169549811 C>T p.Arg534Gln	g.29945521 A>T	g.39248147 C>T	g.53211761 G>A p.Ser427Leu	g.84286011 G>A p.Leu461=	g.21178615 T>C p.Val174Ala	g.233693631 G>T p.Val209=
BBS-001	CC	AA	CT	GG	GG	TT	GG
BBS-002	CC	AA	CT	GG	GG	TT	GG
BBS-003	CC	AA	CC	GG	GG	TT	GG
BBS-004	CC	AA	CC	GG	GG	TC	GG
BBS-005	CC	AA	CC	GG	GG	TT	GG
BBS-006	CC	AA	CC	GG	AA	TT	GG
BBS-007	CC	AA	CC	GG	GG	TT	GG
BBS-008	CC	AA	CC	GG	GA	TT	GG
BBS-009	CC	AA	CC	GG	GG	TT	GT
BBS-010	CC	AA	CT	GG	GA	TT	GG
BBS-011	CC	AA	CT	GG	GA	TT	GG
BBS-012	CC	AA	CC	GG	GA	TT	GG
BBS-013	CC	AA	TT	GG	GG	TT	GG
BBS-014	CC	AA	CC	GG	GG	TT	GT
BBS-015	CC	AA	CT	GG	GG	TT	GG
BBS-016	CC	AA	TT	GG	GA	TT	GT
BBS-017	CC	AA	TT	GG	GA	TT	GT
BBS-018	CC	AA	CC	GG	GG	TC	GG

*Table 5.12 Genotypes for F5, HLA-A *31:01, IFNL3, RARG, SLC28A3, SLCO1B1 and UGT1A6 in BBS cohort*

Gene	Diplotype	Phenotype	Agency	Drug	Risk	Recommendation
CYP2C19	*1/*2	Intermediate metaboliser	CPIC	Amitriptyline, clomipramine, doxepin, imipramine, trimipramine	Cardiotoxicity, anticholinergic effects, seizures, altered consciousness, delirium, coma, death	Advice given in conjunction with *CYP2D6* diplotype (*1/*1-normal metaboliser). Prescribe standard dose, monitor for adverse drug effects. If did not know *CYP2D6* diplotype, would give normal dose and monitor more closely for adverse drug effects
CYP2C19	*1/*2	Intermediate metaboliser	CPIC	Citalopram, escitalopram	Emesis, seizures, arrhythmia, reduced level of consciousness, confusion, coma, rarely death	No dose change but monitor for adverse drug effects
CYP2C19	*1/*2	Intermediate metaboliser	DPWG	Citalopram, escitalopram	Emesis, seizures, arrhythmia, reduced level of consciousness, confusion, coma, rarely death	No dose change but monitor for adverse drug effects
CYP2C19	*1/*2	Intermediate metaboliser	CPIC	Clopidogrel	Reduced bioactivation of clopidogrel resulting in reduced levels of active metabolites. Get reduced platelet inhibition and increased residual platelet inhibition. Increased risk for adverse cardiovascular events	Consider using an alternative antiplatelet agent such as prasugrel or ticagrelor
CYP2C19	*1/*2	Intermediate metaboliser	DPWG	Clopidogrel	Reduced bioactivation of clopidogrel resulting in reduced levels of active metabolites. Get reduced platelet inhibition. Increased risk for adverse cardiovascular events	Consider alternative drug e.g. prasugrel which is not metabolised (or possibly little metabolised) by *CYP2C19* but is associated with an increased risk of bleeding
CYP2C19	*1/*2	Intermediate metaboliser	DPWG	Esomeprazole, lansoprazole, omeprazole, pantoprazole	Dry mouth, ptosis, vomiting, sedation, seizures, coma	No dose adjustment but monitor more closely for adverse drug effects
CYP2C19	*1/*2	Intermediate metaboliser	DPWG	Imipramine	Cardiotoxicity, anticholinergic effects, seizures, altered consciousness, delirium, coma, death	Insufficient evidence to calculate dose adjustment. Consider another drug such as mirtazapine or fluvoxamine
CYP2C19	*1/*2	Intermediate metaboliser	DPWG	Moclobemide	Serotonin syndrome (hyperthermia, sweating, agitation, tremor, diarrhoea, dilated pupils), dizziness, headache, dry mouth, nausea	None as insufficient evidence to calculate alternate dosing schedule. Monitor for adverse drug effects
CYP2C19	*1/*2	Intermediate metaboliser	DPWG	Rabeprazole	Dry mouth, ptosis, vomiting, sedation, seizures, coma	Insufficient evidence to make recommendation
CYP2C19	*1/*2	Intermediate metaboliser	DPWG	Sertraline	Emesis, seizures, arrhythmia, reduced level of consciousness	Insufficient data to consider dose adjustment. Monitor closely for adverse drug effects
CYP2C19	*1/*2	Intermediate metaboliser	CPIC	Sertraline	Emesis, seizures, arrhythmia, reduced level of consciousness	No dose adjustment but monitor closely for adverse drug effects
CYP2C19	*1/*2	Intermediate metaboliser	CPIC	Voriconazole	Hepatotoxicity, seizures, visual disturbance, salivation, shortness of breath, weakness, altered level of consciousness	Initiate treatment at standard dose and monitor
CYP2C19	*1/*2	Intermediate metaboliser	DPWG	Voriconazole	Hepatotoxicity, seizures, visual disturbance, salivation, shortness of breath, weakness, altered level of consciousness	Monitor serum levels and be alert for adverse effects

Gene	Diplotype	Phenotype	Agency	Drug	Risk	Recommendation
CYP2C9	*1/*1	Normal metaboliser	DPWG	Acenocoumarol	Risk of sub-therapeutic INR with risk of thrombosis	None
CYP2C9	*1/*1	Normal metaboliser	DPWG	Phenprocoumon	Risk of sub-therapeutic INR with risk of thrombosis	None
CYP2C9	*1/*1	Normal metaboliser	CPIC	Phenytoin	None	As *HLA-B* 15.02 negative, phenytoin at standard dose
CYP2C9	*1/*1	Normal metaboliser	DPWG	Phenytoin	None	None
CYP2C9	*1/*1	Normal metaboliser	DPWG	Sulfonylureas-glibenclamide, gliclazide, glimepiride, tolbutamide	Risk of hyperglycaemia	None
CYP2C9	*1/*1	Normal metaboliser	CPIC	Warfarin	Risk of sub-therapeutic INR with risk of thrombosis	Depends on *VKORC1,* in this case *VKORC1* wild type so dose 5-7mg daily. As CT at rs2108622, could consider 5-10% dose increase. Should use GAGE or IPWC calculator to take account of age, weight etc.
CYP2D6	*1/*4	Normal metaboliser	DPWG	Amitriptyline, clomipramine, doxepin, imipramine, nortriptyline	Normal drug metabolism	Give standard dose
CYP2D6	*1/*4	Normal metaboliser	DPWG	Amitriptyline, clomipramine, doxepin, imipramine, nortriptyline	Normal drug metabolism	Give standard dose
CYP2D6	*1/*4	Normal metaboliser	CPIC	Amitriptyline, clomipramine, doxepin, imipramine, trimipramine	Normal metabolism	Advice given in conjunction with CYP2C19 diplotype (*1/*2-intermediate metaboliser). Prescribe standard dose, monitor for adverse drug effects
CYP2D6	*1/*4	Normal metaboliser	DPWG	Aripiprazole	Normal drug metabolism	Give standard dose
CYP2D6	*1/*4	Normal metaboliser	DPWG	Atomoxetine	Normal drug metabolism	Give standard dose
CYP2D6	*1/*4	Normal metaboliser	DPWG	Carvedilol	Normal drug metabolism	Give standard dose
CYP2D6	*1/*4	Normal metaboliser	DPWG	Clozapine	Normal drug metabolism	Give standard dose
CYP2D6	*1/*4	Normal metaboliser	DPWG	Codeine	Normal drug metabolism	Give standard dose
CYP2D6	*1/*4	Normal metaboliser	CPIC	Codeine	Normal metabolism	Give standard dose
CYP2D6	*1/*4	Normal metaboliser	CPNDS	Codeine	Normal drug metabolism	Give standard dose
CYP2D6	*1/*4	Normal metaboliser	CPIC	Desipramine, fluvoxamine, nortriptyline	Normal metabolism	Give standard dose

Gene	Diplotype	Phenotype	Agency	Drug	Risk	Recommendation
CYP2D6	*1/*4	Normal metaboliser	DPWG	Duloxetine	Normal drug metabolism	Give standard dose
CYP2D6	*1/*4	Normal metaboliser	DPWG	Flecainide	Normal drug metabolism	Give standard dose
CYP2D6	*1/*4	Normal metaboliser	DPWG	Flupenthixol	Normal drug metabolism	Give standard dose
CYP2D6	*1/*4	Normal metaboliser	DPWG	Haloperidol	Normal drug metabolism	Give standard dose
CYP2D6	*1/*4	Normal metaboliser	DPWG	Metoprolol	Normal drug metabolism	Give standard dose
CYP2D6	*1/*4	Normal metaboliser	DPWG	Mirtazapine	Normal drug metabolism	Give standard dose
CYP2D6	*1/*4	Normal metaboliser	DPWG	Olanzapine	Normal drug metabolism	Give standard dose
CYP2D6	*1/*4	Normal metaboliser	CPIC	Ondansetron, tropisetron	Normal drug metabolism	Give standard dose
CYP2D6	*1/*4	Normal metaboliser	DPWG	Oxycodone	Normal drug metabolism	Give standard dose
CYP2D6	*1/*4	Normal metaboliser	CPIC	Paroxetine	Normal drug metabolism	Give standard dose
CYP2D6	*1/*4	Normal metaboliser	DPWG	Paroxetine	Normal drug metabolism	Give standard dose
CYP2D6	*1/*4	Normal metaboliser	DPWG	Propafenone	Normal drug metabolism	Give standard dose
CYP2D6	*1/*4	Normal metaboliser	DPWG	Risperidone	Normal drug metabolism	Give standard dose
CYP2D6	*1/*4	Normal metaboliser	DPWG	Tamoxifen	Normal drug metabolism	Give standard dose
CYP2D6	*1/*4	Normal metaboliser	DPWG	Tramadol	Normal drug metabolism	Give standard dose
CYP2D6	*1/*4	Normal metaboliser	DPWG	Venlafaxine	Normal drug metabolism	Give standard dose
CYP2D6	*1/*4	Normal metaboliser	DPWG	Zuclopenthixol	Normal drug metabolism	Give standard dose
CYP3A5	*3/*3	Poor metaboliser	CPIC	Tacrolimus	Increased chance of achieving therapeutic levels with reduced risk of undertreatment and transplant rejection	Standard dosing and monitoring
CYP3A5	*3/*3	Poor metaboliser	DPWG	Tacrolimus	Reduced chance of achieving therapeutic levels, risk of undertreatment with increased risk of transplant rejection	No dose recommendation. Dose should be adjusted according to therapeutic monitoring
DPYD	*1/*6	Normal metaboliser	DPWG	Fluorouracil, capecitabine	None	Give standard dose
DPYD	*1/*6	Normal metaboliser	CPIC	Fluorouracil, capecitabine, tegafur	None	Give standard dose
DPYD	*1/*6	Normal metaboliser	DPWG	Tegafur	None	Give standard dose
F5	CC, wild type	Normal factor V inactivation. Non-prothrombotic genotype	DPWG	Hormonal contraceptive	No increased risk of thrombotic events	Can prescribe hormonal contraception
G6PD	Wild type, B	Normal levels of G6PD	CPIC	Rasburicase	No increased risk of acute haemolytic anaemia	Rasburicase may be used
HLA-A	*31:01 negative	Normal interaction between carbamazepine and HLA-B	CPNDS	Carbamazepine	No increased risk of carbamazepine hypersensitivity	No need for altered dose of carbamazepine unless patient is also *HLA-B* *15:02 positive. In this case patient negative so can be prescribed at standard dose

Gene	Diplotype	Phenotype	Agency	Drug	Risk	Recommendation
HLA-B	*57:01 negative	Normal interaction with abacavir	CPIC	Abacavir	None	Standard dosing
HLA-B	*57:01 negative	Normal interaction with abacavir	DPWG	Abacavir	None	Standard dosing
HLA-B	*15:02 negative	Normal interaction between carbamazepine and HLA-B	CPIC	Carbamazepine	No increased risk of SCAR including toxic epidermal necrolysis, Stevens-Johnson syndrome, eosinophilia. Also hepatitis and acute renal failure	Carbamazepine may be used
HLA-B	*15:02 negative	Normal interaction between carbamazepine and HLA-B	CPNDS	Carbamazepine	No increased risk of SCAR including toxic epidermal necrolysis, Stevens-Johnson syndrome, eosinophilia. Also hepatitis and acute renal failure	Carbamazepine may be used (patient is *HLA-A* *51:01 negative)
HLA-B	*15:02 negative	Normal interaction between phenytoin and HLA-B	CPIC	Phenytoin	No increased risk of SCAR, toxic epidermal necrolysis, Stevens-Johnson syndrome, eosinophilia, hepatitis, renal failure, ataxia, nystagmus, dysarthria and sedation	Phenytoin may be used and should be initiated at standard dose (*CYP2C9* normal metaboliser)
IFNL3 (IL28B)	CT	Unfavourable genotype	CPIC	Peginterferon α-2a, peginterferon α-2b, Ribavirin	Less likely to respond to treatment but still exposed to risks of treatments	Reduced chance of response when used alone or in combination with protease inhibitor and patients unlikely to be eligible for shortened therapy regimes. Makes prescription less favourable. Weigh up risks/ and benefits
RARG	GG, wild type	Normal risk of AAC	CPNDS	Daunorubicin, doxorubicin	No increased risk of AAC	In paediatric patients only: No high risk alleles in *RARG* or *UTG1A6.*No protective allele in *SLC28A3*. Patient is at moderate risk. Needs increased follow up and echocardiography. Monitor for cardiotoxicity
SLC28A3	GG, wild type	Non-protective allele	CPNDS	Daunorubicin, doxorubicin	Not protective against AAC	In paediatric patients only: No high risk alleles in *RARG* or *UTG1A6.*No protective allele in *SLC28A3*. Patient is at moderate risk. Needs increased follow up and echocardiography. Monitor for cardiotoxicity
SLCO1B1	Wild type	Fast metaboliser	CPIC	Simvastatin	None	None
TPMT	*1/*3A	Intermediate metaboliser	CPIC	6-Mercaptopurine	Reduced risk of insufficient immune-suppression or treatment. Increased risk of fatal myelosuppression and bone marrow toxicity, liver toxicity and risk of discontinuation	Start at 30-70% of normal dose, monitor and wait 2-4 weeks before adjusting
TPMT	*1/*3A	Intermediate metaboliser	DPWG	6-Mercaptopurine	Reduced risk of insufficient immune-suppression or treatment. Increased risk of fatal myelosuppression and bone marrow toxicity, liver toxicity and risk of discontinuation	Select alternative or reduce dose by 50%. Increased monitoring

Gene	Diplotype	Phenotype	Agency	Drug	Risk	Recommendation
TPMT	*1/*3A	Intermediate metaboliser	CPIC	Azathioprine	Reduced risk of insufficient immune-suppression or treatment. Increased risk of fatal myelosuppression and bone marrow toxicity, liver toxicity and risk of discontinuation	Start at 30-70% of normal dose, monitor and wait 2-4 weeks before adjusting
TPMT	*1/*3A	Intermediate metaboliser	DPWG	Azathioprine	Reduced risk of insufficient immune-suppression or treatment. Increased risk of fatal myelosuppression and bone marrow toxicity, liver toxicity and risk of discontinuation	Select alternative or reduce dose by 50%. Increased monitoring
TPMT	*1/*3A	Intermediate metaboliser	CPIC	Thioguanine	Reduced risk of insufficient immune-suppression or treatment. Increased risk of fatal myelosuppression and bone marrow toxicity, liver toxicity and risk of discontinuation	Start at 30-50% of normal dose, monitor and wait 2-4 weeks before adjusting
TPMT	*1/*3A	Intermediate metaboliser	DPWG	Thioguanine	Reduced risk of insufficient immune-suppression or treatment. Increased risk of fatal myelosuppression and bone marrow toxicity, liver toxicity and risk of discontinuation	Alternative drugs should be chosen as evidence if insufficient to adjust dosage
TPMT	*1/*3A	Intermediate metaboliser	CPNDS	Cisplatin	Risk of ototoxicity in **paediatric** patients	Consider alternative drug or use of otoprotectants
UGT1A1	*1/*1	Normal metaboliser	CPIC	Atazanavir	None	None
UGT1A1	*1/*1	Normal metaboliser	DPWG	Irinotecan	None	Standard dosing
UGT1A1	*1/*1	Normal metaboliser	PRO	Irinotecan	None	Standard dosing. Can consider dose intensification to >240mg/m2
UGT1A6	GG, wild type	Normal risk of AAC	CPNDS	Daunorubicin, doxorubicin	No increased risk of AAC	In paediatric patients only: No high risk alleles in *RARG* or *UTG1A6.*No protective allele in *SLC28A3*. Patient is at moderate risk. Needs increased follow up and echocardiography. Monitor for cardiotoxicity
VKORC1	Wild type	Normal expression	DPWG	Acenocoumarol, phenprocoumon	Normal sensitivity to warfarin	None
VKORC1	Wild type	Normal expression	CPIC	Warfarin	Normal sensitivity to warfarin	Advise with *CYP2C9*. As is normal metaboliser and wild type *VKORC1* will require higher dose of 5-7mg daily. As CT at rs2108622, could consider 5-10% dose increase. Dose should be calculated from GAGE or IWPC calculator

Table 5.13 Prescribing advice for BBS-001

Pharmacogenomics Report (1/1)

Name: BBS-001 **Hospital number: BBS-001**
Date of birth: xx/xx/xxxx **Date of report:01/09/2018**

Drugs to be avoided

The patient is a CYP2C19 intermediate metaboliser (*1/*2). The following drugs should be avoided:
Clopidogrel- Consider using an alternative antiplatelet agent such as prasugrel or ticagrelor

Drugs to be used with caution

The patient is a *CYP2C19* intermediate metaboliser (*1/*2). Use the following drugs with caution:

- **Tricyclic antidepressants**- Amitriptyline, clomipramine, doxepin, imipramine, trimipramine. Monitor closely for adverse drug effects. Can start at standard dose as patient is *CYP2D6* normal metaboliser
- **Citalopram, escitalopram**- Monitor closely for adverse drug effects
- **Moclobemide**- Monitor closely for adverse drug effects
- **Proton pump inhibitors-** Esomeprazole, lansoprazole, omeprazole, pantoprazole. Monitor closely for adverse effects. Insufficient evidence to give advice for Rabeprazole
- **Sertraline**- Monitor closely for adverse drug effects
- **Voriconazole**- Start at standard dose, monitor plasma levels and monitor closely for adverse drug effects

The patient has an unfavourable *IFNL3* genotype. Use the following drugs with caution:

- **Anti-virals**- Peginterferon, ribavirin. Patient is less likely to benefit from treatment and will be ineligible for shortened treatment regimes but may still suffer side-effects. Weigh risks and benefits before prescribing

The patient has no high risk alleles in *RARG* or *UGT1A6* and no protective alleles in *SLC28A3*. The following drugs should be used with caution in **paediatric** patients:

- **Daunorubicin, Doxorubicin**- patient should have additional echocardiography, monitoring and follow-up

The patient is a *TPMT* intermediate metaboliser. Use the following drugs with caution:

- **6-Mercaptopurine**- Select alternative drug or reduce dose to 30-70% of recommended. Monitor closely and wait 2-4 weeks before increasing dose
- **Azathioprine**- Select alternative drug or reduce dose to 30-70% of recommended. Monitor closely and wait 2-4 weeks before increasing dose
- **Thioguanine**- Select alternative drug or reduce dose to 30-50% of recommended. Monitor closely and wait 2-4 weeks before increasing dose
- **Cisplatin**- **Paediatric** patients only- use alternative drug or consider otoprotectants

The patient is a *VKORC1* normal expressor and a *CYP2C9* normal metaboliser. Use the following drugs with caution:

- **Warfarin**- the patient is likely to require higher doses of warfarin e.g. 5-7mg per day. Exact warfarin dose should be calculated using a calculator such as IWPC which can be downloaded at https://cpicpgx.org

Drugs to be used as directed

Drugs predominantly metabolised by the following genes with the exception of those mentioned above should be prescribed at standard doses:

- *CYP2C9*
- *CYP2D6*
- *CYP3A5*
- *DPYD*
- *F5*
- *G6PD*
- *HLA-A *31:01*
- *HLA-B *15:02*
- *HLA-B *57:01*
- *SLCO1B1*
- *UGT1A1*

DISCLAIMER: Rare haplotypes may not be identified. Prescribing guidelines may change. For complete list of drugs metabolised by each gene and most recent prescribing guidelines consult www.PharmGKB.org

Figure 5.1 Summary prescribing advice for BBS-001

Gene	Drug	See also
CYP2C19	Tricyclic antidepressants including amitriptyline, clomipramine, doxepin, imipramine, trimipramine	CYP2D6
CYP2C19	Selective serotonin reuptake inhibitors including citalopram, escitalopram, sertraline	CYP2D6
CYP2C19	Clopidogrel	
CYP2C19	Proton pump inhibitors including esomeprazole, lansoprazole, omeprazole, pantoprazole, rabeprazole	
CYP2C19	Moclobemide	
CYP2C19	Voriconazole	
CYP2C9	Coumarin anticoagulants including acenocoumarol, phenprocoumon, warfarin	VKORC1
CYP2C9	Sulfonylureas including glibenclamide, gliclazide, glimepiride, tolbutamide	
CYP2C9	Phenytoin	
CYP2D6	Tricyclic antidepressants including amitriptyline, clomipramine, desipramine, doxepin, fluvoxamine imipramine, trimipramine	CYP2C19
CYP2D6	Atypical antipsychotics including aripiprazole, clozapine, mirtazapine, olanzapine, risperidone	
CYP2D6	Atomoxetine	
CYP2D6	β-blockers including carvedilol, metoprolol	
CYP2D6	Opioid analgesics including codeine, oxycodone, tramadol	
CYP2D6	Serotonin-norepinephrine reuptake inhibitors including duloxetine, venlafaxine	
CYP2D6	Anti-arrhythmics including flecainide, propafenone	
CYP2D6	Typical antipsychotics including flupenthixol, haloperidol, zuclopenthixol	
CYP2D6	2nd generation tricyclic antidepressants including nortriptyline	CYP2C19
CYP2D6	Selective serotonin reuptake inhibitor- paroxetine	CYP2C19
CYP2D6	Tamoxifen	
CYP3A5	Tacrolimus	
DPYD	Fluorpyrimidine chemotherapeutic agents including capecitabine, fluorouracil, Tegafur	
F5	Hormonal contraceptive	
G6PD	Rasburicase	
HLA-A 31:01	Carbamazepine	HLA-B 15:02
HLA-B 15:02	Carbamazepine	HLA-A 31:01
HLA-B 15:02	Phenytoin	
HLA-B 44	Ribavirin	IFNL3
HLA-B 57:01	Abacavir	
HLA-B 58:01	Allopurinol	
IFNL3	Peginterferon α-2a	
IFNL3	Ribavirin	HLA-B 44
RARG	Daunorubicin (paediatric patients only)	SLC28A3, UGT1A6
RARG	Doxorubicin (paediatric patients only)	SLC28A3, UGT1A6
SLC28A3	Daunorubicin (paediatric patients only)	RARG, UGT1A6
SLC28A3	Doxorubicin (paediatric patients only)	RARG, UGT1A6
SLCO1B1	Simvastatin	
TPMT	Thiopurine immunosuppressants including 6-mercaptopurine, azathioprine, thioguanine	
TPMT	Cisplatin (paediatric patients only)	
UGT1A1	Atazanavir	
UGT1A1	Irinotecan	
UGT1A6	Daunorubicin (paediatric patients only)	RARG, SLC28A3
UGT1A6	Doxorubicin (paediatric patients only)	RARG, SLC28A3
VKORC1	Coumarin anticoagulants including acenocoumarol, phenprocoumon, warfarin	CYP2C9

Figure 5.2 Reverse side of summary prescribing sheets

Pharmacogenomics Report (1/2)

Name: BBS-007 **Hospital number: BBS-007**
Date of birth: xx/xx/xxxx **Date of report:01/09/2018**

Drugs to be avoided

The patient is a *CYP2D6* poor metaboliser (*4/*4). The following drugs should be avoided:

- **Opioids-** avoid codeine, hydrocodone, oxycodone and tramadol
- **Metoprolol**- select alternate drug e.g. carvedilol or bisoprolol, especially if treating heart failure. If prescribing metoprolol reduce dose by 75% and be aware of risk of adverse drug effects
- **Paroxetine**- Consider alternative drug not predominantly metabolised by CYP2D6 or reduce dose by 50% and titrate to response
- **Tricyclic antidepressants**- these include amitriptyline, clomipramine, desipramine, doxepin, fluvoxamine, imipramine, nortriptyline. If prescribing tricyclic antidepressants reduce dose by at least 50% and plasma levels monitored (see www.pharmgkb.org for drug-specific guidance)
- **Venlafaxine**- Consider alternative e.g. sertraline, citalopram or if using venlafaxine monitor plasma O-desmethylvenlafaxine levels and adverse drug effects and adjust dose accordingly

Drugs to be used with caution

The patient is a *CYP2D6* poor metaboliser (*4/*4). The following drugs should be used with caution:

- **Aripiprazole**- reduce dose to 67% of recommended daily dose
- **Atomoxetine**- monitor for adverse drug effects
- **Carvedilol**- monitor for adverse drug effects
- **Clozapine-** monitor for adverse drug effects
- **Duloxetine**- monitor for adverse drug effects
- **Flecainide**- Reduce dose by 50% monitor plasma drug levels and ECG
- **Flupenthixol**- monitor for adverse drug effects
- **Haloperidol**- Reduce dose by 50% or consider selecting alternative drug e.g. pimozide, quetiapine, olanzapine, clozapine, fluphenazine
- **Mirtazapine**- monitor for adverse drug effects
- **Olanzapine**- monitor for adverse drug effects
- **Ondansetron/tropisetron-** monitor for adverse drug effects
- **Propafenone**- Reduce dose by 70% and monitor plasma drug levels and ECG
- **Risperidone**- Monitor plasma levels. Monitor for adverse drug effects and titrate to clinical response. Consider selecting alternative drug e.g. clozapine, quetiapine, olanzapine
- **Tamoxifen**- consider aromatase inhibitor in post-menopausal women
- **Zuclopenthixo**l- Reduce dose by 50% and monitor for adverse drug effects. Alternatively consider another antiopsychotic such as flupenthixol, clozapine, olanzapine or quetiapine

The patient has no high risk alleles in *RARG* or *UGT1A6* and no protective alleles in *SLC28A3*. The following drugs should be used with caution in **paediatric** patients:

- **Daunorubicin, Doxorubicin**- patient should have additional echocardiography, monitoring and follow-up

Patient is a *UGT1A1* intermediate metaboliser. Use the following drugs with caution:

- **Atazanavir**- prescribe at standard dose but warn of possible side effects and monitor closely for adverse drug effects
- **Irinotecan**- For standard dosing (180-230mg/m^2) and intensification regimes (>230mg/m^2) there should be rigorous biological and clinical surveillance. Dose may need to be reduced

The patient is a *VKORC*1 intermediate expressor and a *CYP2C9* normal metaboliser. Use the following drugs with caution:

- **Warfarin**- the patient is likely to require higher doses of warfarin e.g. 5-7mg per day. Exact warfarin dose should be calculated using a calculator such as IWPC which can be downloaded at https://cpicpgx.org

DISCLAIMER: Rare haplotypes may not be identified. Prescribing guidelines may change. For complete list of drugs metabolised by each gene and most recent prescribing guidelines consult www.PharmGKB.org

Figure 5.3 Summary prescribing advice for patient BBS-007, page 1 of 2

Pharmacogenomics Report (2/2)

Name: BBS-007
Date of birth: xx/xx/xxxx

Hospital number: BBS-007
Date of report: 01/09/2018

Drugs to be used as directed

Drugs predominantly metabolised by the following genes with the exception of those mentioned above should be prescribed at standard doses:

- *CYP2C19*
- *CYP2C9*
- *CYP3A5*
- *DPYD*
- *F5*
- *G6PD*
- *HLA-A *31:01*
- *HLA-B *15:02*
- *HLA-B *57:01*
- *IFNL3*
- *SLCO1B1*
- *TPMT*

DISCLAIMER: Rare haplotypes may not be identified. Prescribing guidelines may change. For complete list of drugs metabolised by each gene and most recent prescribing guidelines consult www.PharmGKB.org

Figure 5.4 Summary prescribing advice for patient BBS-007, page 2 of 2

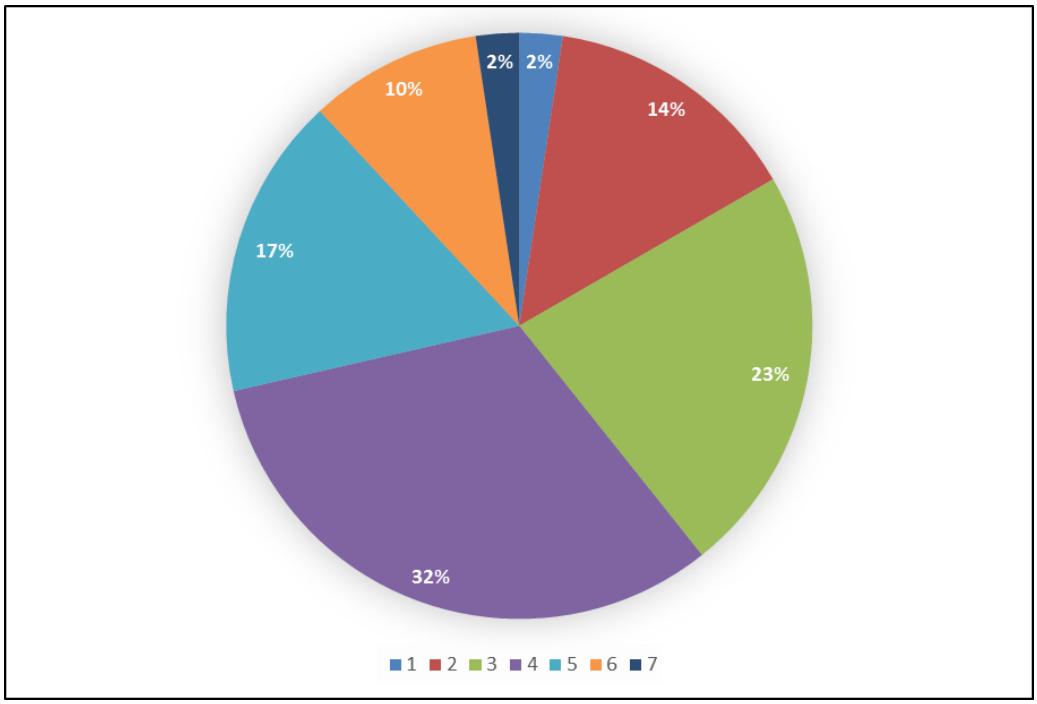

Figure 5.5 Numbers of genes per patient that would require a change in management

5.2.2.2 Present prescribing advice

Information about drugs that patients in the cohorts had been prescribed was extracted from the phenotyping database. There were 41 instances of a patient being prescribed a relevant drug at present or previously (Figure 5.6). 17 were prescribed lansoprazole or omeprazole but only at a prophylactic dose rather than for helicobacter eradication, so the prescribing guidelines were not relevant. The individuals taking codeine, the combined oral contraceptive pill or simvastatin were normal metabolisers for *CYP2D6*, *F5* and *SLCO1B1* respectively. One of the individuals taking amitriptyline and one taking azathioprine had variants in relevant genes (Table 5.14). All other patients prescribed these drugs were normal metabolisers for the relevant genes. The individual taking amitriptyline was on a dose of 17.5mg aged eight when her weight was 23kg (maximum dose 1mg/kg twice daily). It is unknown why this dose was chosen. It was continued for 3.5 years and there is no record of side effects in the notes. The individual on azathioprine was first prescribed it at the age of 7.5 years. Their weight at the time is unknown and there is no record of *TPMT* testing in advance of starting treatment. They were given a dose of 30mg (recommended starting dose 2mg/kg daily). Azathioprine had been discontinued by the time the child was seen 2.5 months later, but there is no reason for this recorded in the notes or letters.

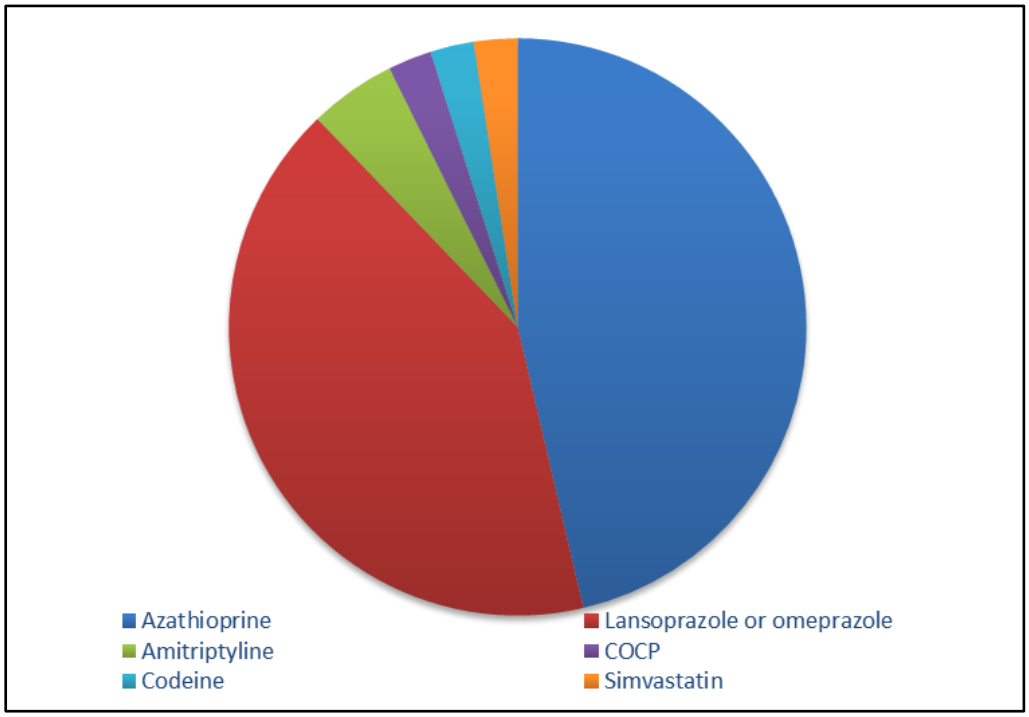

Figure 5.6 Numbers of patients prescribed drugs with relevant prescribing guideline

Patient ID	Drug	Pharmacogenomic prescribing advice
IBD-016	Amitriptyline	*CYP2C19* intermediate metaboliser and *CYP2D6* intermediate metaboliser. Reduce starting dose by 25%, monitor drug levels and for adverse drug effects
IBD-002	Azathioprine	*TPMT* intermediate metaboliser. Start at 30-70% of normal dose, monitor and wait 2-4 weeks before adjusting dose

Table 5.14 Prescribing advice for patients prescribed a drug with variants in the relevant pharmacogene

5.2.3 Haplotype Frequency

The overall haplotype frequencies were determined and compared with published frequencies and the statistical likelihood of obtaining the results was calculated (Tables 5.15 to 5.33). Where possible, UK allele frequencies were used. If unavailable, Western European or European allele frequencies were used. One of each set of monozygotic twins was excluded, as were SRS children, as their diplotypes were inevitably a combination of their parents and were therefore not a random sample. Numbers were rounded to the nearest integer. The cohort diplotype or genotype distribution is represented graphically (Figures 5.7 to 5.25). Again, this excludes one of each pair of monozygotic twins and SRS children.

5.2.3.1 *CYP2C9*

Haplotype	Published European (%)	Published West European (%)	Published British (%)	Haplotype number total	Haplotype total (%)	95% CI (%)	Two proportion z test (z value, p value)
*1	80	80	79	111	77	69-83	z=0.9, p=0.37
*2	12	12	12.5	16	11	6-17	z=0.4, p=0.7
*3	7	7.5	8.5	14	10	6-16	z=0.65, p=0.7
*5	0	n/a	n/a	0	0	n/a	n/a
*6	0	n/a	n/a	0	0	n/a	n/a
*8	<1	n/a	n/a	0	0	n/a	n/a
*9	<1	n/a	n/a	1	1	n/a	n/a
*11	<1	n/a	n/a	1	1	n/a	n/a
*12	<1	n/a	n/a	0	1	n/a	n/a
Total	100	100	100	144	100		

Table 5.15 Haplotype frequency calculation for CYP2C9. European figures are from www.pharmGKB.org, Western European figures from Sistonen et al. and British figures from Stubbins et al.(513, 618)

The p value obtained by performing a Fisher's exact test comparing observed haplotype frequencies to published British haplotype frequencies was 0.8384, so the results were not significantly different from population frequencies(618). For commoner haplotypes, the expected figure was within the 95% confidence interval of the observed and there was no significant difference using a two proportion z-test. The cohort diplotype distribution is shown in Figure 5.7.

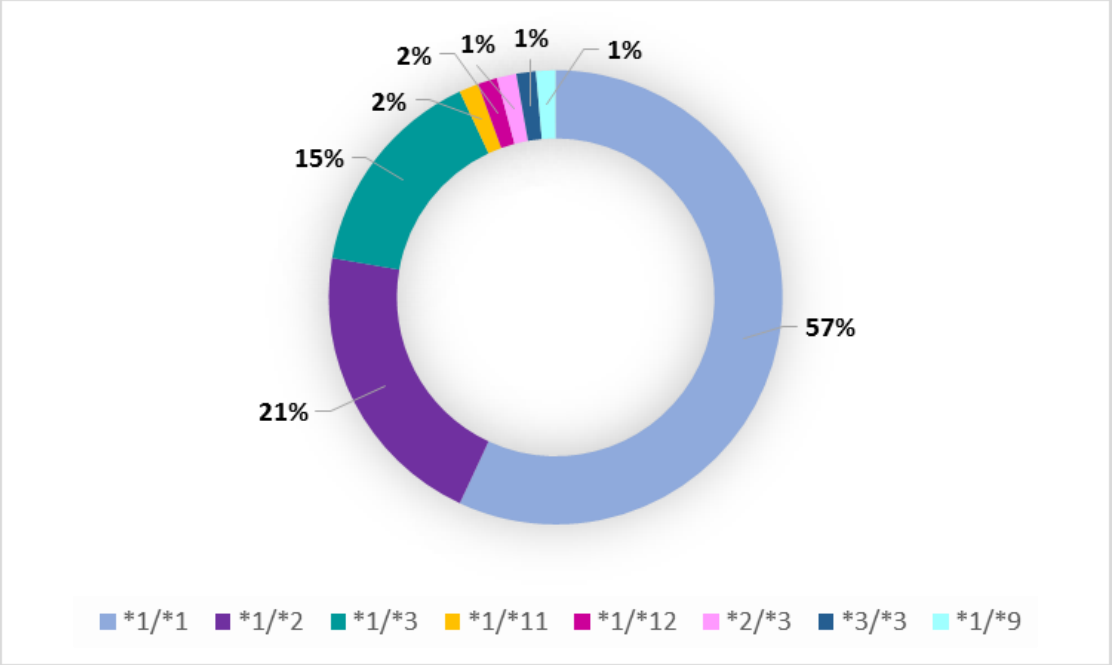

Figure 5.7 Diplotype distribution for CYP2C9

5.2.3.2 *CYP2C19*

Haplotype	Published European (%)	Published West European (%)	Haplotype number total	Haplotype total (%)	95% CI (%)	Two proportion z test (z value, p value)
*1	59	n/a	85	59	51-67	z=0, p=1
*2	18	14	34	24	13-26	Z=1.4, p=0.2
*3	0	<1	0	0	n/a	n/a
*4A	0	n/a	0	0	n/a	n/a
*4B	0	n/a	0	0	n/a	n/a
*5	0	n/a	0	0	n/a	n/a
*6	0	n/a	0	0	n/a	n/a
*7	0	n/a	0	0	n/a	n/a
*8	0	n/a	0	0	n/a	n/a
*15	0	n/a	1	1	n/a	n/a
*17	22	n/a	24	17	11-24	z=1.448, p=0.2
Total	100	n/a	144	100		

Table 5.16 Haplotype frequency calculation for CYP2C19. European figures are from Zhou et al., Western European figures from Sistonen et al.(513, 619)

The p value obtained by performing a Fisher's exact test comparing observed haplotype frequencies to published British haplotype frequencies was 0.4743, so the results were not significantly different from population frequencies(619). For commoner haplotypes, the expected figure was within the 95% confidence interval of the observed and there was no significant difference using a two proportion z-test. The cohort diplotype distribution is shown in Figure 5.8.

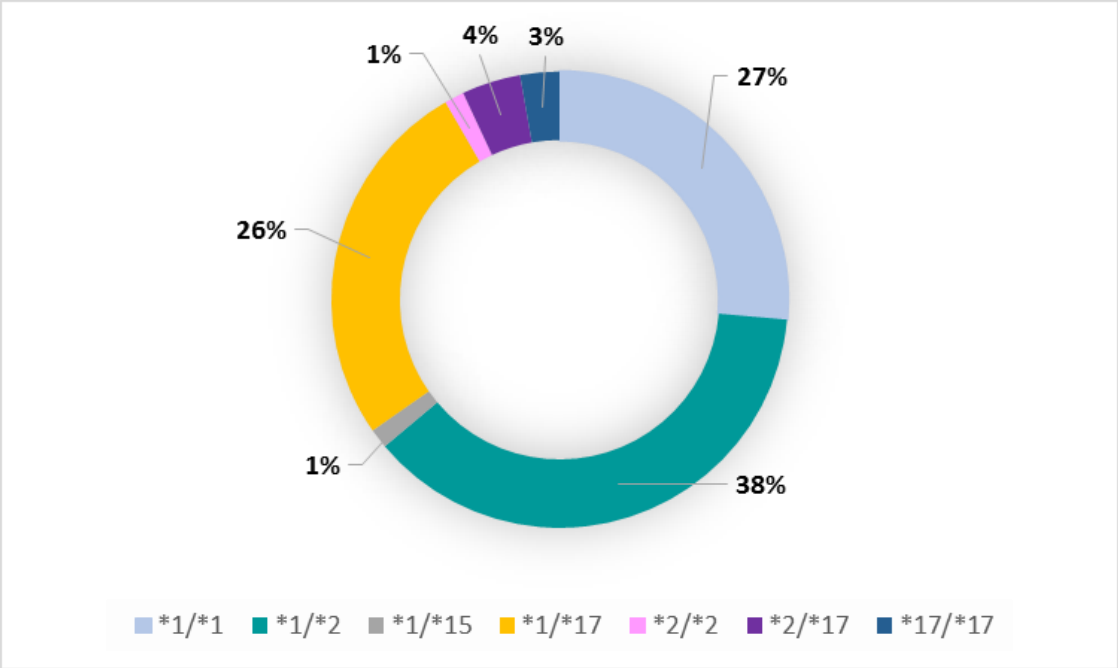

Figure 5.8 Diplotype distribution for CYP2C19

5.2.3.3 *CYP2D6*

Haplotype	Published European (%)	Haplotype number total	Haplotype total (%)	95% CI (%)	Two proportion z test (z value, p value)
*1	37	54	38	30-46	z=0.3, p=0.8
*2	26	28	20	14-27	z=1.6, p=0.1
*3	1	1	1	0.1-4	z=0, p=1
*4	18	23	16	10-23	z= 0.6, p=0.5
*5	3	1	1	0.1-4	z=1.4 , p=0.2
*6	1	1	1	0.1-4	z=0, p=1
*9	2	5	3	0.7-7	z=0.8 , p=0.4
*10	3	2	1	0.7-4	z=1.4 , p=0.2
*17	<1	1	1	n/a	n/a
*29	<1	1	1	n/a	n/a
*35	2	2	1	0.7-4	z=0.9 , p=0.4
*41	7	21	14	8-20	z=3.3 , p=0.001
*1N	1	3	2	0.4-6	z=1.2 , p=0.2
*2N	1	0	0	n/a	n/a
*41N	0	1	0	n/a	n/a
Total	100	144	100		

Table 5.17 Haplotype frequency calculation for CYP2D6. European figures are from www.pharmGKB.org(471)

The p value obtained by performing a Fisher's exact test comparing observed haplotype frequencies to published European haplotype frequencies was 0.501, so the results were not significantly different to population frequencies(471). With the exception of *41, for commoner haplotypes, the expected figure was within the 95% confidence interval of the observed and there was no significant difference using a two proportion z-test. *41 was within the 95% confidence interval, but the z-test score showed the results were significantly different. This is discussed in section 5.3.3.3.2. The cohort diplotype distribution is shown in Figure 5.9.

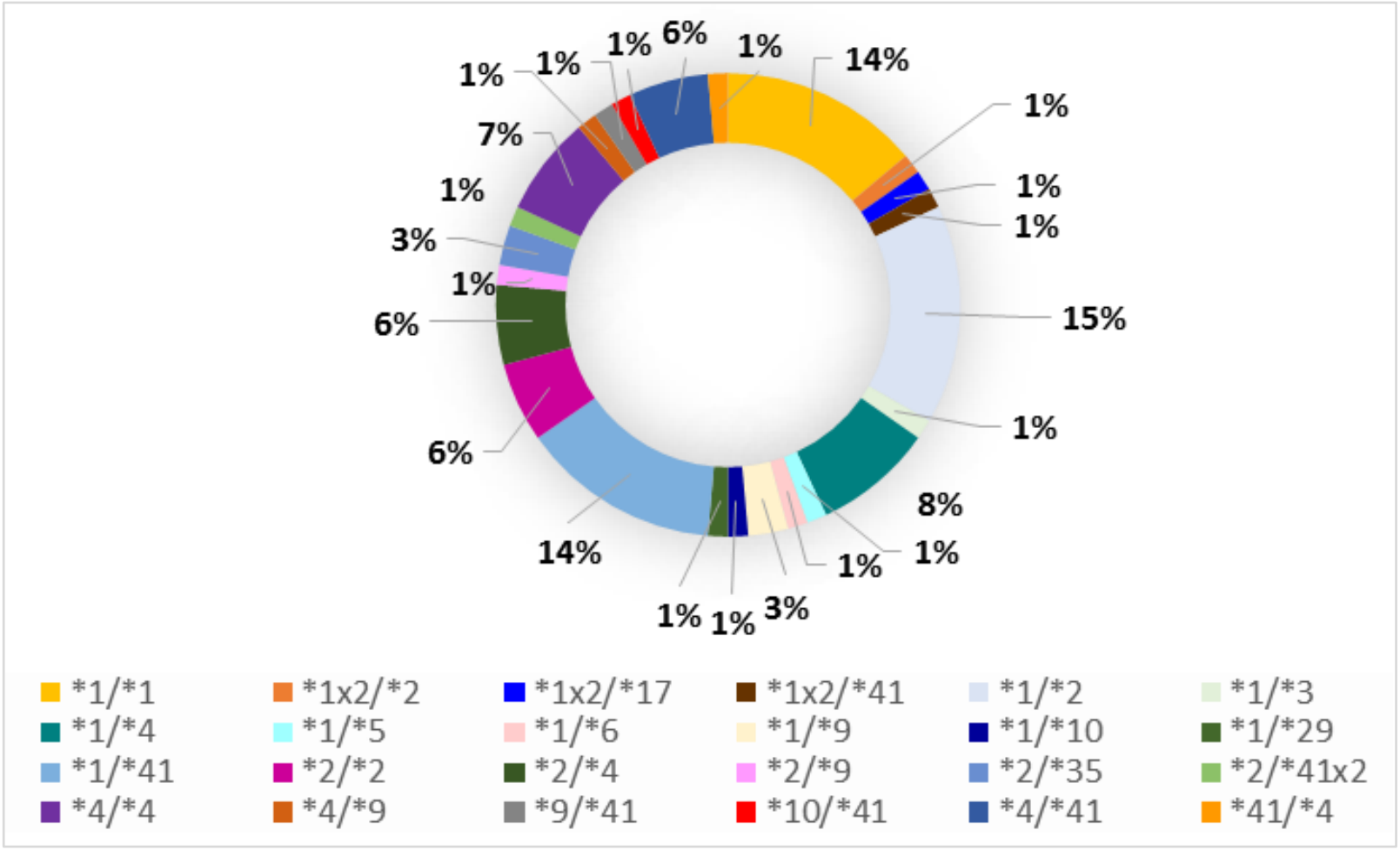

Figure 5.9 Diplotype distribution for CYP2D6

5.2.3.4 *CYP3A5*

Haplotype	Published European (%)	Published British (%)	Haplotype number total	Haplotype total (%)	95% CI (%)	Two proportion z test (z value, p value)
*1	8	6	11	8	4-14	z=0.9, p=0.4
*2	1	0	0	0	n/a	n/a
*3	91	94	130	90	88-97	z=1.6 , p=0.1
*4	n/a	n/a	0	0	n/a	n/a
*5	n/a	n/a	0	0	n/a	n/a
*6	0	n/a	3	2	n/a	n/a
*7	n/a	n/a	0	0	n/a	n/a
*8	n/a	n/a	0	0	n/a	n/a
*9	n/a	n/a	0	0	n/a	n/a
Total	100	100	144	100		

Table 5.18 Haplotype frequency calculation for CYP3A5.European figures are from www.pharmGKB.org, British figures from King et al.(620)

The p values obtained by performing a Fisher's exact test comparing observed haplotype frequencies to published European and British haplotype frequencies were 0.5426 and 0.3305 respectively so the results were not significantly different to population frequencies(486, 620). The British figures did not include the *6 haplotype. For commoner haplotypes, the expected figure was within the 95% confidence interval of the observed and there was no significant difference using a two proportion z-test. The cohort diplotype distribution is shown in Figure 5.10.

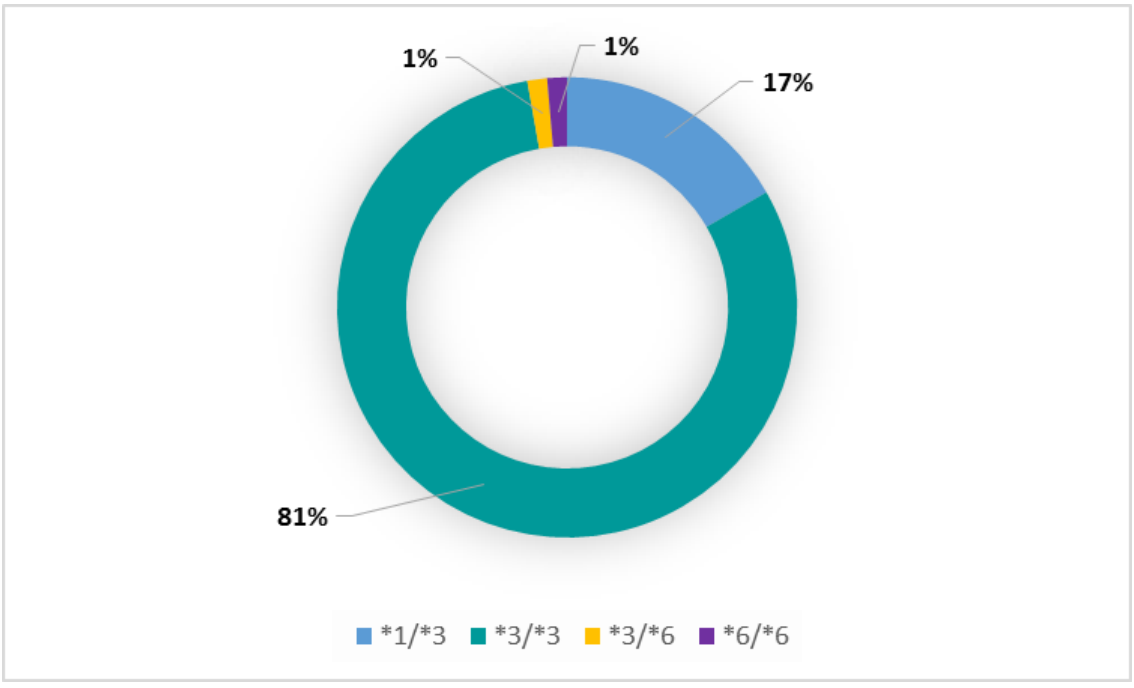

Figure 5.10 Diplotype distribution for CYP3A5

5.2.3.5 *CYP4F2*

Allele	Published European (%)	Allele total	Allele frequency (%)	95% CI (%)	Two proportion z test (z value, p value)
C	70	103	72	64-79	z=0.5, p=0.6
T	30	41	28	21-36	z=0.5, p=0.6
Total	100	144	100		

Table 5.19 Allele frequency calculation for CYP4F2. European figures are from Ross et al.(584)

The p value obtained by performing a Fisher's exact test comparing observed allele frequencies to published European allele frequencies was 0.7583, so the results were not significantly different to population frequencies(584). For both alleles, the expected figure was within the 95% confidence interval of the observed and there was no significant difference using a two proportion z-test. The cohort genotype distribution is shown in Figure 5.11.

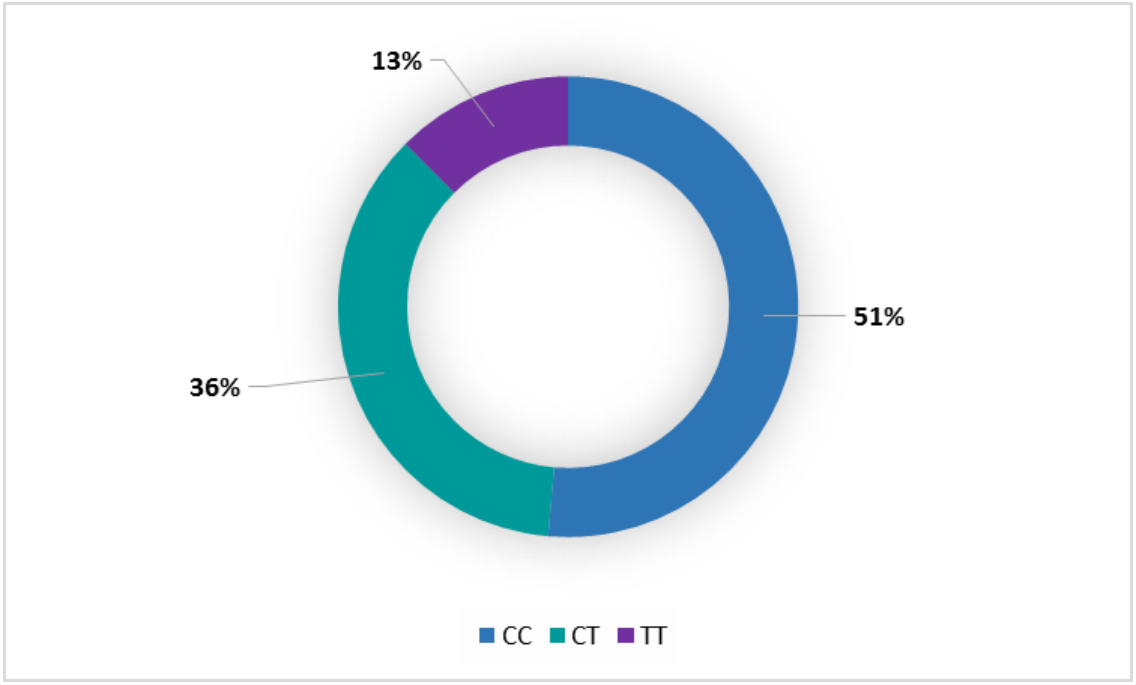

Figure 5.11 Genotype distribution for CYP4F2

5.2.3.6 *DPYD*

Haplotype	Published European (%)	Haplotype number total	Haplotype total (%)	95% CI (%)	Two proportion z test (z value, p value)
*1	60	80	58	50-66	z=0.5, p=0.6
*2A	<1	0	0	n/a	n/a
*2B	n/a	0	0	n/a	n/a
*3	0	0	0	n/a	n/a
*4	2	1	1	0.1-4	z=0.58, p=0.4
*5	15	27	20	14-28	z=1.6, p=0.1
*6	4	3	2	0.4-6	z=1.2, p=0.2
*7	<1	0	0	n/a	n/a
*8	n/a	0	0	n/a	n/a
*9A	18	27	20	14-28	z=0.6, p=0.1
*9B	n/a	0	0	n/a	n/a
*10	n/a	0	0	n/a	n/a
*11	n/a	0	0	n/a	n/a
*12	0	0	0	n/a	n/a
*13	<1	0	0	n/a	n/a
Total	100	138	100		

Table 5.20 Haplotype frequency calculation for DPYD. European figures are from Caudle et al.(621)

The p value obtained by performing a Fisher's exact test comparing observed haplotype frequencies to published European haplotype frequencies was 0.765, so the results were not significantly different to population frequencies(621). The total haplotype number was smaller because the three individuals with uninterpretable haplotypes were excluded. The cohort diplotype distribution is shown in Figure 5.12.

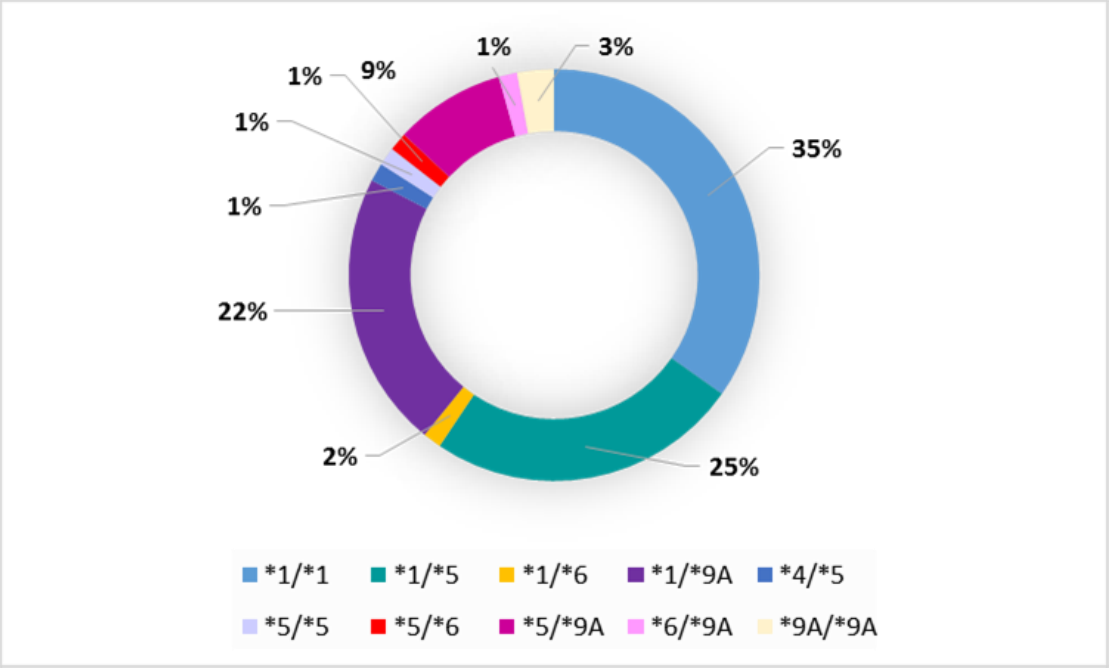

Figure 5.12 Diplotype distribution for DPYD

5.2.3.7 *F5*

Allele	Published White British (%)	Allele total	Allele frequency (%)	95% CI (%)	Two proportion z test (z value, p value)
C	96	141	98	94-99	z=1.2, 0.2
T	4	3	2	0.4-6	z=1.2, 0.2
Total	100	144	100		

Table 5.21 Allele frequency calculation for F5. UK figures are from Rees et al.(622)

The p value obtained by performing a Fisher's exact test comparing observed allele frequencies to published British allele frequencies was 0.4472 so the results were not significantly different to population frequencies(622). Figures from this paper included only people identified as white British, while other, smaller samples of unselected Britons gave lower allele frequencies(542). Using the minor allele frequency of 1.6% from the Beauchamp paper gives a p value of >0.9999. For both alleles, the expected figure was within the 95% confidence interval of the observed and there was no significant difference using a two proportion z-test. The cohort genotype distribution is shown in Figure 5.13.

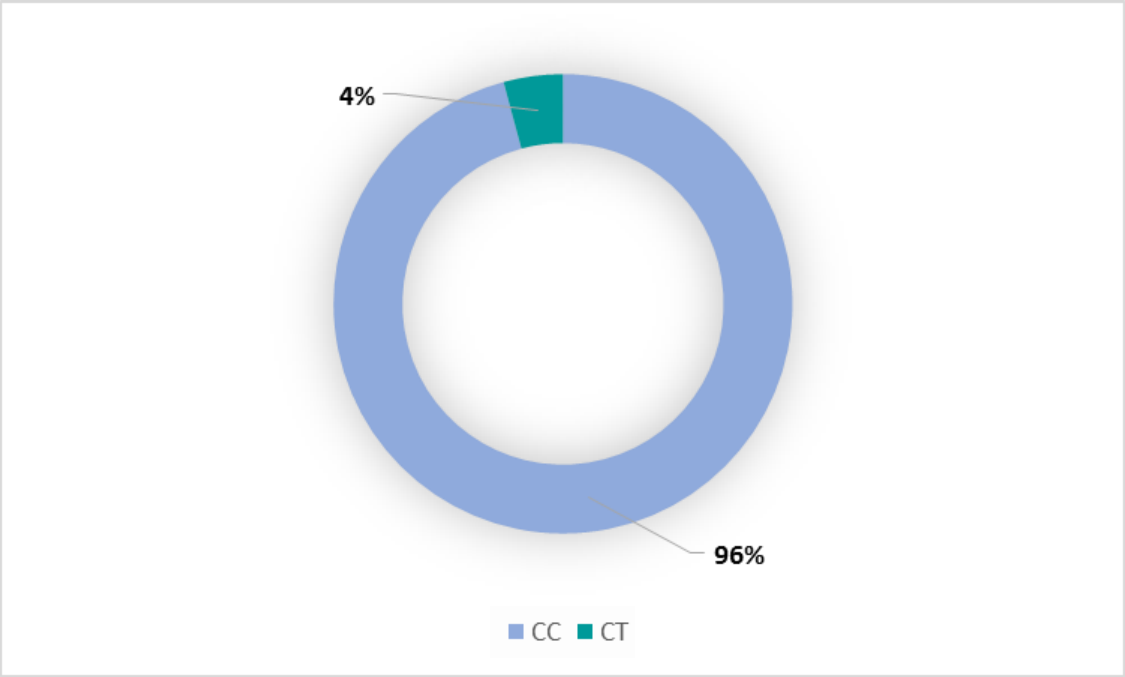

Figure 5.13 Genotype distribution for F5

5.2.3.8 *G6PD*

Clinical status	Published clinical status UK (%)	Clinical status total number	Clinical status total (%)	95% CI (%)	Two proportion z test (z value, p value)
Unaffected	100	72	72	n/a	n/a
Affected	0	0	0	n/a	n/a
Total	100	72	100		

Table 5.22 Haplotype frequency calculation for G6PD. UK figures are from Erdohazi et al.*(623)*

There were 11 patients from high-risk populations: four African, one Middle Eastern and four Asian females and two Asian males. One African female was homozygous for the African (2) (A(2)) allele, but did not have any of the additional SNPs required to make this a pathogenic variant (section 5.3.2.3.3) so did not have G6PD deficiency. A further two females were listed as heterozygotes. One, of Middle Eastern ethnicity, was heterozygote for the Mediterranean SNP, and the other, whose ethnicity was listed as African, was heterozygote for the A(2) allele, but again, did not have any of the additional SNPs. The prevalence of G6PD deficiency in the cohort was zero. Few published figures are available for the UK. Erdohazi et al. conducted a study of 204 patients of British and Irish origin and found the prevalence of G6PD deficiency to be zero(623). This study did not look at individuals who might carry a deficient allele but be clinically unaffected and figures are not available for this in the UK population. Using this population frequency and the finding of zero affected individuals in the control cohort, the p value using Fisher's exact test is >0.9999. Confidence intervals and z-scores were not calculated in this case as the percentages were 100 and zero. Mean allele frequency in malaria-endemic countries has been estimated at 8%, but may be as high as 30% in some regions(624). Given the small number of individuals from high risk areas and the range of allele frequencies, it was not possible to compare allele frequencies more exactly. The cohort diplotype distribution is shown in Figure 5.14.

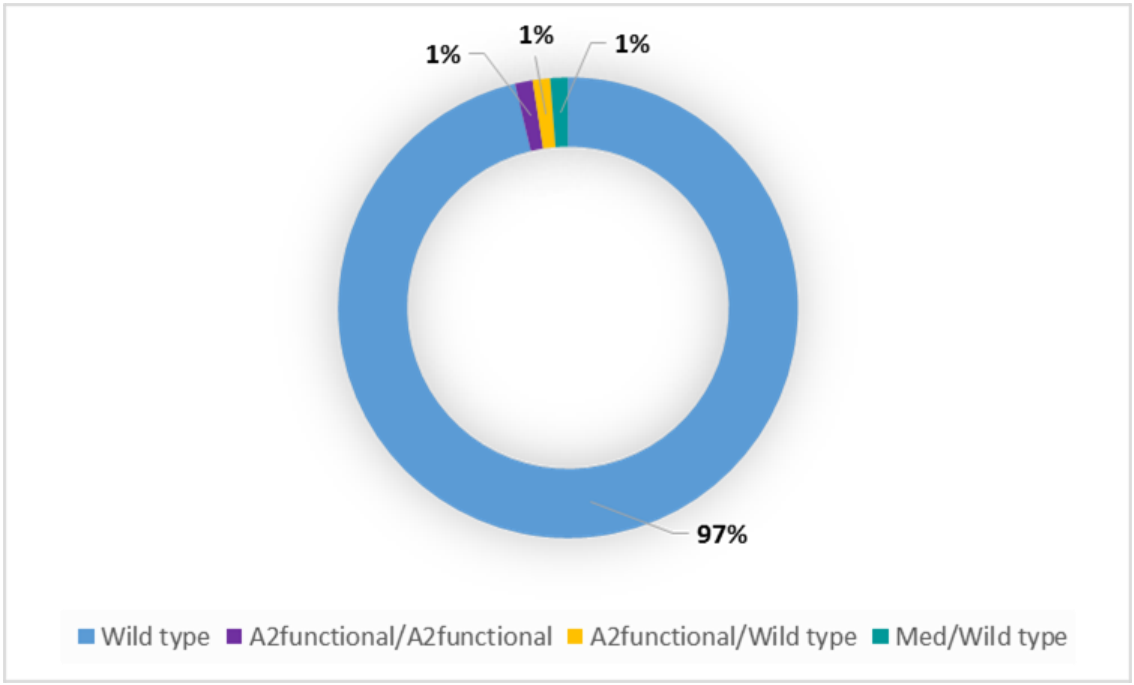

Figure 5.14 Diplotype distribution for G6PD

5.2.3.9 *HLA-A *31:01*

Genotype	Published clinical status European (%)	Published clinical status UK (%)	Clinical status number	Clinical status frequency %	95% CI (%)	Two proportion z test (z value, p value)
Negative (AA)	95-97%	94	67	93	84-98	z=0.4, p=0.7
Positive (AT/TT)	3-5%	6	5	7	2-15	z=0.4, p=0.7
Total	100	100	72	100		

*Table 5.23 Genotype frequency calculation for HLA-A *31:01. UK population frequencies from www.allelefrequencies.net(562)*

The p value obtained by performing a Fisher's exact test comparing observed allele frequencies to published British allele frequencies was 0.784 so the results were not significantly different to UK population frequencies(562). For both groups, the expected figure was within the 95% confidence interval of the observed and there was no significant difference using a two proportion z-test. The cohort haplotype distribution is shown in Figure 5.15.

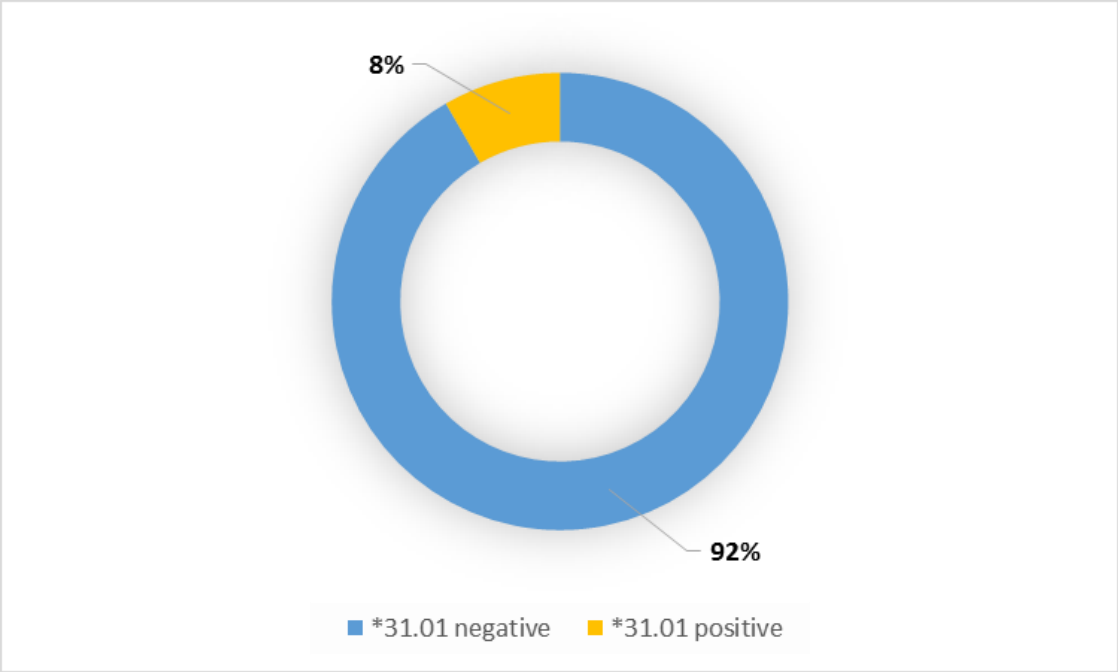

*Figure 5.15 Haplotype distribution for HLA-A *31:01*

5.2.3.10 *HLA-B* *15:02

Clinical status	Published clinical status England (%)	Clinical status number	Clinical status frequency %	95% CI (%)	Two proportion z test (z value, p value)
*15:02 Low risk	>99	72	100	n/a	n/a
*15:02 High risk	<1	0	0	n/a	n/a
Total	100	72	100		

*Table 5.24 Genotype frequency calculation for HLA-B *15:02. UK population frequencies from www.allelefrequencies.net(562)*

For HLA-B *15:02, a single positive SNP was not considered enough to call the haplotype as in the papers used to determine the SNPs, pairs of SNPs are used(625, 626). Therefore, no individuals positive for the *HLA-B *15:02* allele were seen in any of the cohorts. This is not unexpected as rates are highest in Thai and Han Chinese populations and none of the patients were identified as having Far Eastern ethnicity. The p value obtained by performing a Fisher's exact test comparing observed allele frequencies to published British allele frequencies was >0.9999 using the frequency of 0.2% of individuals having the *15:02 allele (English blood donors, mixed ethnicity) so the results were not significantly different to UK population frequencies(562). Confidence intervals and z-scores were not calculated in this case as the percentages were 100 and zero. The cohort haplotype distribution is shown in Figure 5.16.

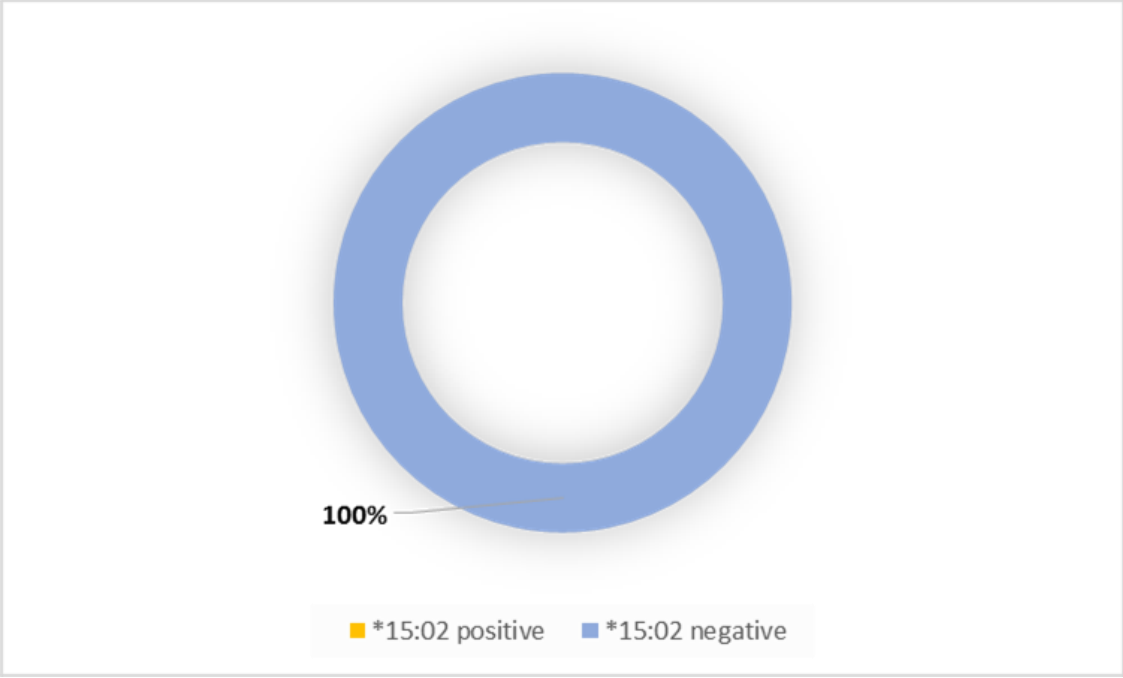

*Figure 5.16 Haplotype distribution for HLA-B *15:02*

5.2.3.11 *HLA-B* *57:01

Clinical status	Published clinical status UK (%)	Clinical status number	Clinical status frequency %	95% CI (%)	Two proportion z test (z value, p value)
*57:01 Low risk	91	67	93	84-98	z=0.6, p=0.5
*57:01 High risk	9	5	7	2-15	z=0.6, p=0.5
Total	100	72	100		

*Table 5.25 Genotype frequency calculation for HLA-B *57:01. UK population frequencies from www.allelefrequencies.net(562)*

The p value obtained by performing a Fisher's exact test comparing observed allele frequencies to published British allele frequencies was 0.6153 using the frequency of 8.9% of individuals having the *57:01 allele (English blood donors, mixed ethnicity) so the results were not significantly different to UK population frequencies(562). For both groups, the expected figure was within the 95% confidence interval of the observed and there was no significant difference using a two proportion z-test. The cohort haplotype distribution is shown in Figure 5.17.

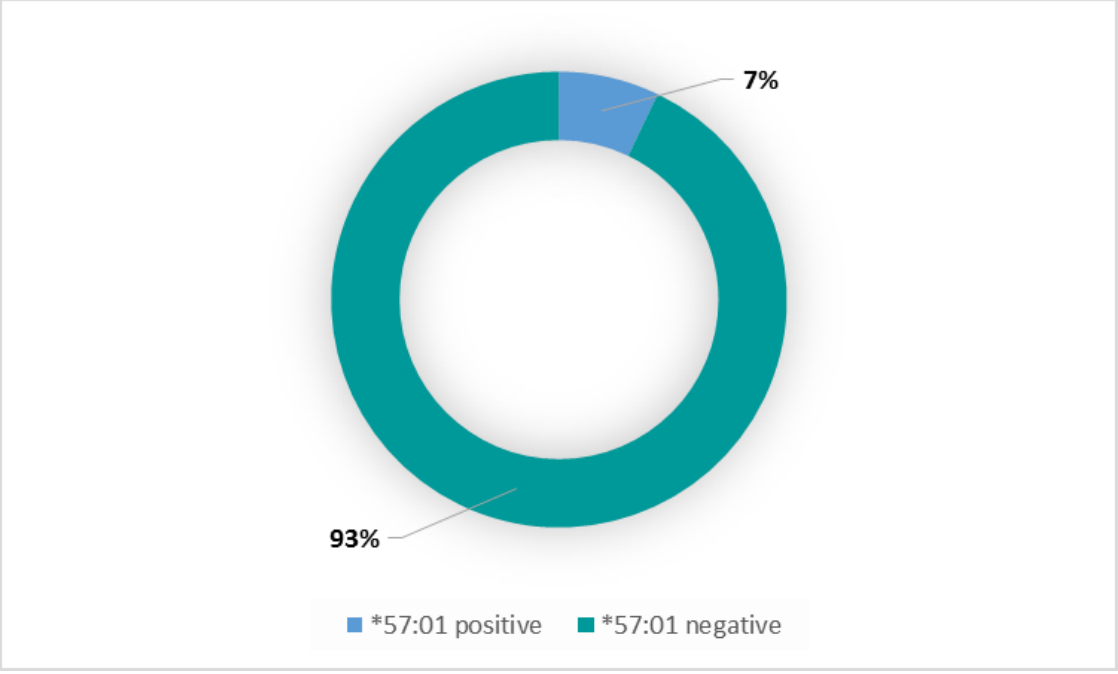

*Figure 5.17 Haplotype distribution for HLA-B *57:01*

5.2.3.12 *IFNL3*

Allele	Published Caucasian (%)	Allele total	Allele frequency (%)	95% CI (%)	Two proportion z test (z value, p value)
C	63	99	69	60-76	z=1.4, p=0.1
T	37	45	31	24-39	z=1.4, p=0.1
Total	100	144	100		

Table 5.26 Allele frequency calculation for IFNL3.Caucasian population frequency from Muir et al.(565)

The p value obtained by performing a Fisher's exact test comparing observed allele frequencies to published Caucasian allele frequencies was 0.3758, so the results were not significantly different to population frequencies(565). For both alleles, the expected figure was within the 95% confidence interval of the observed and there was no significant difference using a two proportion z-test. No UK or mixed European ethnicity frequencies were available. The cohort genotype distribution is shown in Figure 5.18.

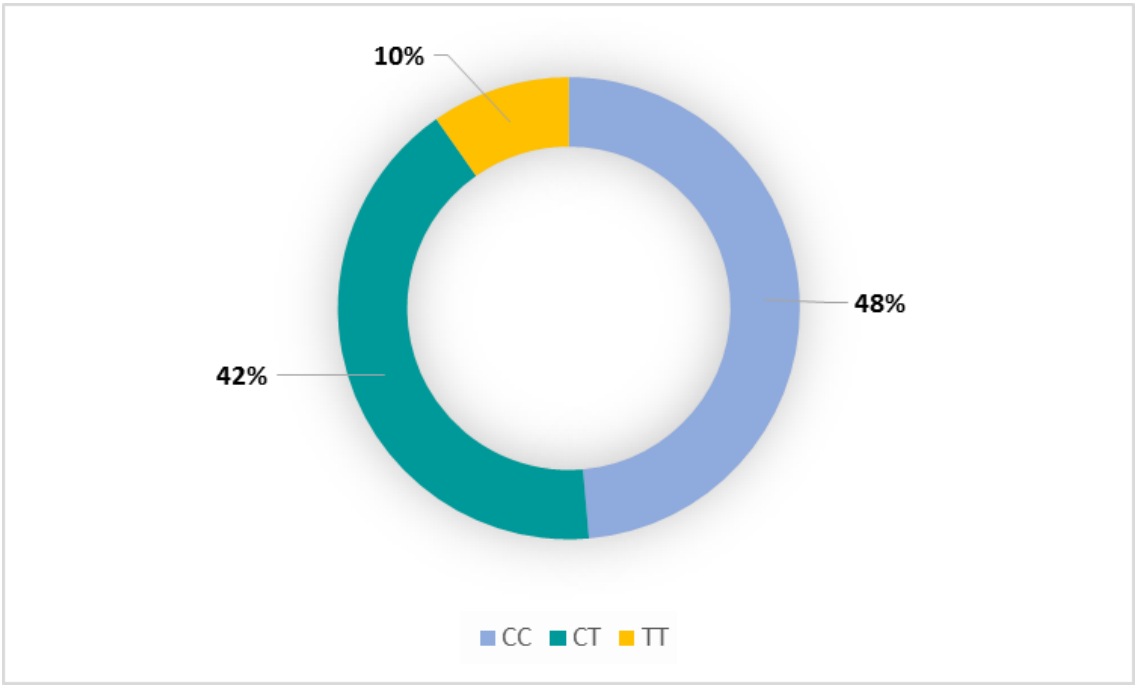

Figure 5.18 Genotype distribution for IFNL3

5.2.3.13 *RARG*

Allele	Published European (%)	Allele Total	Allele Frequency (%)	95% CI (%)	Two proportion z test (z value, p value)
C	92	135	94	87-96	z=0.6, p=0.4
T	8	9	6	3-11	z=0.6, p=0.4
Total	100	144	100		

Table 5.27 Allele frequency calculation for RARG. Caucasian population frequency from Aminkeng et al.(627)

The p value obtained by performing Fisher's exact test comparing observed allele frequencies to published Caucasian allele frequencies was 0.5948, so the results were not significantly different to population frequencies(627). For both alleles, the expected figure was within the 95% confidence interval of the observed and there was no significant difference using a two proportion z-test. The cohort genotype distribution is shown in Figure 5.19.

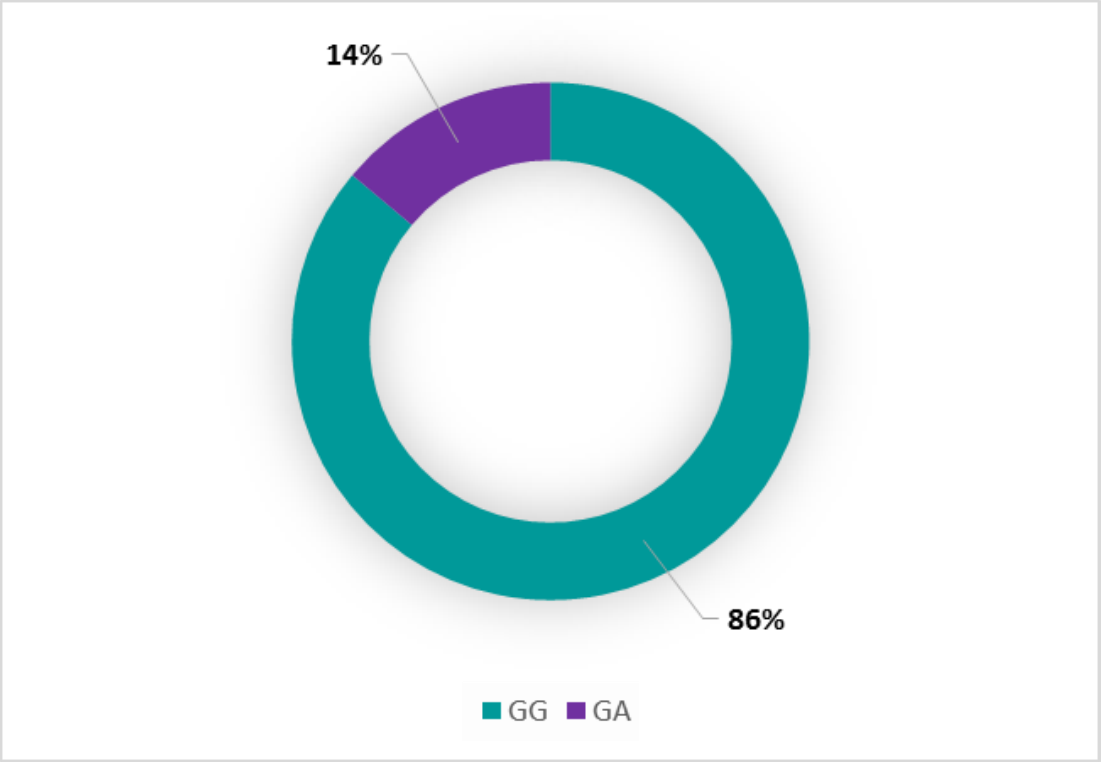

Figure 5.19 Genotype frequencies for RARG

5.2.3.14 *SLC28A3*

Allele	ExAC (%)	Allele total	Allele frequency (%)	95% CI (%)	Two proportion z test (z value, p value)
G	87	118	82	75-88	z=1.8, p=0.08
A	13	26	18	12-25	z=1.8, p=0.08
Total	100	144	100		

Table 5.28 Allele frequency calculation for SLC28A3. *European population frequency from ExAC database(77)*

The p value obtained by performing a Fisher's exact test comparing observed allele frequencies to published European allele frequencies was 0.3378, so the results were not significantly different to population frequencies(77). For both alleles, the expected figure was within the 95% confidence interval of the observed and there was no significant difference using a two proportion z-test. Published frequencies compared cardiomyopathy and non-cardiomyopathy control groups, so ExAC population data were chosen instead. The cohort genotype distribution is shown in Figure 5.20.

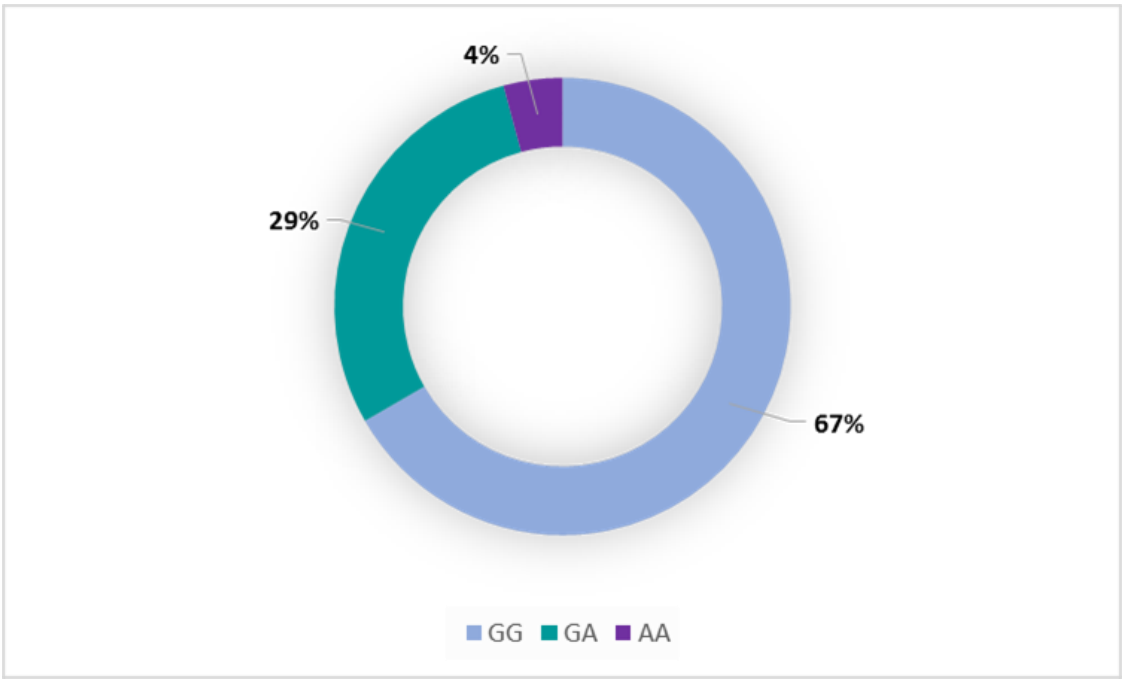

Figure 5.20 Genotype distribution for SLC28A3

5.2.3.15 *SLCO1B1*

Allele	ExAC (%)	Allele total	Allele frequency (%)	95% CI (%)	Two proportion z test (z value, p value)
T	84	128	89	83-94	z=1.6, p=0.1
C	16	16	11	6-17	z=1.6, p=0.1
Total	100	144	100		

Table 5.29 Allele frequency calculation for SLCO1B1. European population frequency from ExAC database(77)

The p value obtained by performing a Fisher's exact test comparing observed allele frequencies to published European allele frequencies was 0.3111, so the results were not significantly different to population frequencies(77). For both alleles, the expected figure was within the 95% confidence interval of the observed and there was no significant difference using a two proportion z-test. The ExAC figures agree with the figures from Oshiro et al. which gives a minor allele frequency of 12-20%(480). The cohort genotype distribution is shown in Figure 5.21. This SNP represents the reduced function haplotypes *15 and *17 as well as the rarer reduced function *5 haplotype. This is discussed further in section 5.3.3.3.7 and in Chapter 6.

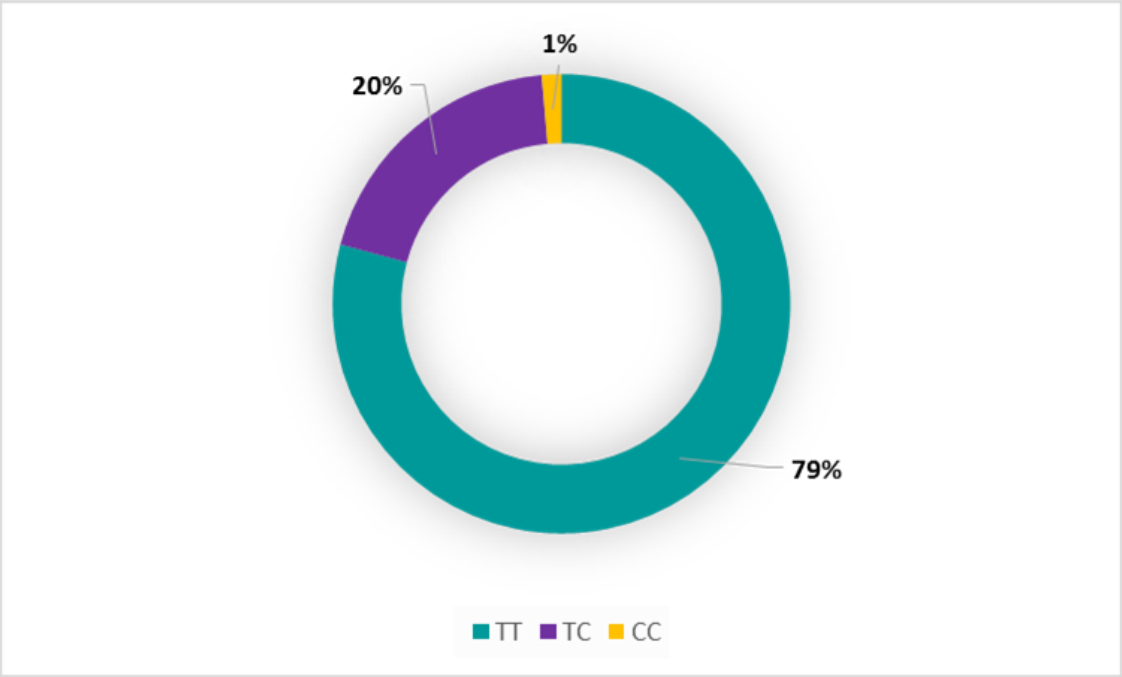

Figure 5.21 Genotype distribution for SLCO1B1

5.2.3.16 *TPMT*

Haplotype	Published European (%)	Haplotype number total	Haplotype total (%)	95% CI (%)	Two proportion z test (z value, p value)
*1	95	136	94	89-97	z=0.6, p=0.6
*2	<1	0	0	n/a	n/a
*3A	5	8	6	3-11	z=0.6, p=0.6
*3B	0	0	0	n/a	n/a
*3C	<1	0	0	n/a	n/a
*4	0	0	0	n/a	n/a
Total	100	144	100		

Table 5.30 Haplotype frequency calculation for TPMT. European figures are from Collie-Duguid et al.(628)

The p value obtained by performing a Fisher's exact test comparing observed haplotype frequencies to published British Caucasian haplotype frequencies was 0.7686, so the results were not significantly different to population frequencies(628). For the observed haplotypes, the expected figure was within the 95% confidence interval of the observed and there was no significant difference using a two proportion z-test. The cohort diplotype distribution is shown in Figure 5.22.

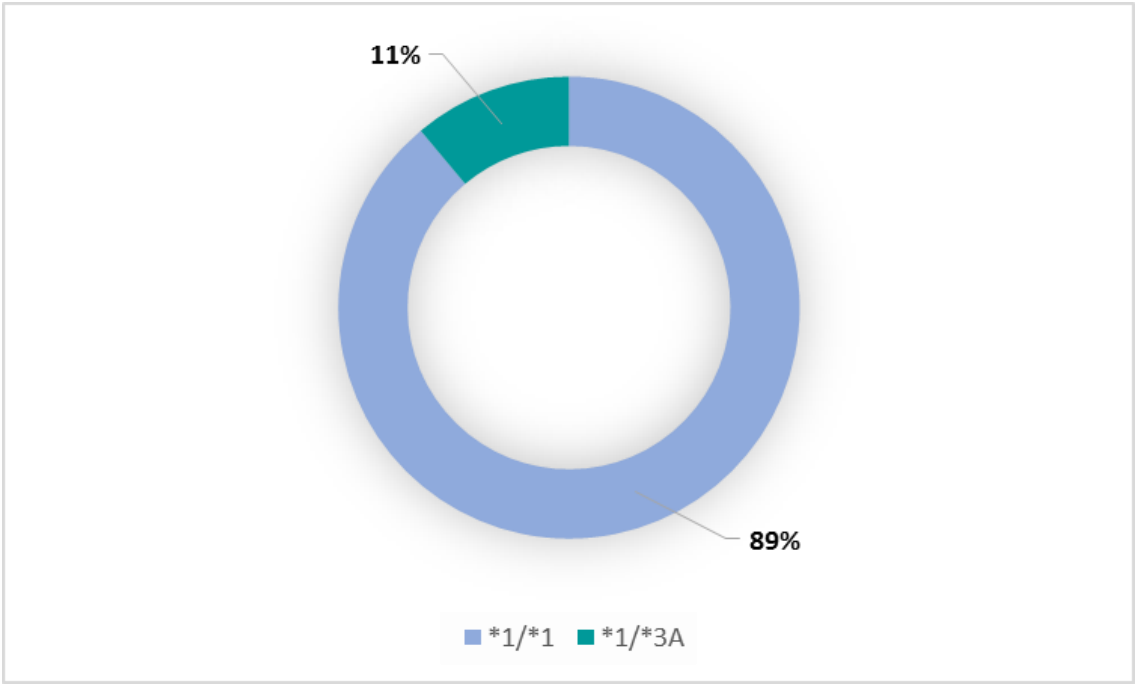

Figure 5.22 Diplotype distribution for TPMT

5.2.3.17 *UGT1A1*

Haplotype	Published Caucasian (%)	Haplotype number total	Haplotype total (%)	95% CI (%)	Two proportion z test (z value, p value)
*1	68	93	65	57-73	z=0.7, p=0.4
*6	n/a	0	0	n/a	n/a
*28	32	50	34	27-42	z=0.5, p=0.6
*36	<1	1	1	n/a	n/a
*37	<1	0	0	n/a	n/a
Total	100	144	100		

Table 5.31 Haplotype frequency calculation for UGT1A1.Caucasian figures are from www.pharmGKB.org (471)

The p value obtained by performing a Fisher's exact test comparing observed haplotype frequencies to published Caucasian haplotype frequencies was 0.6553, so the results were not significantly different to population frequencies(471). For commoner haplotypes, the expected figure was within the 95% confidence interval of the observed and there was no significant difference using a two proportion z-test. The frequency of the *28 haplotype is greater in African and Asian populations than in Caucasians, so published population frequencies may not exactly reflect a mixed population such as this one. The cohort diplotype distribution is shown in Figure 5.23.

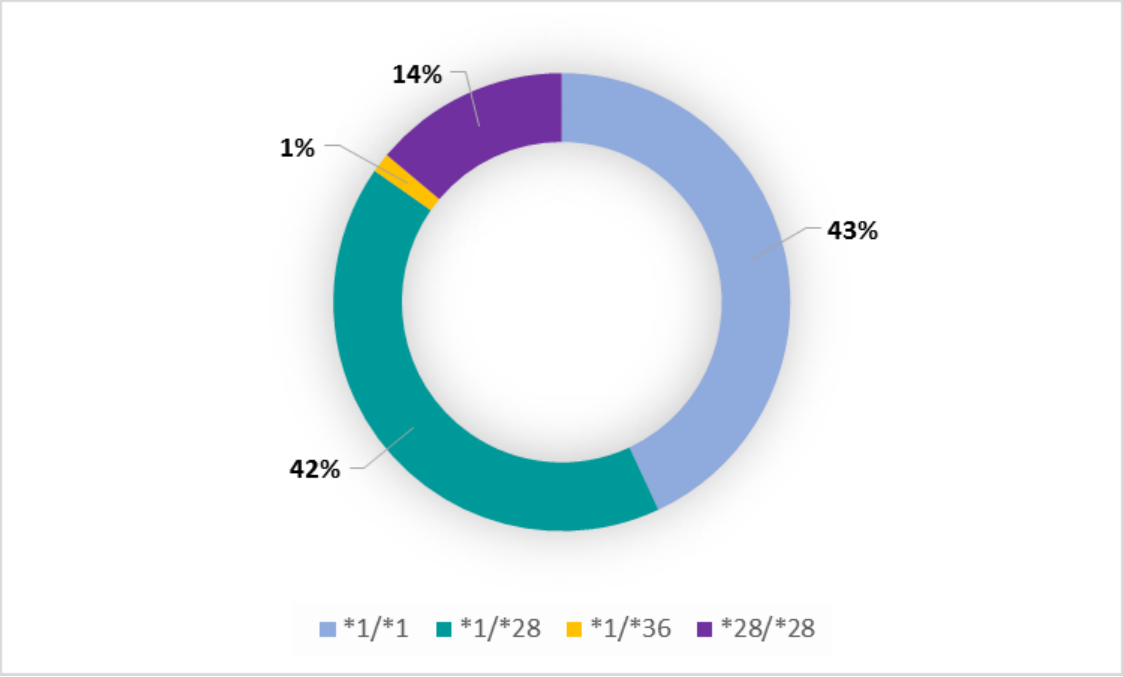

Figure 5.23 Diplotype distribution for UGT1A1

5.2.3.18 *UGT1A6*

Allele	ExAC (%)	Allele total	Allele frequency (%)	95% CI (%)	Two proportion z test (z value, p value)
G	97	134	93	88-97	z=2.8, p=0.05
T	3	10	7	3-12	z=2.8, p=0.05
Total	100	144	100		

Table 5.32 Allele frequency calculation for UGT1A6. European genotype frequencies from ExAC database(77)

The p value obtained by performing a Fisher's exact test comparing observed haplotype frequencies to published European haplotype frequencies was 0.1023, so the results were not significantly different to population frequencies(77). When European allele frequencies are used to calculate the z score the results are significantly different, although the ExAC values fall within the confidence interval for the observed values. However, when total ExAC frequency is used, the results are not significantly different (z=1.8, p=0.06). The frequency of the T allele is greater in African and Asian populations than in Caucasians, and therefore the European figures may not exactly reflect a population such as this one. The cohort genotype distribution is shown in Figure 5.24.

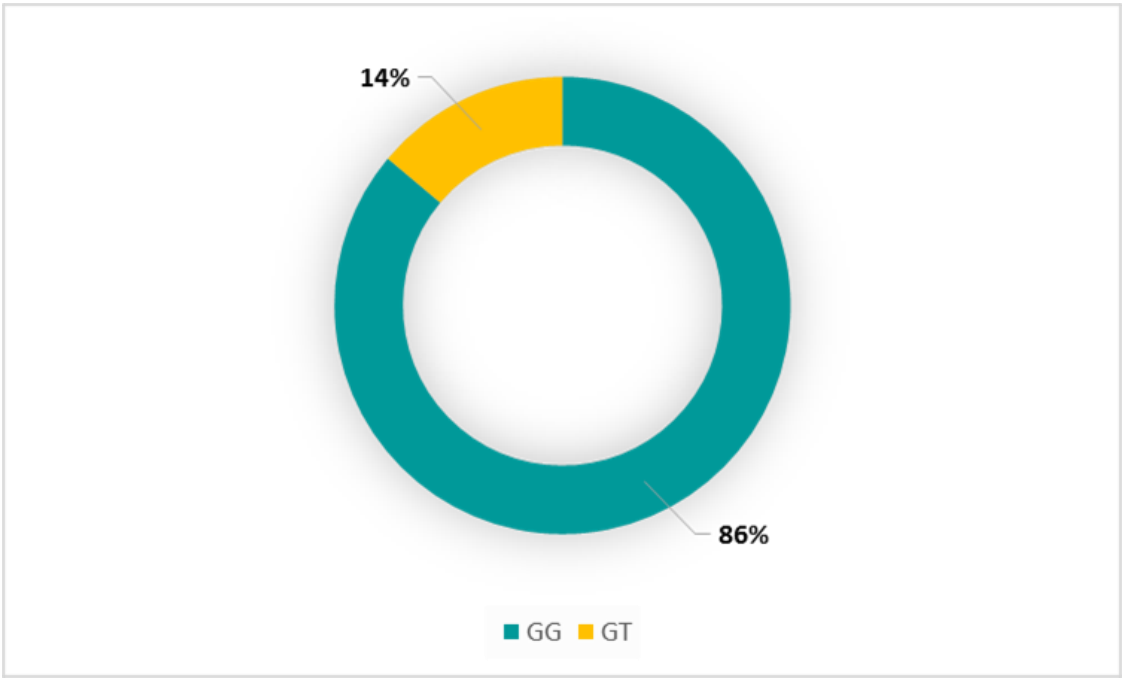

Figure 5.24 Genotype distribution for UGT1A6

5.2.3.19 *VKORC1*

Allele	Published Caucasian (%)	Allele total	Allele frequency (%)	95% CI (%)	Two proportion z test (z value, p value)
C	59	96	67	58-75	z=1.9, p=0.06
T	41	48	33	25-41	z=1.9, p=0.06
Total	100	144	100		

Table 5.33 Allele frequency calculation for VKORC1. Caucasian population frequency from Johnson et al.(510)

The p value obtained by performing a Fisher's exact test comparing observed allele frequencies to published Caucasian allele frequencies was 0.3053, so the results were not significantly different to population frequencies(510). For both alleles, the expected figure was within the 95% confidence interval of the observed and there was no significant difference using a two proportion z-test. The cohort genotype distribution is shown in Figure 5.25.

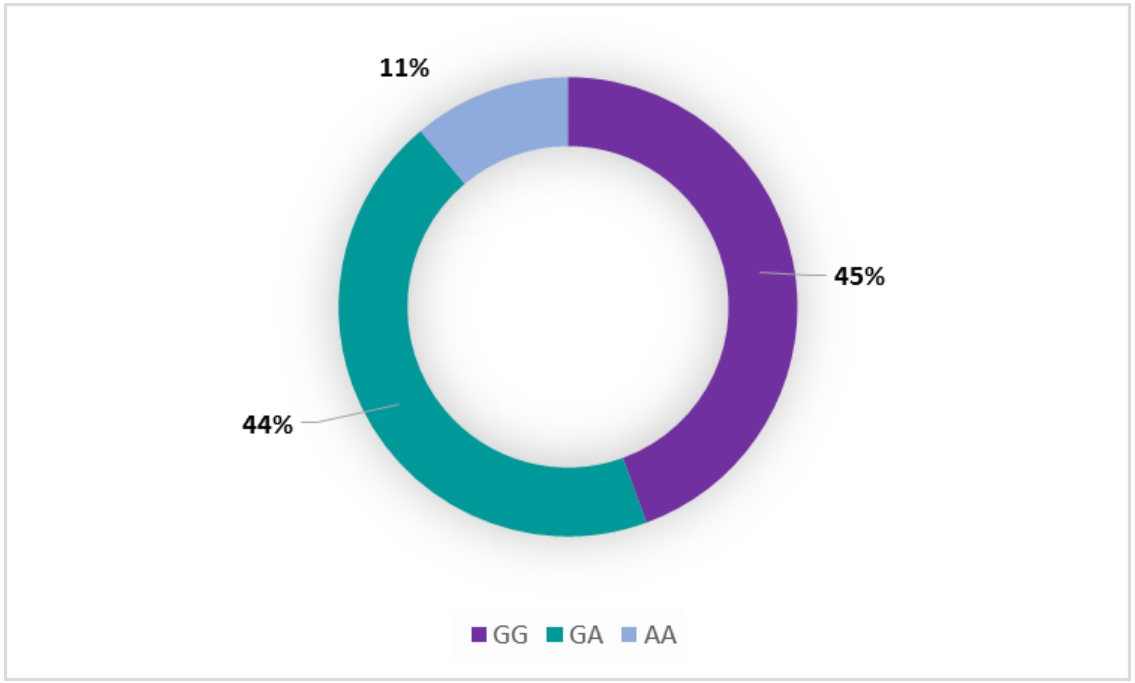

Figure 5.25 Genotype distribution for VKORC1

5.2.4 *CYP2C9, CYP2C19 and CYP2D6* haplotype confirmation using Astrolabe

Astrolabe was used to confirm *CYP2C9, CYP2C19* and *CYP2D6* haplotypes and to detect possible copy number variants (CNVs) in *CYP2D6*. In general all haplotypes were confirmed. Ones that were not are listed in Tables 5.34 to 5.36 below.

5.2.4.1 *CYP2C9* haplotype confirmation

There were 2 disagreements (Table 5.34). In one case, that of JDM-013, Astrolabe called *1/*26. The SNP associated with *CYP2C9* *26, rs200965026, is not present in WGS data. Coverage is good. Therefore this appears to be an error on the part of Astrolabe. The same was true of the second. Astrolabe called *12, but the SNP was not present. However, in the case of IBD-008, Astrolabe called the diplotype as *1/*9. When WGS data were rechecked, the SNP for the *9 haplotype was present, so the correct diplotype is *1/*9. As *1/*9 is a normal metaboliser, prescribing was unchanged.

Patient	Astrolabe	WGS	Reason
IBD-008	*1/*9	*1/*1	Astrolabe call correct. Did not look at *9 originally
JDM-013	*1/*26	*1/*1	Disagree with Astrolabe calls- SNP for *26 is not present
SRS-011	*1/*12	*1/*1	Disagree with Astrolabe calls- SNP for *12 is not present

Table 5.34 Conflicting Astrolabe and whole genome sequence calls for CYP2C9. Probable correct calls highlighted in green

5.2.4.2 *CYP2C19* haplotype confirmation

There were 15 disagreements (Table 5.35), 11 of which occurred when Astrolabe called the *3 haplotype of *CYP2C19*. The main SNP for *CYP2C19* *3 is rs4986893, which is not present in any of the samples. This is further discussed in section 5.3.2.2.3. In the case of IBD-010, Astrolabe correctly called the SNP for haplotype *15. This was not looked at originally as there was no associated prescribing advice. Results and prescribing sheets were amended to reflect this. In the cases of IBD-012 and JDM-012, Astrolabe called *2/*2. However, both patients are heterozygous. In both cases, the read depth was good, 42 and 23 respectively and the G:A ratio was 20:22 and 9:13 respectively. Astrolabe calls are incorrect but it is unclear why this is the case. In the case of IBD-014, Astrolabe called *34. However, none of the 3 SNPS for *34 are present and it is unclear why this call was made.

Patient	Astrolabe	WGS	Reason
BBS-016	*1/*3	*1/*1	Incorrect call by Astrolabe. Main SNP for *3 not present
BBS-017	*1/*3	*1/*1	Incorrect call by Astrolabe. Main SNP for *3 not present
IBD-001	*2/*3	*1/*2	Incorrect call by Astrolabe. Main SNP for *3 not present
IBD-003	*3/*3	*1/*1	Incorrect call by Astrolabe. Main SNP for *3 not present
IBD-006	*1/*3	*1/*1	Incorrect call by Astrolabe. Main SNP for *3 not present
IBD-010	*1/*15	*1/*1	Astrolabe call correct. Did not look at *15 originally
IBD-011	*2/*3	*1/*2	Incorrect call by Astrolabe. Main SNP for *3 not present
IBD-012	*2/*2	*1/*2	Incorrect call by Astrolabe. Read count 42. Ratio 20:22 A:G, so correct call appears to be *1/*2
IBD-014	*17/*34	*1/*17	Incorrect call by Astrolabe. No *34 SNPs present
IBD-015	*3/*17	*1/*17	Incorrect call by Astrolabe. Main SNP for *3 not present
JDM-001	*1/*3	*1/*1	Incorrect call by Astrolabe. Main SNP for *3 not present
JDM-003	*3/*17	*1/*17	Incorrect call by Astrolabe. Main SNP for *3 not present
JDM-005	*2/*3	*1/*2	Incorrect call by Astrolabe. Main SNP for *3 not present
JDM-009	*2/*3	*1/*2	Incorrect call by Astrolabe. Main SNP for *3 not present
JDM-012	*2/*2	*1/*2	Incorrect call by Astrolabe. Read count 23. Ratio 10:13 G:A, so correct call appears to be *1/*2

Table 5.35 Conflicting Astrolabe and whole genome sequence calls for CYP2C19. Probable correct calls highlighted in green

5.2.4.3 *CYP2D6* haplotype confirmation

There were 5 disagreements (Table 5.36). In three of these, Astrolabe correctly called a SNP associated with a rare haplotype that did not have associated prescribing advice, and was therefore not checked originally. Results and advice sheets were amended to reflect this. In the case of BBS-006, Astrolabe failed to call a heterozygous CCT deletion associated with the *9 haplotype. The read depth was 25, and the ratio was 10:15 CTT:delCTT, so in this case the Astrolabe call appears to be incorrect. In the case of SRS-017, Astrolabe called *56/*91. The tag SNP (SNP that identifies a haplotype) for *56 was not present. A single SNP seen in association with *56A was present, but not any of the other associated SNPs. A single SNP for *91 was present, but is was one that is seen in association with other haplotypes also. The tag SNP for *91 was not present. Therefore the Astrolabe data were judged to be incorrect.

Patient	Astrolabe	WGS	Reason
BBS-006	*1/*1	*1/*9	Incorrect call by Astrolabe, CTT deletion from 22:42128174-22:42128176 is present
BBS-015	*2/*35	*2/*2	Astrolabe call correct. Originally did not look at *35
JDM-007	*1/*29	*1/*2	Astrolabe call correct. Originally did not look at *29
JDM-012	*2/*35	*2/*2	Astrolabe call correct. Originally did not look at *35
SRS-017	*56/*91	*1/41	Disagree with Astrolabe- SNP for *56 not present, though single SNP for *56A is. Only 1 of 2 SNP for *91 present

Table 5.36 Conflicting Astrolabe and whole genome sequence calls for CYP2D6. Probable correct calls highlighted in green

5.2.4.4 *CYP2D6* copy number calls

Astrolabe also looked at whether there were any gene deletions or duplications in *CYP2D6.* The results are detailed in Table 5.37. A single deletion was identified in IBD-013. This appeared to be a true heterozygous deletion as the coding areas had a lower read depth than the average for the chromosome. In addition, there were no heterozygous SNPs. Seven duplications were identified. Five of them appeared to be real, in that the read depth of the coding area of CYP2D6 was greater than the average for the chromosome. Heterozygous SNPs were seen in a 2:1 ratio which is consistent with a duplication. In two cases however, the duplications did not appear to be real. In the case of JDM-005, the average coding-region read depth was 51 compared with an average chromosomal coding region read depth of 41. The SNPs were all biallelic and had a 1:1 ratio. In the case of JDM-009, the average coding-region read depth was 39 compared with an average chromosomal coding region read depth of 32. The SNPs were all biallelic and had a 1:1 ratio. Results were amended to take account of the duplications and deletions that were thought real.

Patient	Duplication or deletion	Average read depth gene	Average read depth chromosome	Homozygosity/ Heterozygosity	Conclusion
IBD-007	Duplication	42	30	SNPs have 1:2 ratio	Likely duplication *1/*41xN
IBD-008	Duplication	71	43	SNPs have 1:2 ratio	Likely duplication *1xN/*17
IBD-013	Deletion	15	33	SNPs homozygous	Likely deletion *1/*5
JDM-005	Duplication	51	41	SNP ratio 1:1.	Unlikely duplication *1/*41
JDM-009	Duplication	39	32	SNP ratio 1:1.	Unlikely duplication *1/*3
SRS-002	Duplication	54	42	SNPs have 1:2 ratio	Likely duplication *1x2/*2
SRS-013	Duplication	47	36	SNPs have 1:2 ratio	Likely duplication *1x2/*1
SRS-015	Duplication	52	39	SNPs have 1:2 ratio	Likely duplication *1x2/41

Table 5.37 CYP2D6 duplications and deletions called by Astrolabe. Probable correct calls highlighted in green

5.3 Discussion

5.3.1 Choice of genes and SNPs

5.3.1.1 Choice of pharmacogenes

Increasing numbers of genes are recognised as having an effect on drug absorption, distribution, metabolism or excretion, but for some of these there is little evidence to say what the clinical effect of genetic variation is, nor what strategies can be used to mitigate this. This limits their utility in clinical practice at present. One of the difficulties with pharmacogenomics is this lack of certainty around clinical implementation. In order to maximise the utility of the data obtained in this project, a decision was made to restrict whole genome analysis with clinical translation advice to genes with robust, evidence-based prescribing guidelines, hence the choice of PharmGKB actionable pharmacogenes. Without such guidelines the data are not useful to the prescriber or patient. However, with published guidelines, translation of the data into clear, clinically actionable results was possible, a key stage for the integration of pharmacogenomics into everyday clinical practice.

5.3.1.2 Choice of SNPs

Initially, the SNPs examined were those of the haplotypes that had prescribing guidelines associated with them, as described in section 5.3.1.1. The advantages of this were threefold. Firstly, it made interpretation easier, as these haplotypes are generally those seen most commonly in European populations. Much of the published haplotype frequency data cover only common haplotypes, so at present it is difficult to say whether additional haplotypes are truly rare or have not been looked for. Secondly, for rarer haplotypes, the functional effect is often unknown,

making prescribing advice impossible to give. Thirdly, for rare haplotypes, it is becoming apparent that sometimes SNPs that have been associated with rare haplotypes are actually reasonably common in some populations, again making data uninterpretable. Therefore, for clarity, a restricted set of SNPs was examined in the first instance. Some of these were later extended, as, for example, in the case of *CYP2C9* with the release of the PharmGKB updated prescribing guidelines which were published in 2017(510). Astrolabe looked at 122 *CYP2D6* haplotypes, and a small number of corrections were made to the extracted haplotypes. Additionally, an attempt was made to examine the WGS data for the entire PharmGKB *CYP2D6* haplotype set, which consisted of 197 SNPs covering 163 haplotypes or subhaplotypes (data not included in thesis). This proved too difficult to do manually, hence the use of Astrolabe. When examining data manually, multiple contradictory SNPs were present. For example, patient BBS-012, originally called as *1/*41, had a SNP that is associated with haplotypes *31 and *35B, but was missing the other associated SNPs for these. However, the SNPs associated with *41 are present. This issue has been identified by other researchers also, where many individuals did not match exactly with any described haplotype(629).

The tag haplotype system has some limitations in that the SNPs associated with particular haplotypes were generally described before the advent of large scale WGS data, and some studies looked at small numbers of individuals from homogenous populations. It is likely that they do not reflect the level of diversity in humans, particularly in ethnic groups other than those of European ancestry. This may be resolved as more data are generated, although without functional tests to accurately define activity levels, the translation of these data into clinically actionable recommendations will be difficult.

Another issue of note is that when examining WGS data there may be additional SNPs present that are not seen in haplotype data sets. Some are located near known SNPs of assumed functional importance. The effect of these may not be certain, and may affect both function and results of PCR-based assays. Currently, the standard in pharmacogenomics is to look at a small number of SNPs to infer function, whereas function may depend on the presence of additional SNPs or indeed the entire gene sequence. An important possible future direction of pharmacogenomic research may be around resolving what exactly a haplotype consists of and how best to identify it.

5.3.2 Extraction of data from whole genome sequences

5.3.2.1 Choice of method of data extraction

Several methods of data extraction were tried before a decision was made to manually visualise each SNP in the Integrative Genomics Viewer (IGV)(71, 72). Both Ingenuity® Variant Analysis™ (IVA) and Sapientia™ were used. The advantage of using programmes such as these were the ability to output data for further analysis and the fact that data such as read depth were displayed in an easy-to-read format. However, IVA requires vcf files and as not all SNPs would be present as variants, haplotypes or genotypes had to be imputed from their absence, which increased the risk of error. Sapientia accepts BAM files but filtering is used to reduce the number of variants. As

the mean allele frequencies (MAF) of many pharmacogenomic SNPs are very high, even when an appropriate gene panel was applied the numbers of SNPs displayed were so great that analysis was very difficult. It was not possible to input lists of SNPs.

Visualising the SNPs directly had several benefits. Firstly, it meant that even if SNPs were not called as being present for example due to low numbers of non-reference alleles, they could still be seen, and a note made for reviewing at the validation stage. Secondly, SNPs located close to the SNP of interest could be noted, again useful for validation. Thirdly, IGV has additional features that help to assess the quality of the data. For example it is easy to visualise read depth over an exon or gene, or view the allele fraction of heterozygous SNPs, as well as the ability to look at paired reads. Fourthly, it allowed the visualisation of parent-child trios and monozygotic twin pairs together.

Overall, the direct visualisation method was both more efficient and more useful than using the software described above, although it was time consuming. All SNPs were visualised twice to reduce the risk of error or mistranscription. The main disadvantage of this method was the fact that it was not automated. In addition, it was not possible to visualise more than three or four SNPs simultaneously without the information failing to load properly. An ideal solution would be to be able to add a list of SNPs to an appropriate variant visualisation software so a list of SNPs could be generated without having to attempt to filter the relevant variants or to automate the calling from BAM files using a custom pipeline.

5.3.2.2 Issues with data extraction

5.3.2.2.1 Phasing of haplotypes

At present, pharmacogenomic tests do not include haplotype phasing to assign SNPs to the maternal or paternal chromosome i.e. whether variants are in *cis* or in *trans*. Assigning the SNPs to haplotypes and combining the haplotypes into diplotypes assumes that the haplotypes and diplotypes can be inferred based on which SNPs are seen. As discussed in section 5.3.1, this may be an oversimplification. However, this is the process used in commercially available tests at present. There is support for this in the fact that population haplotype frequencies appear to be consistent with expected rates of ADRs in some studies(630). Resolving this issue in pharmacogenomics is problematic(631). Often the SNPs within a single gene are situated at a distance from one another and so cannot be resolved by methods such as looking at paired end reads. Methods such as chromosome sorting are expensive and time consuming.

One method of resolving the issue is to use parent-child trios. WGS information was available for 10 parent-child trios in this study (SRS cohort). While this number was small, it did allow consideration of whether assigning diplotypes on the presence or absence of SNPs was accurate. In general, the trio data confirmed that the approach was effective, with the vast majority of child diplotypes being a combination of parental diplotypes. The only exception was *DPYD*, where in two parent child trios the diplotypes did not make sense (Table 5.38). A similar uninterpretable *DYPD* diplotype was seen in a single patient from the IBD cohort. As can be seen below, even without trio data, the haplotypes are clearly abnormal for SRS-002, SRS-009 and IBD-020, but

the trio data also suggests that the data for SRS-001 and SRS-007 may also be incorrect. The sequencing data from IGV for SRS-002 is shown in Figure 5.26. The read depth data for IBD-020 suggest that one possible explanation is a heterozygous deletion of the region containing the *9A SNP, but this does not appear to be the case for the SRS trios. The *5 haplotype is further discussed in section 5.3.3.3.4. In all other patients and for all other genes, the trio data worked well, with the child diplotype or genotype being a combination of parental haplotypes or alleles. Trio data are non-informative where all members are heterozygous for a particular SNP.

One solution for the problem of phasing in pharmacogenomics is the introduction of long-read technology(632). Long reads will, in theory, allow the determination of phase directly from sequencing data. Currently, this technology is expensive and not in routine use, but as with all genetic technologies, costs will decrease. Recent studies show, however, that this may not be straightforward(373).

Patient ID	**Diplotype**	**Read depth for homozygous SNP**	**Read depth for heterozygous SNP**
SRS-001 (child, trio 1)	*5/*9A	n/a	n/a
SRS-002 (father, trio 1)	*5/*5/*9A (hom for *5, het for *9A)	41	44
SRS-003 (mother, trio 1)	*1/*1	n/a	n/a
SRS-007 (child, trio 3)	*5/*6	n/a	n/a
SRS-008 (father, trio 3)	*1/*1	n/a	n/a
SRS-009 (mother, trio 3)	*5/*5/*6 (hom for *5, het for *6)	41	32
IBD-020	*6/*9A/*9A (het for *6, hom for *9A)	27	44

Table 5.38 Uninterpretable diplotypes in DPYD

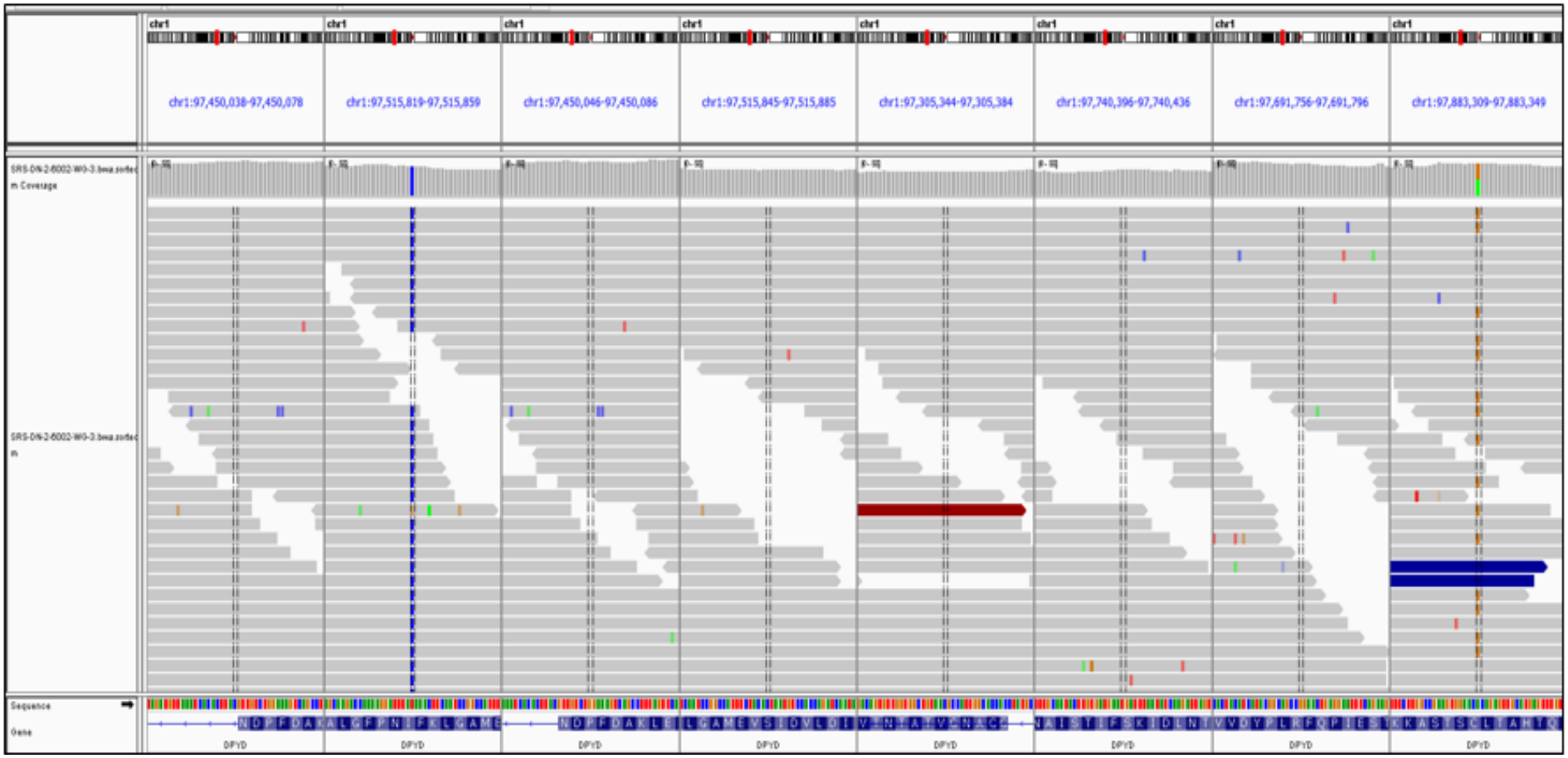

*Figure 5.26 Screenshot of SNPs for DPYD haplotypes for patient SRS-002 (image from IGV). Rs1801189, the tag SNP for *5 is homozygous, while rs1801265, the tag SNP for *9 is heterozygous (brown)*

5.3.2.2.2 Calling copy number variants in *CYP2D6*

Astrolabe was used to identify CNVs (section 5.2.4.4)(633). It operates by looking for areas where read depth for coding regions of *CYP2D6* differs from read depth of a pre-specified control region. In addition it looks at whether the sequence is homozygous or heterozygous for SNPs. Deleted regions, as seen in the *CYP2D6* *5 gene deletion have lower read depth and loss of heterozygosity. In the case of duplications, the read depth is greater than expected and heterozygosity is preserved, but SNPs may be seen at allele fraction ratios of other than 1:1. The Astrolabe calls were then checked manually to see whether, based on read depth and variation, these calls appeared real. Out of eight calls, six were real on visualisation of BAM files. In the other two, both duplications, read depth was not unusually high and the allele fraction of the SNPs was approximately 50%. Astrolabe is a useful tool for identifying possible CNVs, but checking was required to determine whether calls were real. Looking at read depth alone generated a greater number of possible CNVs which then had to be checked. Read depth alone did not identify any CNVs that were missed by Astrolabe. The optimum strategy appears to be to use Astrolabe to identify possible CNVs and then to confirm these manually.

5.3.2.2.3 Calling diplotypes using Astrolabe

In two of the 23 disagreements between Astrolabe and WGS calls Astrolabe appeared to call a rare haplotype based on a single SNP that was not the tag SNP for the haplotype. In a further 11 cases the issue was with the *CYP2C19* *3 haplotype. PharmGKB lists only a single SNP for *3 which was not present in the WGS data. When the developer was informed of this, they said it was a problem that they were aware of and that they had encountered when running their own samples. Their response was “These false positive calls are usually due to a variant in the definition sequence that, as far as the nomenclature is concerned, only appear with that specific uncommon haplotype, but when sequencing large and/or more diverse population sets is actually much more frequently seen as part of a normal functional variation background. These errors usually originate from smaller studies and/or large genes (such as *2C9* and *2C19*) when groups sequence areas that are usually not covered creating a hidden information imbalance between the haplotype definitions”. PharmGKB lists only a single SNP for *3 (rs4986893), while PharmVar lists two (rs4986893 and rs3758581). This latter SNP is present in the samples called as *3 by Astrolabe, but the other SNP used by PharmGKB is not. It appears that Astrolabe is calling *3 on the basis of one of the SNPs. The other SNP is the tag SNP for the *3 haplotype. The rs3758581 SNP has a population frequency of 7%, which is higher than the population frequency of *3.

In two disagreements, Astrolabe called rare haplotypes where none of the variants listed in PharmGKB was present. It is unclear from the output which SNPs Astrolabe looks at but the paper cites the old location of the Human Cytochrome P450 Nomenclature Committee allele list before its migration to PharmVar (www.cypalleles.ki.se), which is no longer functional. However the list has been migrated to PharmVar (www.pharmvar.org) and when checked the SNPs were found to be the same as in PharmGKB, so it is unclear where the error has arisen. There were three additional disagreements where Astrolabe appeared to miscall something present in WGS, in one case a heterozygous CCT deletion and in two a heterozygous SNP. In all cases read depth was

good and allele fraction consistent with the heterozygous state. It is unclear why these were missed. In every case, Astrolabe developers have been informed of disagreements.

In five cases, Astrolabe identified a rare haplotype that was not originally checked in WGS. In the case of a *CYP2C9* *9 haplotype, it was identified in a patient of African ancestry. The *9 haplotype appears to be rare. It is not mentioned in PharmGKB figures for European, British or African haplotype frequencies. In ExAC, the global frequency is less than 1%, but the frequency in those of African ancestry is higher, at 7.5%.

While Astrolabe is a useful tool, these results show that it cannot be used in isolation to call haplotypes, especially for *CYP2C19*. Its success rate in this study appears to be less good than that quoted by the developers in their paper(633). The fact that it identified some rarer haplotypes that would have been missed using the original strategy of checking only haplotypes with published guidance, and the fact that it is possible to use the PharmGKB website to look up these rarer haplotypes and infer prescribing guidance raises a question, namely whether an extended set of haplotypes should be looked for. As discussed in section 5.3.1.2, the more SNPs used, the harder the data are to interrogate manually. When interrogating the data computationally, as was confirmed by the Astrolabe dataset, there needs to be a prioritising of SNPs so that a haplotype is not called without the tag SNP being present. However, there is certainly an argument for including some of the commoner rare haplotypes, such as *CYP2D6* *35, whose population frequency appears to be in the order of 2%(471).

5.3.2.3 Other issues

5.3.2.3.1 *CFTR*

CFTR is considered an actionable pharmacogene, but is only of relevance if a patient has a diagnosis of CF and treatment with Ivacaftor is being considered. No patient in the study has a diagnosis of CF. In addition, checking for the p.Gly551Asp mutation or any of the additional eight mutations for which Ivacaftor is now licensed would constitute a carrier test for cystic fibrosis. In 2013, the ACMG recommended the return of secondary findings for WES or WGS studies(5). However, the findings it describes are those for potentially fatal diseases with some treatment or screening process of which patients could avail. They do not mention carrier testing. The current position in clinical genetics is that carrier testing, including for CF, requires explicit informed consent and should not be carried out on minors. Individuals in this study were not consented for CF or indeed any carrier testing. In addition, because there was no plan to return findings from the study to patients, and patients were informed of this, there was no discussion of the possibility of secondary findings, and no option to opt in or out. Many of the patients in this study are minors. For all of these reasons, CFTR was excluded from analysis.

5.3.2.3.2 *CYP2D6*

While there were not any obvious issues in extracting data for *CYP2D6* in this study, previous publications have noted that because of high levels of sequence homology with pseudogenes, high levels of variation, GC content, copy number variants and repetitive sequences, *CYP2D6* is a difficult gene to sequence and extract data from(633). In the 1000 Genomes project, it was

considered to be part of the inaccessible genome(604). Reasons why the same issues might not have arisen might include better quality sequence data or that extracting data manually is more reliable that computational methods. An alternative is that the pseudogenes or homologous regions contained the same sequences as *CYP2D6* for each SNP examined in the study. While this does not mean that it is impossible to extract data from WGS, it does mean that this requires further careful validation with larger numbers of samples. As previously discussed, long-read sequencing technologies may solve this and other issues in the extraction of pharmacogenomic information from WGS data.

5.3.2.3.3 *G6PD*

There were several issues with *G6PD.* The first is that in females heterozygous for G6PD deficiency alleles, the exact clinical effect is very variable. Therefore the recommendation is that a functional test should be performed as clinical consequences cannot be accurately predicted.

The second is that in the case of the A(2) allele, seen in two females in this study, one in the homozygous state, the allele is functional unless accompanied by certain other SNPs, namely rs1050828, rs137852328 or rs76723693. None of the individuals who had the A(2) allele in the study had any of the additional SNPs, so they are not considered to be G6PD deficient.

A third issue is that there is a small chance of discovering that a male has an XXY, or indeed other multiple of X, karyotype when looking for variants in *G6PD*, if heterozygous SNPs are identified on the X chromosome. This is an example of why, when individuals are being consented for genetic tests, they are warned that there may be unexpected findings. In this study, no secondary findings were identified, and in addition, patients were informed that results would not be returned, but it raises an issue for other pharmacogenomic studies and their consent processes.

A final issue is that patients who have G6PD deficiency will be at risk from drugs and substances other than rasburicase, including some foods, and may suffer from haemolytic anaemia during infections. This may mean that there is a duty to warn beyond informing the patient about the risk with rasburicase. A list of additional drugs can be found in the CPIC *G6PD* guidelines paper by Relling et al.(548).

5.3.2.3.4 *HLA-B*

There was some difficulty in determining which SNPs best indicated the presence of *HLA-B* *15:02, with the literature being directly contradictory in many cases(625, 626). PharmGKB does not suggest an allele set for *HLA-B* *15:02. SNPs from both publications were checked, namely rs2395148, rs10484555, rs144012689 and rs2524160, and in the study population, no *HLA-B* *15:02 positive individuals were identified. *HLA-B* *44 and *HLA-B* *58 are both identified from the sequence rather than from any specific SNPs. These data were not extracted from the whole genome sequence, although it might be done computationally.

5.3.2.3.5 *IFNL3*

In the case of *IFNL3*, patients are determined to have a favourable or unfavourable genotype with respect to having interferon and related treatments for viral hepatitis. Patients heterozygous or homozygous for rs12979860 are considered to have an unfavourable genotype and are less likely to respond to treatment(565). However, having an unfavourable genotype does not mean that they will not respond to treatment (30%, increasing to 60% if combined with a protease inhibitor, as compared to 70% increasing to 90% for those with a favourable genotype). It is possible that patients could be refused treatment based on their pharmacogenomic test results, and currently there are few, if any, alternatives available for those with viral hepatitis. There is a debate to be had about whether this is ethical in the absence of other treatments, especially since, because of population frequencies, this will disadvantage certain ethnic groups.

5.3.3 Haplotype frequency analysis

When the haplotype frequency analysis was performed, in no case was the overall observed haplotype frequency different to the population frequency. However, some points should be noted.

5.3.3.1 Population frequencies and ethnicity

Many of the published haplotype frequencies are calculated for ethnic groups. In section 5.2.3, many of the published population haplotype frequencies used for comparison are for white Europeans or Caucasians, i.e. people of white European or Caucasian ancestry, rather than being reflective of the ethnically diverse population now resident in Europe and the UK. The population in the case of this study is mixed, although the majority of participants described themselves as White British. Self-reported ethnicity figures by cohort can be seen in Table 5.39. Overall, 85% of the study participants describe themselves as white British, Irish or European, with 5% describing themselves as African, 7% describing themselves as Asian, 1% describing themselves as Middle Eastern and 2% describing themselves as other. Of note, no one in the JDM cohort described themselves as White British, raising the possibility that this was not an option on the ethnicity list. The UK Government figures from the 2011 census indicate that this is broadly reflective of the UK population, with 87% of people describing themselves as white, 7% as Asian, 3% as African and 3% for mixed/other(634).

The fact that the study population was broadly reflective of the UK population in featuring individuals of non-white ethnicity provides a possible explanation for the fact that there are some differences between the observed and published haplotype frequencies. Haplotype frequencies that are rare in people of white ancestry may be commoner in individuals of other ethnicity. A case in point is *CYP2C9*, where the *3 haplotype is rare in Europeans but more common in South-East Asians, while the *5 and *8 haplotypes are much more common in those of African ancestry(619).

Cohort	Ethnicity	Number (%)
BBS	White British	16 (89%)
	White Irish	2 (11%)
IBD	African	2 (10%)
	Asian	6 (30%)
	Middle Eastern	1 (5%)
	Other	1 (5%)
	White British	4 (20%)
	White European	6 (30%)
JDM	African	2 (15%)
	Mixed other	1 (8%)
	White	10 (77%)
SRS	White British	26 (87%)
	White European	4 (13%)
USH	White British	3 (100%)

Table 5.39 Self-described ethnicity by cohort

5.3.3.2 Other issues with published population frequencies

As mentioned in section 5.3.3, there are issues with published population frequencies beyond the fact that the data have been stratified by ethnic origin, rather than reflecting mixed populations. Many of the studies from which the data have been extracted are small. This means that there is a risk of missing rarer haplotypes and also a risk that any rare haplotypes found will skew population data. Many ethnic groups are not well covered by population data and figures are most likely to be reliable for people of white ancestry. One solution to this is large scale population studies and an obvious way of doing this in the UK would be to extract pharmacogenomic data from the 100,000 Genomes project, something that Genomics England is planning. Generally when doing genetic studies, groups of individuals that have been pre-selected for having medical issues do not make good subjects. But in pharmacogenomic research, medical diagnoses should not make a difference to the underlying pharmacogenomic profile, with the exception of disease-causing genes such as *CFTR* and *G6PD.* In addition, more accurate figures for particular ethnicities, certainly for individual SNPs, could be extracted from population databases such as ExAC.

5.3.3.3 Issues with observed haplotype frequencies

Although none of the overall haplotype or genotype frequencies were significantly different to published frequencies, some individual haplotypes for particular genes were less close to the published frequencies than others.

5.3.3.3.1 *CYP2C19*

A slightly higher percentage of *2 haplotypes and a slightly lower percentage of *17 haplotypes were seen in *CYP2C19*. This is likely to be accounted for, at least in part, by the mixed ethnicity cohort; *2 alleles are seen at a higher frequency in Africans and Asians, while *17 is seen at a lower frequency in these groups(619). Another possible explanation is skewing due to the small size of the cohort.

5.3.3.3.2 *CYP2D6*

One of the haplotype frequencies that was most different to expected frequencies was the *41 haplotype of *CYP2D6*. While the expected figure was within the confidence interval of the observed, the two proportion z score was significantly different. The frequency of *41 was 14% while in the PharmGKB allele frequency data it was 7%, z=3.3, p=0.001(471). In other published data, the frequency is even lower. Zhou et al. state it as 3-3.5%, with a higher frequency in South Asians of 14%(635). The results seen cannot be explained by ethnicity, as apart from one person describing themselves as Middle Eastern, all the individuals with the *41 haplotype were of White British or European ancestry. The important SNP for *41 is rs28371725, which is also seen in *61 and *69, both of which are rare, and have other SNPs associated with them. *41 also has many additional associated SNPs, all of which are seen in other haplotypes, especially *2. The questions therefore are, can rs28371725 be seen not in association with *41, in which case the haplotype calls may be wrong, are the population data for the *41 allele incorrect, or do the small numbers of individuals in the study result in some skewing? The additional 7% is 10 additional haplotypes out of the 144 haplotype total. However, in general, figures for most haplotypes/genotypes of most genes differed by only a few percentage points, so a 100% increase is unusual. The answer to this remains unclear. Possible methods of elucidation might include Sanger sequencing of rs2837172, validation by other methods (see Chapter 6) or functional studies to determine whether this did indeed represent a reduced function haplotype.

5.3.3.3.3 *CYP3A5*

The population frequency of the *CYP3A5* *6 allele in a European population is 0. However, it is higher than 15% in people of African ancestry(619). In the study, the individual with the *6/*6 diplotype is of African ancestry, while the other individual with a *6 haplotype describes their ancestry as mixed other. This is likely to account for the increased percentage seen. It is still a high proportion given the small number of people of African ancestry in the cohort, but such numbers are easily skewed given the small number of individuals involved.

5.3.3.3.4 *DPYD*

A higher percentage of *5 haplotypes was seen in *DPYD*, 20% as opposed to 15% in the literature(621). However, the frequency in individuals of African origin is 27%, and in those of Asian origin, 28%. This is may account for at least some of the observed difference. However as discussed in section 5.3.2.2.1, there may be a problem with *DPYD* as some of the observed diplotypes did not make any sense. Some of these included the *5 haplotype, so it is possible that there may be an error in calling *DPYD* haplotypes. In a recent large study by Reisberg et al,

difficulty was encountered in distinguishing DPYD *5, *6 and *9 alleles(636). Alternatively, the solution may be that the SNP thought to be associated with the *5 haplotype is in fact occasionally seen in other haplotypes or that other SNPS are required along with rs1801159 for the *5 haplotype to be called. In addition, according to ExAC, the SNP associated with the *5 haplotype is seen in about 20% of Europeans, which would fit with the data in this study, making the most likely explanation either that the frequency of the *5 haplotype has been underestimated or that the SNP is seen in association with other haplotypes.

5.3.3.3.5 *HLA-A*

HLA-A *31:01 was seen at a slightly higher frequency than expected. The tag SNP checked for *HLA-A* *31:01 was rs1061235. The sensitivity of using this SNP as a proxy for *HLA-A* *31:01 was evaluated by He *et al.(637)*. They estimated the sensitivity to be 100%, i.e. all individuals truly positive for *HLA-A* *31:01 would be picked up by testing this SNP, and the specificity at 84%. This means that some false positives will be seen. Applying these figures to the numbers obtained in this study would indicate that approximately one false positive would be expected. This would bring the numbers closer to expected, but still in the order of 6% (UK population frequency 3-5%).

5.3.3.3.6 *IFNL3*

In the case of *IFNL3,* a higher proportion of favourable genotypes was seen than expected (70% as opposed to 63%(565). Higher frequencies are seen in Asians and lower frequencies in Africans. This might account for some of the difference. Other possible explanations include the possibility that the American population with European ancestry from whom these numbers were obtained does not accurately reflect the allele frequency in the UK or that the numbers are skewed due to the small size of the cohort.

5.3.3.3.7 *SLCO1B1*

While there are small differences between the observed population frequencies and the ExAC data, the figures observed are within the 12-20% minor allele frequency range quoted in the PharmGKB paper for *SLCO1B1*(480). However, the haplotype frequency of the *5 allele represented by rs4149056 is approximately 2% (www.pharmGKB.com). This SNP also identifies haplotypes *15 and *17, other reduced function haplotypes, whose combined population frequency in Caucasians is up to 13%. Therefore this SNP represents several reduced function alleles, not just *5. This is further discussed in Chapter 6.

5.3.3.3.8 *UGT1A1*

The slightly increased frequency of the *28 haplotype may be accounted for in part by the increased frequency of this haplotype in African populations, where frequencies may be as high as 40%. 2 of the 6 individuals of African ancestry are homozygous for the *28 haplotype, and another 2 are heterozygous.

5.3.3.3.9 *UGT1A6*

The increased frequency of the T allele may be accounted for in part by the increased frequency of this allele in African populations, where frequencies may be as high as 12% (European

frequency 2%). 3 of the 6 individuals of African or mixed origin carried a T allele, accounting for 1/3 of T alleles seen, while making up 1/12 of the cohort.

5.3.3.3.9 *VKORC1*

While not significantly different, the C allele is seen at a higher than expected frequency. The C allele is commoner in those of African and South Asian ancestry, so this may account for some of the difference.

5.3.4 Presentation of data

5.3.4.1 Long form prescribing guidance

Initially, it was intended that this was the only form in which guidance would be prepared. It contains the complete pharmacogenomic profile for the patient, the functional effect of the genotype or diplotype, the possible clinical effect of the genotype or diplotype, the source of the recommendation and the prescribing advice. However, it quickly became clear that it was not easy to navigate, as it ended up as a 78-row table, spread over four printed pages per patient. As guidance often came from more than one source, it was confusing in terms of ability to see easily what the prescribing recommendations were and which ones were important. This risk of important recommendations being overlooked was felt to be high. However, it was a good source of data, and was helpful to have for each patient. An example is shown above in Table 5.13.

5.3.4.2 Summary prescribing guidance

A decision was made to produce a summary advice sheet. This used the traffic light system used by many pharmacogenomics companies. Traffic light labelling is now widely understood by the general public, due to its use on food labelling(638). Red drugs should be avoided by the patient, orange drugs should be used with caution and may require dose or monitoring adjustments, and green drugs are can be prescribed as usual, without additional monitoring or initial dose adjustment. In order to keep this as succinct as possible, a brief summary of the guidance was given, and CPIC guidance was used in preference to DPWG where guidelines conflicted (section 5.3.5.7). Also for succinctness, when it came to green drugs, the genes which did not require a change in prescribing guidance were listed. On the back of each page is a comprehensive list of which drugs were metabolised by which gene. In addition, any genes for which prescribing guidance could not be given were noted, as was the case with *DPYD* for a small number of patients. A disclaimer was added to state that rare haplotypes may not be checked and that prescribing guidance was liable to change. It also gave the PharmGKB website link so that recent guidance could be obtained (Figures 5.1 to 5.4). For the majority of patients, the guidance ran to a single A4 page, with the gene and drug list on the reverse; in approximately 30% of patients it continued onto a second page. This was mainly the case for *CYP2D6* poor or ultra-rapid metabolisers, where a large number of drugs required prescribing guidance. This summary guidance was easier to follow and it was clear which recommendations were important.

5.3.4.3 Improving clarity and longevity of prescribing guidance

Giving large numbers of prescribing recommendations simultaneously may lead to a lack of clarity. In addition, while all of the guidance may be relevant to a patient at some point in their lives, it is unlikely that more than one or two recommendations would be of interest at any given point in time. A web-based platform such as that available on www.PharmGKB.org would be a better solution. When a drug is prescribed, the drug and the patient's diplotype or genotype could be entered and advice could be returned. This would avoid giving too much information at once and have the additional benefit that advice could be kept up-to-date on the web-based platform, avoiding the patient or physician keeping superseded guidance in use. However, the platform would have to be clear and easy to use, and the patient's diplotypes would need to be recorded or kept to hand.

An alternative method is that of the UPGX consortium, a group who have implemented the **Pre**-emptive Pharmacogenomic Testing for **P**reventing **A**dverse Drug **E**ffects (PREPARE) study(639). The study aims to implement pharmacogenomic testing in multiple European centres in a crossover design. Each patient is issued with a small card, which has a QR code. When they present at a hospital, medical centre or pharmacy, the code is scanned. It links to a website where patient-specific guidance is available. In the age of smartphones, it is also conceivable that patients might carry data on their phone, with automatic app updates when guidance changes. Various platforms are being developed, such as the Janusmed Interaktioner in Sweden (http://www.janusinfo.se/Beslutsstod/Janusmed-interaktioner-och-riskprofil/) and the ePGA platform (www.epga.gr). Many these have been developed as research tools and are not yet licensed for clinical use.

Whichever method is chosen, it is important that data are kept current and the interface is user-friendly, as, unless the information is easy to interpret, clinicians are unlikely to change prescribing.

5.3.5 Issues with guidelines

A decision was made to use guidelines from CPIC, DPWG and CPNDS as published on the PharmGKB website. Additional guidelines are available, such as those from the Food and Drug administration (FDA), European Medicines Agency (EMA) and in the literature. However, as discussed in section 5.3.1.1, for guidelines to be useful they have to be from a reputable source, be up-do-date, easy to locate and easy to implement. The guidelines published on the PharmGKB website meet all of these criteria. They are reviewed and updated, and their evolution is easy to follow on the website. They are collated in a single location. Online tools allow for ease of use and the prescribing guidance is easy to follow. However, the use of these guidelines is not without issue.

5.3.5.1 Limitations of selected guidelines

One of the limitations of restricting the prescribing advice to a limited set of guidelines is that other pharmacogenes with a good evidence base may be inadvertently excluded. As more guidelines

are developed, this may become less of an issue. One reason why guidelines are slow to be published is the level of evidence required. In general, large, good quality studies are needed for guideline development.

5.3.5.2 Restriction to certain haplotypes

As discussed in section 5.3.1.2, a decision was made to restrict to haplotypes where prescribing guidance was available. In fact, in many cases, if the functional effect of a rare haplotype is known, prescribing advice can be inferred, and the PharmGKB website gives advice for many rarer haplotypes not specifically mentioned in the original papers. This strengthens the case for looking at a wider range of haplotypes. Missing a rare haplotype risks the provision of incorrect prescribing advice. There are ethical as well as medicolegal issues with this, especially as there is the potential to cause greater harm than if the test had never been done.

5.3.5.3 Limitation to certain drugs

Again, generally because of a lack of robust evidence, the PharmGKB guidelines are restricted to a small number of drugs in any given class. In the guideline for *CYP2D6* and codeine, tramadol is mentioned as an alternative to be used with caution, but there are no specific guidelines relating to it. The same applies to *G6PD*, where multiple drugs can cause problems in G6PD-deficient patients, but guidelines only cover rasburicase(548). The guidelines for prevention of anthracycline toxicity are currently limited to daunorubicin and doxorubicin(568). Another example is that of the SSRIs. The paper covers both *CYP2C19* and *CYP2D6,* but prescribing guidelines that take account of both diplotypes are not available(516).

5.3.5.4 Lack of combined guidelines

There are very few prescribing guidelines that take account of more than one pharmacogene at a time. Examples include *CYP2C19* and *CYP2D6* and their effect on tricyclic antidepressants, and *CYP2C9* and *VKORC1* and the effect on warfarin prescribing. For example, sertraline is metabolised by multiple CYP450 enzymes including *CYP2C9*, *CYP2C19* and *CYP2D6*, but currently specific guidelines only exist for sertraline and *CYPC19*(640). However sertraline is among the drugs mentioned in the *CYP2C19* and *CYP2D6* SSRI guideline so prescribing advice could be inferred from this(516). A further issue is that drugs can inhibit or induce enzymes and alter the phenotypic metaboliser status of an individual as in the case of *CYP2D6* and fluoxetine, which three individuals in the study, all *CYP2D6* normal metabolisers, are taking(641). Fluoxetine inhibits *CYP2D6*, so these individuals may respond to other drugs as intermediate metabolisers.

5.3.5.5 Non-pharmacogenomic factors

Many non-pharmacogenomic factors affect how individuals will react to a drug, but currently, these are rarely taken into account when prescribing. These factors may include gender, height and weight, renal or hepatic function, diet and so on. There is also limited evidence that gender can play a role in gene expression, and therefore maybe in pharmacogenomics(642). Again, none of these are included in pharmacogenomic guidelines, although calculators are available for calculating drugs with very variable doses such as warfarin.

5.3.5.6 Adult and paediatric guidelines

The majority of available pharmacogenomic guidelines are for adults only, and are not available for children, although there is no reason to think that children do not exhibit pharmacogenomic differences in drug metabolism. There are some noticeable exceptions. The CPNDS was set up to develop paediatric guidelines, and there are CPNDS guidelines available on the PharmGKB website that are for paediatric patients only. These include the guidelines for *RARG*, *SLC28A3* and *UGT1A6* and the anthracyclines daunorubicin and doxorubicin and the guidelines for *TPMT* and cisplatin(568, 576). The daunorubicin/doxorubicin guideline specifically states that pharmacogenomic testing is not recommended in adult patients, as studies had not confirmed the associated with cardiotoxicity(643, 644). Cisplatin ototoxicity is a bigger problem in children than in adults, as children exposed to cisplatin are more likely to develop hearing loss, and early hearing loss is more likely to affect speech development(576). The guidelines do not recommend testing adults due to a lack of evidence.

Many of the adult guidelines mention paediatric patients. Examples include the guidelines for codeine and *CYP2D6* and the guidelines for *CYP2C19* and *CYP2D6* and SSRIs. In the case of the codeine guideline, it specifically mentions that codeine is not recommended in children under two or after adenoidectomy or tonsillectomy(527). It also mentions risks to breastfed infants of ultra-rapid metaboliser mothers after a reported death(645). It does not, however, give any recommendations for dose adjustment. Similarly in the case of SSRIs, guidelines state that paediatric evidence is scarce, that *CYP2D6* guidelines may be extrapolatable to adolescents and children with close monitoring, but *CYP2C19* guidelines may not be, as children may have increased levels of *CYP2C19* activity compared to adults. Prescribing in children is more complex because dose often varies with age and weight and levels of enzyme activity may vary throughout childhood, meaning that pharmacogenomics may play a greater or lesser role at various stages(646, 647). Most pharmacogenomic studies, and indeed most drug trials, are performed in adults, so limited evidence is a recurring problem in extrapolating pharmacogenomic guidelines to children. However, some paediatric studies exist, and there may be an argument for looking beyond the limited range of studies from the CPIC or DPWG when considering prescribing in paediatric patients(648, 649). In addition, more paediatric studies are required to further understanding of these issues and to allow safer prescribing in children.

5.3.5.7 Differences between CPIC and DPWG guidelines

As previously discussed, the majority of guidelines used in this study have been from the CPIC or DPWG. In general, guidelines are concordant, with similar advice given for the same diplotypes by each organisation. Differences between the organisations and their methods of guideline development, as well as differences in guidelines, are discussed in a paper written by members of both consortia(87). Differences of methodology include the scales on which the levels of evidence are judged and the source of evidence for inclusion. There are also some differences in terminology. For example, both consortia originally used the term *extensive metaboliser.* The CPIC recently revised that to the less confusing *normal metaboliser*. The DPWG still use the former term, but in this project, the latter term has been used for both sets of guidelines in the

interest of clarity. Differences may be due to use of different methodologies, due to timing, where one organisation has published or updated guidelines more recently and so taken account of recent additions to the literature, or even due to different international practices. For example, the DPWG gives no guidance as to maintenance doses of warfarin as all warfarin prescribing in the Netherlands is done according to INR, and it was considered unnecessary.

5.3.5.7.1 Differences in drugs with guidelines

One difference is that there are various drugs that are dealt with by only one or other organisation. For example, the DPWG discusses drugs that the CPIC does not, but generally concludes that there is insufficient evidence to alter dosage. It may recommend increased monitoring for some diplotypes.

5.3.5.7.2 Differences in haplotype, diplotype or genotype function

There are a few genes where haplotypes, diplotypes or genotypes are considered as having different functional effects by each organisation, and therefore may have conflicting prescribing advice associated with them. Discordant guidelines are a problem for the translation of pharmacogenomics into clinical practice, and one solution is to have an accepted set of guidelines for the UK. This is further discussed in Chapter 6.

A gene with this issue is *CYP2C19*, where *1/*17 is considered a rapid metaboliser by CPIC, but a normal metaboliser by DPWG. Historically, the CPIC had considered *1/*17 an ultra-rapid metaboliser, but subsequently introduced the term rapid metaboliser for *1/*17, as the effect lies somewhere between that of a normal metaboliser and ultra-rapid metaboliser. This is only the case in guidelines published since the end of 2016(520). The earlier literature does not make this distinction, which makes giving prescribing advice difficult. The difference between the CPIC and DPWG classifications occurred despite the same reference literature being used. While the difference in prescribing guidance was set out in the long prescribing guidance with a note of the conflict, only the CPIC guidance was given in the summary prescribing sheet. This is because the CPIC guidelines are more up-to-date than the English language version of the DPWG guidelines, which date from 2011 and may not include updates(87). However, the DPWG guidelines are being used across Europe as part of the PREPARE study, so this might add weight to choosing DPWG guidelines as the default UK guidelines(639).

Another difference was the *8 haplotype of *CYP2C9*. While irrelevant in this study because it was not seen and is rare in those not of African ancestry, it is classified as a reduced function haplotype by CPIC, and as an increased function haplotype by the DPWG(650-652). This difference has arisen because the CPIC guideline is based on more recently published literature than the DPWG guideline and highlights the need both for consensus guidelines and a system of updates.

There are also some differences in the assignment of *CYP2D6* activity levels to diplotypes. This does not have an effect in this study, where the classification of the diplotypes into metaboliser types was used, but would make a significant difference if the activity level was used(87). In

addition, there is discordance in the classification of the rare *36 haplotype, but it was not seen in this study.

Prior to 2017, the CPIC described *DPYD* in terms of haplotypes, while the DPWG used activity levels, while indicating which haplotypes fell into which activity category. The updated CPIC guidelines now use activity levels instead of haplotypes, bringing them into line with DPWG guidelines. All patients in this study fell into the category of normal metabolisers, with the exception of those for whom advice could not be given due to uninterpretable diplotypes.

5.3.5.7.3 Differences in prescribing advice

In many cases, even when there is no difference in functional assignation of haplotypes, there are differences in prescribing advice. Full details can be seen in the comparison by Bank et al.(87). Very often the differences are only for a single metaboliser type, or may be relatively minor such as one guideline recommending therapeutic level monitoring and the other not, or the CPIC discussing a class of drugs, where the DPWG restricts advice to a single drug. Occasionally, they are more major, such as in the case of *CYP2D6* and doxepin, where the DPWG does not recommend any change in starting dose regardless of metaboliser type, while the CPIC recommends using an alternative drug or monitoring levels closely for ultra-rapid metabolisers and avoiding or reducing dose by 50% for poor metabolisers.

5.3.5.8 Updating Guidelines

As per section 5.3.5.7.2, guidelines may change to reflect new knowledge. A plan for periodic review of guidelines is essential for keeping then up-to-date and relevant. Both the CPIC and DPWG have review procedures in place, 2 yearly and as required but at most 4 yearly respectively, and guidelines are updated to reflect this. As noted above, the DPWG do not update the English-language guidelines as frequently as the Dutch ones. When pharmacogenomic data are generated and prescribing advice returned to patients, thought should be given as to whether and how that information will be updated to take account of changes in guidelines. In addition, the CPIC has a list of guidelines it intends to publish shortly, including *NUDT15* and possible dosing recommendations for thiopurines. It is unclear at time of writing whether this will be incorporated into the existing *TPMT* guidelines or will stand alone. This is discussed in Chapter 6.

5.3.6 Alteration of prescribing

Even in this small cohort, there were individuals who had already been prescribed drugs that have pharmacogenomic guidelines associated with them. Two individuals had variants in the relevant genes that would result in altered prescribing. As mentioned in section 5.3.5.6, few paediatric pharmacogenomic guidelines exist, and but in the case of thiopurines, testing of paediatric patients and alteration of dose on that basis is common(616). It could be hypothesised that these children might benefit from starting on a lower dose of these medications.

5.4 Conclusions and future directions

Validation of these data is discussed in Chapter 6. Altering prescribing based on pharmacogenomic data have the potential to reduce adverse drug effects and increase prescribing efficacy. Currently only a small number of pharmacogenomic genes have robust prescribing guidelines associated with them but this will increase. Utilising WGS to obtain pharmacogenomic data is beneficial. Not only does it have the advantage of adding value to WGS but it allows for the identification of even very rare haplotypes, the retrospective calling of novel haplotypes or genes for which prescribing guidelines are later developed, and the potential to be used in research to look for novel pharmacogenes or variants.

It is possible to extract pharmacogenomic data from WGS. This is time consuming to do manually, and would benefit from being automated. Automation might also allow a wider range of haplotypes to be called. Software such as Astrolabe can help with extracting copy number variants and identifying rare *CYP2C9*, *CYP2C19* and *CYP2D6* haplotypes, although manual confirmation of these may be required. The data extracted during this study appear to be correct, as in general, the haplotype frequencies are not significantly different to published data. Pharmacogenomic guidance does not cover all relevant drugs and haplotypes. Ideally, a consensus set of haplotypes should be chosen, based on the ethnic make-up of the population being tested.

Presenting a full pharmacogenomic profile is difficult, as it may contain prescribing recommendations for many drugs. Summary guidance is more legible and reduces the risk of missing important data, but it is likely that an app or web-based platform would be better still, both in allowing the information to be accessed when required and allowing it to be updated as necessary. Whichever is chosen, prescribing guidance needs to be clear and easy to implement.

Several sets of mainly concordant guidelines exist. To avoid confusion, a single set should be chosen. As there is no UK consensus, in this study both CPIC and DPWG guidelines were used, along with additional guidance where available.

Future directions include automating the SNP calling process and reviewing published guidelines regularly. More study will be needed into which SNPs should be looked at when considering *HLA-B* *15:02 haplotypes, particularly in mainly Caucasian populations, where *HLA-B* *15:02 positive individuals are likely to be extremely rare. In addition, it has not yet been possible to use WGS data to determine *HLA-B* *44 and *58:01 haplotypes. Further work needs to be done to look at whether this is possible. If additional DNA could be obtained, Sanger sequencing could be done to determine the *DPYD* haplotypes for individuals for whom it was unclear, and try to understand why these results were indeterminate.

Chapter 6 Validation of whole genome sequence pharmacogenomic data

6.1 Introduction

While extraction of pharmacogenomic data from whole genome sequences (WGS) appeared to be straightforward, it was important that this was confirmed using another method. Confirmation of genetic test results obtained in a research setting is widely practised in the NHS, for example, when 100,000 Genomes results are confirmed with Sanger sequencing prior to their return to patients. It was particularly important in the context of this study for two reasons. Firstly, there are limited published data on the extraction of pharmacogenomic data from WGS. Secondly, pharmacogenes are complex and many have a high degree of homology to one another, or to various pseudogenes, thereby increasing the potential difficulties in interpreting the data.

6.1.1 Methods of validating data

6.1.1.1 Commercially available non-pharmacogenomic SNP arrays

Commercially available SNP arrays are one possible method of validating data. Many are available, with more coming to the market all the time. The benefits of these are that they cover very large numbers of SNPs and that the SNPs are often well validated. The disadvantages include the fact that they are not customisable, do not always include pharmacogenomic SNPs, may require moderate volumes of DNA and may be prohibitively expensive for small sample numbers. At the time that validation was under consideration for this project, the coverage of pharmacogenomic SNPs was poor, though this has subsequently improved.

6.1.1.2 Commercially available SNP-genotyping pharmacogenomic testing

The advantages of these tests are that they are cheaper than the very large scale SNP arrays, they require very small amounts of DNA, the individual SNP tests are validated and they cover many of the SNPs of interest, including some not covered by the larger non-pharmacogenomic SNP arrays. The disadvantages are that again, they are not customisable and do not allow the identification of CNVs, such as those seen in *CYP2D6*.

6.1.1.3 Custom SNP-genotyping pharmacogenomic testing

Various customisable commercial genotyping tests are available, such as the Canon Biomedical genotyping tests. As with other tests mentioned above, these are available on a research basis only. The advantages are that tests can be customised to reflect SNPs of interest and specific pharmacogenes can be targeted. Disadvantages include the high cost, the high DNA requirement as they involve large numbers of individual assays and that they often do not work well without positive and negative controls, which have to be purchased separately.

6.1.1.4 Sanger sequencing of individual variants

Sanger sequencing has long been considered the gold standard for confirming sequence variants in the NHS. However, this method is time consuming and expensive, requiring primers to be designed for each SNP unless they are located in close proximity. DNA requirements for Sanger sequencing each variant discussed in Chapter 5 would be high. Furthermore, Sanger sequencing of genes that are highly homologous to other genes or pseudogenes may be technically difficult.

6.1.2 Choice of validation method

There were two main limitations when determining how to validate data. The first of these was the amount of DNA available for validation. Small amounts of DNA had been made available for WGS, and very little remained, often only 50-100ng. This was not enough for most validation methods, as large SNP arrays, custom genotyping and Sanger sequencing all required more DNA than was available without whole genome amplification. The second issue was cost, as there were limited funds available. Based on these constraints, the only feasible validation method was commercial testing that would look only at pharmacogenomic SNPs, reducing the cost and amount of DNA required. Congenica Ltd, a spin-out company from the Wellcome Sanger Genome Centre in Cambridge was setting up a pharmacogenomics assay in the UK. They were using Thermofisher TaqMan assays on the 12Kflex platform, and the test included 177 SNPs, 152 of which were pharmacogenomic SNPs and 25 of which were non-pharmacogenomic SNPs for sample tracking. The assay included most of the actionable pharmacogenes discussed in Chapter 5. However, it excluded some genes that are involved in the metabolism of chemotherapeutic or anti-viral agents and any gene in which testing could result in a carrier or diagnostic test. Therefore *CFTR, DPYD, G6PD and RARG, SLC28A3* and *UGT1A6,* which are considered together, were not included. *CFTR* had already been excluded from analysis (section 5.3.2.3.1).

All of the SNPs examined in the other actionable pharmacogenomics genes were included, with the exception of three genes. The first of these was *HLA-B* *15:02, where alternative SNPs were used (see section 5.3.2.3.4 for discussion of lack of consensus around *HLA-B* *15:02 tag SNPs). The second was *TPMT* where only SNPs for *2, *3A, *3B, *3C and *4 were included, the remainder being very rare, especially in individuals of European ancestry. The third was *UGT1A1* where rs817534 was not present. This is because this SNP is for the insertion of a variable number of repeats, and so cannot be checked in SNP genotyping assays. However, rs887829, which represents *UGT1A1* *80, is in high linkage disequilibrium with *28 and *37, both represented by the number of repeats at rs817534. If only *80 is interrogated, an intermediate or poor metaboliser phenotype can be inferred from heterozygosity and homozygosity respectively (www.pharmGKB.com), meaning that this exclusion was not detrimental to the study. In the case of all other genes discussed in Chapter 5, all the SNPs checked during WGS data extraction were present, and in many cases, additional SNPs also. Of the SNPs identified by Astrolabe (section 5.2.4), both the SNPs for the *29 and *35 haplotypes of *CYP2D6* were included in the assay. The SNP for the *15 haplotype of *CYP2C19* was not included in the assay. SNPs not covered are listed in Table 6.1. Some of the SNPs had not been available as off the shelf assays but had been

designed and validated by Congenica Ltd. SNPs for which this was the case are listed in Table 6.2.

Congenica Ltd agreed to allow me to use their equipment and assay to run my validation testing in order to see how the system worked and what the quality control data were like before offering it as a commercial service. The Congenica Ltd pharmacogenomics assay (Congenica PGx) was therefore the most cost effective and efficient solution, worked for the limited amount of DNA that remained and allowed the confirmation of most of the variants of interest. It also allowed the samples to be run and interpreted by us rather than a commercial company.

rs number/ Transcript	Gene	Haplotype	Solution
rs55752064	*CYP2C19*	*15	No solution found, WGS data (Astrolabe) cannot be confirmed
rs2395148	*HLA-B*	*15:02	Congenica SNPs checked in WGS data
rs10484555	*HLA-B*	*15:02	Congenica SNPs checked in WGS data
rs144012689	*HLA-B*	*15:02	Congenica SNPs checked in WGS data
rs2524160	*HLA-B*	*15:02	Congenica SNPs checked in WGS data
rs72552740	*TPMT*	*5	Not seen in WGS data
rs75543815	*TPMT*	*6	Not seen in WGS data
rs72552736	*TPMT*	*7	Not seen in WGS data. Not in PharmGKB haplotype set
rs56161402	*TPMT*	*8	Not seen in WGS data
rs151149760	*TPMT*	*9	Not seen in WGS data
rs72552737	*TPMT*	*10	Not seen in WGS data
rs72552738	*TPMT*	*11	Not seen in WGS data
rs200220210	*TPMT*	*12	Not seen in WGS data
rs72552742	*TPMT*	*13	Not seen in WGS data
rs9333569	*TPMT*	*14	Not seen in WGS data
rs9333570	*TPMT*	*15	Not seen in WGS data. Not in PharmGKB haplotype set
rs144041067	*TPMT*	*16	Not seen in WGS data. Not in PharmGKB haplotype set
NM_000367.2: c.124C>G	*TPMT*	*17	Not seen in WGS data
NM_000367.2: c.124C>G	*TPMT*	*17	Not seen in WGS data
rs8175347	*UGT1A1*	*28, *36, *37	Data can be inferred from rs887829

Table 6.1 SNPs checked in whole genome data not present in Congenica Ltd PGx assay

rs number	Gene	Haplotype
rs72558189	*CYP2C9*	*14
rs138100349	*CYP2D6*	*44
rs72549352	*CYP2D6*	*21
rs72549353	*CYP2D6*	*19
rs72549354	*CYP2D6*	*20
rs138105638	*CYP3A4*	*26
rs67666821	*CYP3A4*	*20
rs1061235	*HLA-A*	*31:01
rs25531	*SLC6A4*	n/a
rs4149015	*SLCO1B1*	*17 and *21
rs1976391	*UGT1A1*	n/a
rs2011425	*UGT1A4*	n/a

Table 6.2 SNPs which required a custom assay design in Congenica Ltd PGx assay

In addition to the actionable pharmacogenes discussed in Chapter 5, 64 additional pharmacogenes had one or more SNPs on the Congenica PGx panel. While these were not examined during original analysis, the data were later extracted from the WGS data using the same method as previously (Chapter 2) and compared with results from the Congenica PGx panel to determine whether they were accurately called from WGS data. Genes included in the assay are listed in Table 6.3, and a full list of SNPs can be found in Supplementary Information 2.4 (CD-ROM).

AARS	COG1	DRD2	HTR2A	NKD2	STK32B
ABCB1	COL4A4	EVC	HTR2C	NPHS2	SUMF1
ABCB11	COMT	F2	IFNL3	OPRD1	TDRD7
ABCG2	CYP1A2	F5	ITGB3	OPRM1	TNFRSF4
ACE	CYP2B6	FERMT1	KCNJ6	PDP1	TPMT
ADRA2A	CYP2C19	FKBP5	KIF6	PLCG1	UGT1A1
ADRB1	CYP2C9	FREM2	L2HGDH	POLG	UGT1A4
ADRB2	CYP2D6	GRIK4	LAMA3	POR	VCAN
AGTR1	CYP3A4	GRIN2B	LILRB5	PRSS53	VKORC1
ANKK1	CYP3A5	GRIN3B	LPA	SLC12A6	WNK1
ATM	CYP4F2	GRK5	MTHFR	SLC6A4	YEATS4
B9D2	DBH	HLA-A	MTR	SLCO1B1	
BDKRB1	DMRT2	HLA-B	MUC6	SNTG2	
CACNA1C	DNMT1	HPSE2	NDUFV3	SOX6	

Table 6.3 Genes with SNPs included in the Congenica Ltd PGx assay

6.1.3 Additional pharmacogenes

Many of the additional genes included in the Congenica PGx are listed on PharmGKB. Some have Food and Drug Administration (FDA) guidelines associated with them, which are used in the USA to guide prescribing. However, many of them do not have the weight of evidence required for the formulation of peer-reviewed prescribing guidelines, such as those developed by the DPWG or CPIC (section 5.1.2.1). This may be because there is limited information about the functional effect of different haplotypes, or because even if haplotype function is well understood, there may be insufficient information about how prescribing could be altered. As discussed in section 5.3.5, guidelines are being developed and updated continuously, so it is likely that some will have guidelines published in the future. Some of the genes listed are not known pharmacogenes and are present to allow sample tracking.

6.1.3.1 *AARS* (alanyl-tRNA synthetase)

AARS is implicated in infantile epileptic encephalopathy, a form of Charcot-Marie-Tooth disease and other neuropathies(653, 654). There is limited evidence to suggest that it may be implicated in methotrexate resistance(655). While it is listed in PharmGKB, there is no clinical, prescribing or variant information available.

6.1.3.2 *ABCB1* (ATP-binding cassette, sub-family B (MDR/TAP), member 1)

ABCB1 is listed as a very important pharmacogene (VIP) in PharmGKB(488). It is a member of the human adenosine triphosphate (ATP)-binding cassette (ABC) transporter superfamily and has an important role in maintaining cell homeostasis by acting as an efflux pump. *ABCB1* was previously known as multi-drug resistance protein 1 (*MDRP1*) because of its role in conferring resistance to multiple drugs due to their interaction with its product, the p-glycoprotein(656, 657). While likely to play a role in the safe and efficient dosing of many drugs, there is currently very limited guidance available for *ABCB1*, although the Europeans Medicines Agency (EMA) and its Canadian equivalent have produced drug labels for *ABCB1* and aliskiren, a renin inhibitor used for treatment of hypertension.

6.1.3.3 *ABCB11* (ATP-binding cassette, sub-family B (MDR/TAP), member 11)

Another member of the human ATP-binding cassette superfamily, *ABCB11* has been implicated in intrahepatic cholestasis and is believed to act as a bile salt export pump(658). ABCB11 has been implicated in drug-induced liver injury and in possibly in anthracycline-related hepatotoxicity in children(659, 660).

6.1.3.4 *ABCG2* (ATP-binding cassette, subfamily G, isoform 2)

A PharmGKB VIP, *ABCG2* encodes the breast cancer resistance protein (BCRP), which is another efflux pump(661). Resistance to chemotherapeutic agents has been found in individuals overexpressing BCRP(662). PharmGKB lists multiple clinical annotations, including for statins, chemotherapeutic agents, immunosuppressants and allopurinol(663-666).

6.1.3.5 *ADRA2A* (adrenoceptor alpha 2A)

ADRA2A encodes an adrenaline (epinephrine) receptor and there is evidence that polymorphisms in this and similar genes may be implicated in attention deficit hyperactivity disorder (ADHD) and type II diabetes(667-669). Pharm GKB lists clinical annotations for *ADRA2A* and drugs including SSRIs, atenolol and methylphenidate, a treatment for ADHD(670-672).

6.1.3.6 *ADRB1* (adrenoceptor beta 1)

ADRB1 encodes the Beta-1 adrenergic receptor, a cardiac G-protein coupled receptor with a role in regulating cardiac contraction rate and force(673). Polymorphisms in *ADRB1* have been associated with drug response in heart failure and dilated cardiomyopathy(674, 675). There are several clinical annotations listed, mainly for antihypertensives including beta-blockers(676, 677).

6.1.3.7 *ADRB2* (adrenoceptor beta 2)

ADRB2 encodes the Beta-2 adrenergic receptor, which is widely expressed in human tissue especially in pulmonary, cardiac, endocrine and central nervous system tissue, making it a treatment target for common conditions including asthma and hypertension(485, 678, 679). The EMA (www.ema.europa.eu) has released a drug label for *ADRB2* and indacaterol, but it does not include prescribing advice. PharmGKB also lists many clinical annotations, including for beta-blockers and other antihypertensives, and asthma treatments including bronchodilators and corticosteroids(677, 680-682).

6.1.3.8 *AGTR1* (angiotensin II receptor, type 1)

AGTR1 has been implicated in renal tubular dysgenesis and essential hypertension and is the target of angiotensin II(683-685). Clinical annotations in PharmGKB are for antihypertensive agents including ACE inhibitors, angiotensin II receptor blockers and thiazide diuretics(686-688).

6.1.3.9 *ANKK1* (ankyrin repeat and kinase domain containing 1)

A polymorphism in *ANKK1* has been linked to dopamine D2 receptor density(689). *ANKK1* has therefore been linked to efficacy of smoking cessation treatments, but also to antipsychotic and anti-epileptic toxicity(689-692).

6.1.3.10 *ATM* (ataxia telangiectasia mutated gene)

ATM has been implicated in ataxia telangiectasia and therefore in malignancy(693). It is a phosphatidyl-3-kinase that has a role in DNA repair. There is some evidence to suggest that variants in or near *ATM* may affect metformin metabolism(694).

6.1.3.11 *BDKRB1* (bradykinin receptor B1)

BDKRB1 is expressed in response to tissue injury(695). Evidence from a large study suggests that a *BDKRB1* variant may affect risk of cardiac events in those taking perindopril(688).

6.1.3.12 *CACNA1C* (calcium channel, voltage-dependent, L type, alpha 1C subunit)

CACNA1C encodes a subunit of a calcium channel important in cardiac conduction, and is implicated in Brugada and Timothy syndromes, in both of which cardiac arrhythmias are seen(696, 697). Polymorphisms have been implicated in the efficacy of calcium channel blockers and may increase the risk of suicide in individuals prescribed citalopram for depression(698-700).

6.1.3.13 *COMT* (catechol-O-methyltransferase)

Variants in *COMT* have been implicated in the development of neuropsychiatric disorders such as schizophrenia and also in response to neuroleptics and the likelihood of developing addictions(701-703). PharmGKB lists many drugs where variants in *COMT* may have an effect(471).

6.1.3.14 *CYP1A2* (cytochrome P450, family 1, subfamily A, polypeptide 1)

CYP1A2 is a member of the cytochrome P450 family. It appears to play a role in metabolism of many drugs including caffeine and related compounds, beta-blockers and antipsychotics(704-706).

6.1.3.15 *CYP2B6* (cytochrome P450, family 2, subfamily B, polypeptide 6)

CYP2B6, another cytochrome P450 gene, is expressed mainly in the liver and the brain(474). It has a role in the metabolism of multiple drugs including efavirenz, for which it has both FDA and EMA labels(707).

6.1.3.16 *CYP3A4* (cytochrome P450, family 3, subfamily A, polypeptide 4)

CYP3A4 is also implicated in the metabolism of multiple drugs. The EMA recommends testing when treating with gefitinib, especially if treating with *CYP3A4* inhibitors or in patients who are *CYP2D6* poor metabolisers. Many other drugs are listed for which *CYP3A4* may alter metabolism, including tacrolimus and atorvastatin(589, 708).

6.1.3.17 *DBH* (dopamine beta-hydroxylase)

DBH is implicated in nicotine requirement, and variants have been associated with heavy smoking, opioid dependence and naltrexone response(709-711).

6.1.3.18 *DRD2* (dopamine receptor D2)

DRD2 is one of a family of dopamine receptors and a target of antipsychotic drugs and dopamine agonists used to treat Parkinson's disease(712-714). PharmGKB lists multiple drugs that have clinical annotations for *DRB2*, though levels of evidence vary; among these are cocaine, antipsychotics and opioids(715-717).

6.1.3.19 *F2* (factor 2)

F2 encodes a clotting factor, prothrombin, which is converted to thrombin, the active form, by factor Xa(718). Some variants can cause dysprothrombinaemia, which predisposes to haemorrhage, while others predispose to thrombosis(719, 720). Drug labels for *F2* include one

for various hormonal contraceptives (Health Canada Santé Canada (HCSC) and also for tamoxifen (FDA), both of which can increase the risk of thrombosis(721). Other variants may affect warfarin dose(722).

6.1.3.20 *FKBP5* (FK506 binding protein 5)

FKBP5 has a role in immunosuppression and binds tacrolimus and rapamycin(723). Polymorphisms appear to affect response to antidepressants and corticosteroids(724, 725).

6.1.3.21 *GRIK4* (glutamate receptor, ionotropic, kainate 4)

GRIK4 encodes a subunit for an ion channel activated by glutamate, an excitatory neurotransmitter(726). Variants have been associated with antidepressant response(727).

6.1.3.22 *GRIN2B* (glutamate receptor, ionotropic, N-methyl D-aspartate 2B)

GRIN2B encodes a subunit of an NMDA receptor activated by glutamate, and mutations in it have been associated with a form of infantile epileptic encephalopathy and with autosomal dominant mental retardation(728-730). Variants are associated with antipsychotic toxicity and efficacy among other things(731, 732).

6.1.3.23 *GRK5* (G protein-coupled receptor kinase 5)

GRK5 encodes a regulator of G protein-coupled receptors and is involved in the regulation of the cell cycle(733). Polymorphisms appear to be protective in heart failure as well as affecting response to antihypertensives and antidepressants(734, 735).

6.1.3.24 *HTR2A* (5-hydroxytryptamine (serotonin) receptor 2A, G protein-coupled)

HTR2A is a serotonin receptor. Variants in this gene have been implicated in antidepressant, antipsychotic and analgesic response(736-738).

6.1.3.25 *HTR2C* (5-hydroxytryptamine (serotonin) receptor 2C, G protein-coupled)

Also a serotonin receptor, variants in *HTR2C* have been associated with antipsychotic toxicity(739).

6.1.3.26 *IFNL4* (interferon, lambda 4)

Encoding a type III interferon, *IFNL4* has a role in protecting the body against viruses, similar to *IFNL3* (Chapter 5). Interestingly, *IFNL4* is a pseudogene in those who lack a particular frameshift variant, rs368234815 at position 19:39248514(740). If the delG variant is present, the frameshift results in the non-production of IFNL4, which is associated with clearance of hepatitis C. This and additional SNPs have been implicated in response to antivirals(741-743).

6.1.3.27 *ITGB3* (integrin, beta 3 (platelet glycoprotein IIIa, antigen CD61)

IGTB3 is a subunit of a glycoprotein receptor found on the surface of platelets and is implicated in various bleeding disorders(744-746). Variants may affect efficacy of anti-platelet agents(747, 748).

6.1.3.28 *KCNJ6* (potassium inwardly-rectifying channel, subfamily J, member 6)

KCNJ6 encodes an ATP-sensitive, G protein-coupled potassium channel, which appears to be involved in insulin secretion(749). Polymorphisms may affect opiate requirements(750, 751).

6.1.3.29 *KIF6* (kinesin family member 6)

KIF6, a component of the microtubule motor, is involved in intracellular transport, and variants have been associated with statin efficacy(752).

6.1.3.30 *LILRB5* (leukocyte immunoglobulin-like receptor, subfamily B, member 5)

The receptor encoded by *LILRB5* is one of a family of immune receptors found on monocytes, B cells and natural killer cells and acts as a negative regulator of the immune system(753). A variant has been implicated in statin-related myopathy(754).

6.1.3.31 *LPA* (lipoprotein, Lp(a))

LPA encodes lipoprotein (a) which has been identified as a risk factor for cardiovascular events including atherosclerosis and stroke(755). Variants have been associated with statin efficacy and toxicity(756).

6.1.3.32 *MTHFR* (methylenetetrahydrofolate reductase)

MTHFR encodes the rate-limiting enzyme in the methyl cycle and is important in homocysteine remethylation(757). There is a drug label for MTHFR and oral contraceptives (HCSC). PharmGKB lists many clinical annotations for a wide range of drugs including chemotherapeutic agents and antipsychotics(758, 759).

6.1.3.33 *NOS1AP* (nitric oxide synthase 1 (neuronal) adaptor protein)

The protein product of *NOS1AP* binds neuronal nitric oxide synthase, and appears to have a role in cardiac repolarisation(760). Variants may increase the risk of QT interval prolongation and arrhythmias in patients receiving drugs such as amiodarone and verapamil(761, 762).

6.1.3.34 *NOS3* (nitric oxide synthase 3)

Also known as endothelial nitric oxide synthase, this gene encodes one of three enzymes that synthesise nitric oxide (NO). *NOS3* is responsible for NO production in the vascular endothelium, where it has multiple roles(763). Variants have been associated with efficacy of antihypertensives and anti-cancer agents among others(677, 764, 765).

6.1.3.35 *OPRD1* (opioid receptor, delta 1)

As suggested by its name, *OPRD1* encodes an opioid receptor important in analgesic response(766). Variants are implicated in opioid efficacy and dependence(767-769).

6.1.3.36 *OPRM1* (opioid receptor, mu 1)

Another opioid receptor, OPRM1 is also implicated in opioid response(750, 770). Variants may affect the risk of substance abuse(771).

6.1.3.37 *POLG1* (polymerase (DNA directed), gamma)

POLG1 is a nuclear gene that encodes a subunit of the mitochondrial DNA polymerase, and pathogenic variants in it have been associated with several mitochondrial diseases including Alpers and MNGIE type DNA depletion syndromes and progressive external ophthalmoplegia(772-774). Both the FDA and HCSC have drug labels for *POLG1* and divalproex sodium and valproic acid in people who have, or are at risk of having, mitochondrial disease. Variants may also increase the risk of liver failure in individuals without mitochondrial disease(775).

6.1.3.38 *POR* (P450 (cytochrome) oxidoreductase)

POR is required for electron donation to all of the microsomal P450 genes, and mutations in it have been implicated in Antley-Bixler syndrome with genital anomalies and disordered steroidogenesis and other disorders of steroid synthesis(776). Variants in *POR* appear to affect immunosuppressant efficacy in combination with other genes(777).

6.1.3.39 *PRSS53* (protease, serine, 53)

PRSS53 encodes a serine protease that appears to act on fibrinogen alpha(778). A variant appears to influence warfarin requirement in individuals of African-American ancestry(779).

6.1.3.40 *SLC6A4* (solute carrier family 6 (neurotransmitter transporter), member 4)

SLC6A4 encodes a transporter responsible for serotonin clearance(780). Variants may be important in mediating toxicity and efficacy of antidepressants(781-783).

6.1.3.41 *UGT1A4* (UDP glucuronosyltransferase 1 family, polypeptide A4)

The product of *UGT1A4*, like that of *UGT1A1*, is involved in the glucuronidation of substrates to make them water soluble and excretable. Along with *UGT1A1*, it is one of a number of genes encoded by a complex locus, where splicing allows a variable first exon to be combined with other invariant exons(784). Variants appear to influence efficacy and toxicity of antiepileptic drugs among others(785, 786).

6.1.3.42 *WNK1* (WNK lysine deficient protein kinase 1)

WNK1 encodes a serine-threonine kinase and has a role in regulating co-transporters of the kidney. Mutations in it can cause a form of pseudohypoaldosteronism and hereditary sensory neuropathy(787, 788). A variant has been implicated in the efficacy of hydrochlorothiazide(789).

6.1.3.43 *YEATS4* (YEATS domain containing 4)

The product of *YEATS4* is believed to be a transcription factor implicated in tumorigenesis(790). A single variant is thought to affect hydrochlorothiazide efficacy(791).

6.1.3.44 Sample tracking SNPs

Other genes listed in Table 6.3 above are not pharmacogenes but are included for sample tracking purposes(792). They include single SNPs in *B9D2, COG1, COL4A4, DMRT2, DNMT1,*

EVC, FERMT1, FREM2, GRIN3B, HPSE2, L2HGDH, LAMA3, MTR, MUC6, NDUFV3, NKD2, NPHS2, PDP1, PLCG1, SLC12A6, SNTG2, SOX6, STK32B, SUMF, TDRD7 and *TNFRSF4.* Sample tracking SNPs allow post-hoc sample identification. They are easy to genotype and found at a population frequency of 20-80% in all major ethnicities.

6.1.4 Samples unavailable for validation

Owing to a shortage of DNA, not all samples were available for validation. No DNA from the children of the SRS parent-child trios was available (SRS-001, SRS-004, SRS-007, SRS-010, SRS-013, SRS-016, SRS-019, SRS-022, SRS-025 and SRS-028). DNA was also unavailable for BBS-016 (although DNA was available for his monozygotic twin, BBS-017, who had the same results from WGS), JDM-011, IBD-004, IBD-020, SRS-005 and SRS-015. All other samples had DNA of sufficient quality and concentration for validation. Overall, 68 samples were validated.

6.1.5 Aims

The aim of this chapter was to use SNP genotyping to confirm the results obtained from WGS data including those that were obtained using Astrolabe, and in particular to identify problematic SNPs or genes. In addition, it aimed to look at some of the benefits, disadvantages and challenges of clinical implementation of pharmacogenomic testing.

6.2 Results

In all results tables, *ref* means homozygous for the reference nucleotide, *het* means heterozygous for the reference and non-reference nucleotides and *hom* means homozygous for the non-reference nucleotide. The SNP calls can be seen in Supplementary Information S2.4 (CD-ROM).

6.2.1 Comparison of SNP and WGS data for genes with prescribing guidelines

6.2.1.1 *CYP2C9*

In the case of *CYP2C9*, there were no disagreements between SNP and WGS data. The *9 haplotype seen in IBD-008 in Astrolabe and retrospectively confirmed in WGS data was also confirmed in SNP data. The *11 haplotype seen in SRS-011 in the Astrolabe data was also confirmed. However, in the case of 6 assays, a single sample failed to amplify. Possible explanations for this are discussed in section 6.3.1.4.3. In each case a different sample failed to amplify. With the exception of IBD-003, which was the only sample where the diplotype in the WGS data was called as *3/*3, another sample with the same diplotype amplified successfully in the same assay. However, multiple samples with a *1/*3 diplotype amplified successfully in that assay. In addition, homozygosity was identified for rs1057910 in sample IBD-003, confirming the *3/*3 diplotype.

rs number	Haplotype	Checked in WGS	Samples concordant	Samples discordant	Samples failing to amplify	Comment
rs1799853	*2	N	68	0	0	
rs1057910	*3	Y	67	0	1	SRS-008 failed to amplify .*1/*1 in WGS
rs56165452	*4	N	67	0	1	IBD-003 failed to amplify. *3/*3 in WGS
rs28371686	*5	Y	68	0	0	
rs9332131	*6	Y	68	0	0	
rs7900194(a)	*8	Y	68	0	0	
rs2256871	*9	N	68	0	0	
rs28371685	*11	Y	68	0	0	
rs9332239	*12	N	67	0	1	SRS-012 failed to amplify. *1/*2 in WGS
rs72558187	*13	N	67	0	1	IBD-016 failed to amplify. *1/*1 in WGS
rs72558189	*14 & *35	N	67	0	1	SRS-006 failed to amplify. *1/*1 in WGS
rs72558190	*15	N	67	0	1	IBD-014 failed to amplify. *1/*1 in WGS
rs72558192	*16	N	68	0	0	

Table 6.4 Validation of SNPs in CYP2C9

6.2.1.2 *CYP2C19*

In the case of *CYP2C19*, there were no disagreements between WGS data and SNP data and all the called diplotypes were confirmed with one exception. The SNP for the *15 haplotype seen in Astrolabe and retrospectively confirmed in WGS data was not included in the panel and therefore could not be confirmed. In the case of two assays a single sample failed to amplify and in the case of a single assay, 2 samples failed to amplify. In each case a different sample failed to amplify and another sample with the same diplotype amplified without difficulty. In all cases, with the exception of BBS-005 where no variants were identified in keeping with the *1/*1 haplotype,

the expected SNPs were seen for each of these samples. Of interest, both samples that failed to amplify for the rs55640102 assay had the *1/*17 diplotype, although other samples with this amplified normally.

rs number	Haplotype	Checked in WGS	Samples concordant	Samples discordant	Samples failing to amplify	Comment
rs4244285	*2	Y	67	0	1	BBS-005 failed to amplify. *1/*1 in WGS
rs4986893	*3	Y	68	0	0	
rs28399504	*4	Y	67	0	1	BBS-001 failed to amplify. *1/*2 in WGS
rs12248560	*4B & *17	Y	68	0	0	
rs56337013	*5	Y	68	0	0	
rs72552267	*6	Y	68	0	0	
rs72558186	*7	Y	68	0	0	
rs41291556	*8	Y	68	0	0	
rs17884712	*9	N	68	0	0	
rs6413438	*10	N	68	0	0	
rs55640102	*12	N	66	0	2	SRS-008 and SRS-030 failed to amplify. Both *1/*17 in WGS

Table 6.5 Validation of SNPs in CYP2C19

.

6.2.1.3 *CYP2D6*

In the case of *CYP2D6*, there were no disagreements between WGS and SNP data and all the called diplotypes were confirmed, except in the case of CNVs which cannot be confirmed with a SNP-based assay. However, it confirmed the haplotypes that would be present if the copy number variants were excluded. rs16947, which is part of the *41 haplotype, did not amplify for BBS-011. In the case of BBS-011, the *1/*41 diplotype can be confirmed without this SNP and it was also confirmed in the monozygotic twin. In the case of two other assays a single sample failed to amplify. In each case a different sample failed to amplify and another sample with the same diplotype amplified without difficulty. In both cases the expected SNPs were seen for each of these samples. The *35 haplotypes picked up by Astrolabe in BBS-015, JDM-012 and USH-002 were confirmed, as was the *29 haplotype in JDM-007. Initially, in patient BBS-002, the data point for rs5030867 did not cluster very well, falling between het and ref clusters. As this was one of

the first assays run, it was repeated on a later run and clustered with the other ref samples in the repeat assay, confirming the diplotype.

rs number	Haplotype	Checked in WGS	Samples concordant	Samples discordant	Samples failing to amplify	Comment
rs16947	*2, *4, *17, *41	N	67	0	1	BBS-011 failed to amplify. *1/*41 in WGS
rs1080985	*2, *35, many	N	68	0	0	
rs35742686	*3	Y	68	0	0	
rs3892097	*4	Y	68	0	0	
rs1065852	*4 & *10	Y	68	0	0	
rs5030655	*6	Y	68	0	0	
rs5030867	*7	N	68	0	0	
rs5030865	*8 & *14	N	68	0	0	
rs5030656	*9	Y	67	0	1	JDM-005 failed to amplify. *1/*41 in WGS
rs201377835	*11	N	68	0	0	
rs5030862	*12	N	68	0	0	
rs28371706	*17	Y	68	0	0	
rs72549354	*20	N	68	0	0	
rs59421388	*29, *70, *109	N	68	0	0	
rs769258	*35	Y	68	0	0	
rs72549351	*38	N	68	0	0	
rs28371725	*41	Y	68	0	0	
rs72549346	*42	N	68	0	0	
rs138100349a	*44	N	68	0	0	
rs147960066	*56	N	68	0	1	BBS-009 failed to amplify. *1/*41 in WGS
rs1135840	Many	Y	68	0	0	
rs72549353	Not in PharmGKB	N	68	0	0	

Table 6.6 Validation of SNPs in CYP2D6

6.2.1.4 *CYP3A5*

In the case of *CYP3A5*, there were no disagreements between WGS and SNP data and all called diplotypes were confirmed. 2 samples failed to amplify in the assay for the *8 haplotype, rs55817950. Both were called as *1/*3 in the WGS data. While other samples with this diplotype amplified normally, it is a rare diplotype and may indicate an area for concern. However the SNP pertaining to the *3 haplotype amplified correctly, confirming the observed diplotype.

rs number	Haplotype	Checked in WGS	Samples concordant	Samples discordant	Samples failing to amplify	Comment
rs10264272	*6	Y	68	0	0	
rs28365083	*2	Y	68	0	0	
rs28383479	*9	Y	68	0	0	
rs41303343	*7	Y	68	0	0	
rs55817950	*8	Y	66	0	2	IBD-017 and JDM-006 failed to amplify. Both *1/*3 in WGS
rs55965422	*5	Y	68	0	0	
rs56411402	*4	Y	68	0	0	
rs776746	*3 & *9	Y	68	0	0	

Table 6.7 Validation of SNPs in CYP3A5

6.2.1.5 *CYP4F2* and rs12777823

In the case of *CYP4F2* and rs12777823, there were no disagreements between WGS and SNP data and no samples failed to amplify.

rs number	Gene	Checked in WGS	Samples concordant	Samples discordant	Samples failing to amplify	Comment
rs2108622	*CYP4F2*	Y	68	0	0	
rs12777823		Y	68	0	0	

Table 6.8 Validation of SNPs in CYP4F2 and rs12777823

6.2.1.6 *F5*

There were no disagreements between WGS and SNP data in the case of *F5*, and no samples failed to amplify.

rs number	Checked in WGS	Samples concordant	Samples discordant	Samples failing to amplify	Comment
rs6025	Y	68	0	0	

Table 6.9 Validation of SNPs in F5

6.2.1.7 *HLA-A* and *HLA-B*

For two samples, SRS-017 and JDM-007, the *HLA-A* *31:01 results disagreed with the WGS results. They were clear heterozygotes in WGS (ratios 26:24 and 38:30 respectively), but clustered as non-reference homozygotes in SNP data, although automated calling software called JDM-007 as heterozygote (green) (Figure 6.1). However, in both cases, other SNPs were seen nearby in WGS data and there were very few true homozygotes in the cohort (section 6.3.1.1.1).

In the case of the assay for the SNP for *HLA-B* *57:01, 2 samples, IBD-015 and SRS-008 failed to amplify. One of these was *HLA-B* *57:01 negative, the other positive. Other samples with these haplotypes amplified normally.

In the case of the assays for HLA-B *15:02 none of the SNPs matched the SNPs chosen for HLA-B *15:02 originally (section 5.3.2.3.4). When the Congenica SNPs were checked, there was no disagreement between SNP and WGS data and no samples failed to amplify. However, multiple samples were heterozygous or homozygous for one or other of the *HLA-B* *15:02 SNPs. As per the original WGS data, samples heterozygous or homozygous for only one SNP were not considered to be *HLA-B* *15:02 positive. Only two samples were heterozygous or homozygous for both SNPs, suggesting they were *HLA-B* *15:02 positive. This disagrees with the WGS data where no samples were called as positive, based on the SNPs chosen originally. IBD-012 was heterozygous for both Congenica SNPs and JDM-001 was heterozygous for rs2844682 and homozygous for rs3909184 but neither was positive for any of the SNPs checked in the WGS data. This is discussed further in section 6.3.1.1.1.

rs number	Gene and haplotype	Checked in WGS	Samples concordant	Samples discordant	Samples failing to amplify	Comment
rs1061235	*HLA-A* *31:01	Y	66	2	0	JDM-007 and SRS-017 disagreed
rs2395029	*HLA-B* *57:01	Y	66	0	2	IBD-015 and SRS-008 failed to amplify. In WGS, IBD-015 was negative, SRS-008 was positive
rs2844682	*HLA-B* *15:02	Y	66	2	0	IBD-012 and JDM-001 suggested to be *HLA-B* *15:02 positive
rs3909184	*HLA-B* *15:02	N	66	2	0	IBD-012 and JDM-001 suggested to be *HLA-B* *15:02 positive

Table 6.10 Validation of SNPs in HLA-A and HLA-B

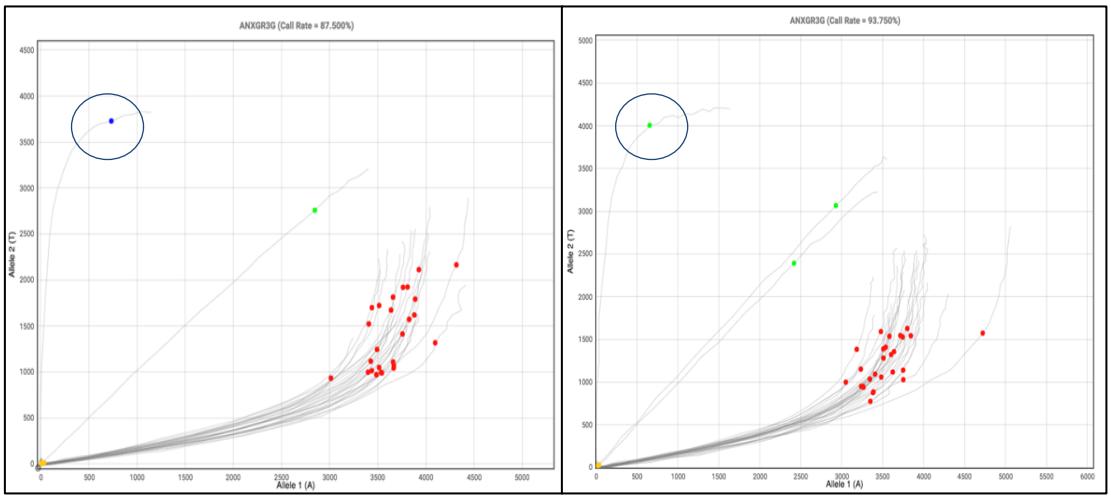

*Figure 6.1 HLA-A *31:01 rs1061235: clustering of samples SRS-017 (left image) and JDM-007 (right image)*

6.2.1.8 *IFNL3*

In the case of *IFNL3*, no samples failed to amplify and there were no disagreements between SNP and WGS data.

rs number	Checked in WGS	Samples concordant	Samples discordant	Samples failing to amplify	Comment
rs12979860	Y	68	0	0	

Table 6.11 Validation of SNPs in IFNL3

6.2.1.9 *SLCO1B1*

In the case of *SLCO1B1*, there was a single disagreement and no failed assays. The disagreement was in IBD-007. It was called as a heterozygote in the assay, but as reference sequence in WGS (Figure 6.2). On re-examination of the WGS data, only 6 reads out of 22 were non-reference (Figure 6.3). An additional SNP was used that was not checked initially during WGS analysis. rs4149056 identifies multiple reduced function alleles and rs4149015 adds some clarity as to which reduced function allele is present. However, several samples did not have the rs4149056 SNP, but did have rs4149015 and other variants which identify the *21 haplotype.

rs number	Haplotype	Checked in WGS	Samples concordant	Samples discordant	Samples failing to amplify	Comment
rs4149056	*5, *15, *17	Y	67	1	0	IBD-007 was called as a het. In WGS, but 6/22 reads were non-ref
rs4149015	*17 & *21	N	68	0	0	

Table 6.12 Validation of SNPs in SLCO1B1

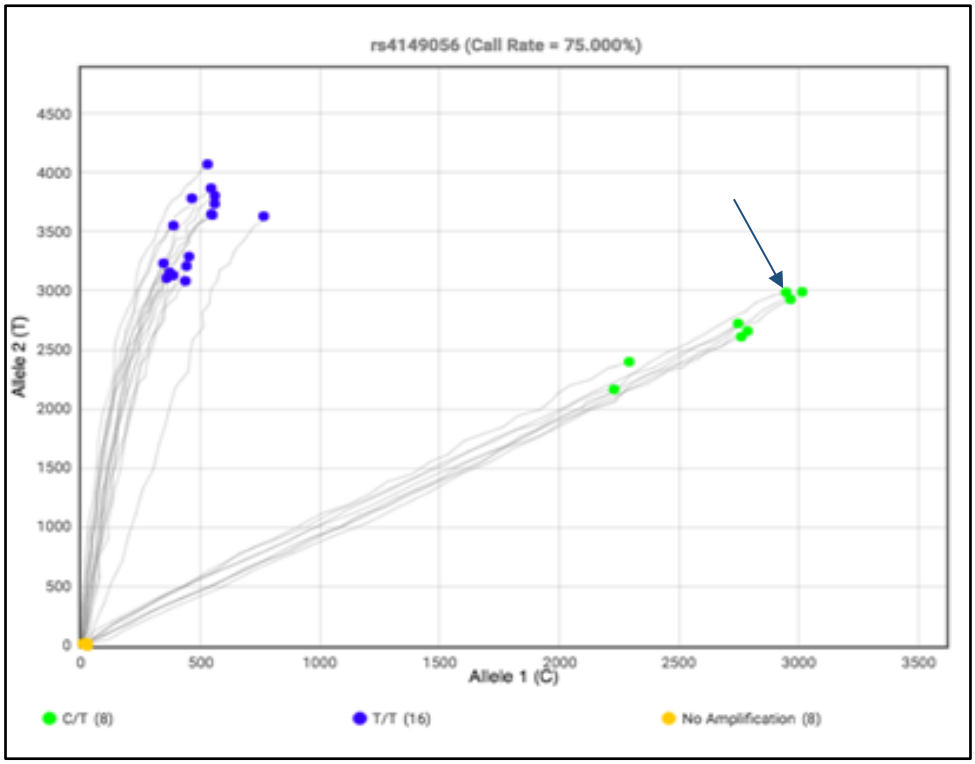

Figure 6.2 SLCO1B1 rs4149056: clustering of IBD-007 (arrowed) with heterozygotes (green)

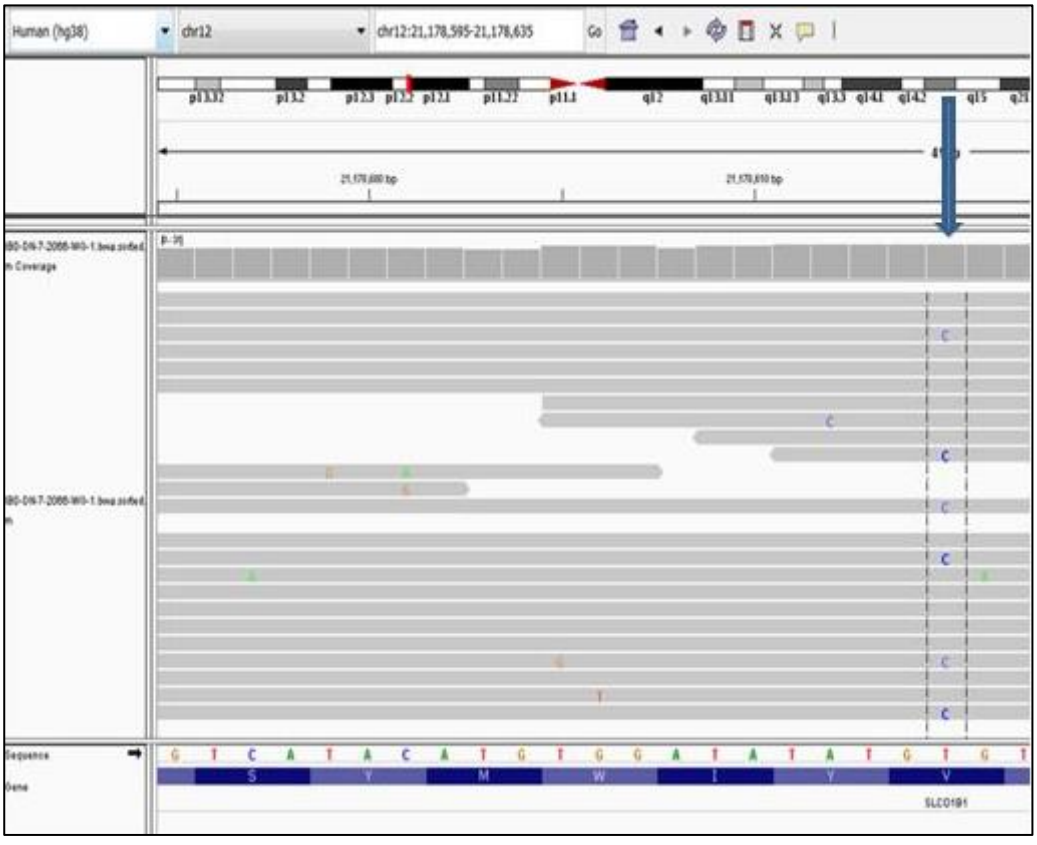

Figure 6.3 SLCO1B1 rs4145106 for IBD-007. 6 of 22 reads are non-reference. Image from IGV

6.2.1.10 *TPMT*

In the case of *TPMT*, there were no discrepancies and no assay failures. Although a smaller number of SNPs were checked than originally checked in the WGS analysis, all identified haplotypes were covered by the Congenica SNPs and all diplotypes were in agreement.

rs number	Haplotype	Checked in WGS	Samples concordant	Samples discordant	Samples failing to amplify	Comment
rs1142345	*3A and *3C	Y	68	0	0	
rs1800460	*3A & *3B	Y	68	0	0	
rs1800462	*2	Y	68	0	0	
rs1800584	*4	Y	68	0	0	

Table 6.13 Validation of SNPs in TPMT

6.2.1.11 *UGT1A1*

In the case of *UGT1A1*, there were no disagreements and no samples failed to amplify for any assay. However, the SNP rs8175347 was not included in the Congenica panel. This is discussed in section 6.3.1.5.3. As no DNA was available for confirmation of IBD-004, the only *36 haplotype seen in the cohort could not be verified.

rs number	Haplotype	Checked in WGS	Samples concordant	Samples discordant	Samples failing to amplify	Comment
rs1976391	Not in PGKB haplotype set	N	68	0	0	
rs4148323	*6	Y	68	0	0	
rs887829	*1, *6, *28, *36, *37	Y	68	0	0	

Table 6.14 Validation of SNPs in UGT1A1

6.2.1.12 *VKORC1*

In the case of *VKORC1*, BBS-003 failed to amplify for rs9934438. In WGS data it was heterozygous. However, as this SNP is in tight linkage disequilibrium with rs9923231, for which BBS-003 amplified and was confirmed to be heterozygous, this can be used as confirmation. Other samples heterozygous for rs9934438 amplified normally. Otherwise there were no disagreements.

rs number	Function	Checked in WGS	Samples concordant	Samples discordant	Samples failing to amplify	Comment
rs9923231	Normal, intermediate, low	Y	68	0	0	
rs9934438	In linkage disequilibrium with above	Y	67	0	1	BBS-003 failed to amplify
rs2359612	Not in PharmGKB haplotype set	N	68	0	0	

Table 6.15 Validation of SNPs in VKORC1

6.2.2 Comparison of SNP and WGS data for additional pharmacogenes

For 12 of the additional 78 pharmacogene assays, a single sample failed to amplify (Table 6.16). In each case another sample with the same genotype amplified normally. In one other assay, rs35694136 (*CYP1A2*), there were five samples that failed to amplify. All but one were ref in WGS.

Gene	rs number	Samples concordant	Samples discordant	Samples failing to amplify	Comment
AARS	rs2070203	68	0	0	
ABCB1	rs2032582	66	2	0	JDM-010 and SRS-14 clustered ambiguously between het and ref. Both het in WGS.
ABCB1	rs1045642	68	0	0	
ABCB11	rs497692	68	0	0	
ABCG2	rs2231142	68	0	0	
ADRA2A	rs1800544	68	0	0	
ADRA2A	rs1800545	67	0	1	SRS-20 failed to amplify. Ref in WGS
ADRB1	rs1801253	68	0	0	
ADRB2	rs1042713	68	0	0	
AGTR1	rs275651	68	0	0	
AGTR1	rs5186	67	0	1	IBD-001 failed to amplify. Ref in WGS
AGTR1	rs5182	68	0	0	
ANKK1	rs1800497	68	0	0	
ATM	rs11212617	68	0	0	
BDKRB1	rs12050217	68	0	0	
CACNA1C	rs1006737	68	0	0	

CACNA1C	rs1051375	68	0	0	
COMT	rs4680	68	0	0	
CYP1A2	rs2069514	68	0	0	
CYP1A2	rs762551	68	0	0	
CYP1A2	rs35694136	63	0	5	BBS-002, BBS-006, IBD-014, USH-001 & JDM-005 failed to amplify. All but one ref in WGS
CYP1A2	rs12720461	68	0	0	
CYP2B6	rs2279343	62	6	0	BBS-002, BBS-012, BBS-014, IBD-003, IBD-005, IBD-006 results clustered between het and ref. All ref in WGS
CYP2B6	rs28399499	68	0	0	
CYP2B6	rs3211371	68	0	0	
CYP2B6	rs35303484	68	0	0	
CYP2B6	rs3745274	67	0	1	BBS-010 failed to amplify. Ref in WGS
CYP2B6	rs8192709	68	0	0	
CYP3A4	rs138105638	68	0	0	
CYP3A4	rs2740574	68	0	0	
CYP3A4	rs35599367	68	0	0	
CYP3A4	rs4646438	68	0	0	
CYP3A4	rs4986910	67	0	1	SRS-011 failed to amplify. Ref in WGS
CYP3A4	rs4987161	68	0	0	
CYP3A4	rs55785340	68	0	0	
CYP3A4	rs55901263	67	0	1	SRS-008 failed to amplify. Ref in WGS
CYP3A4	rs55951658	67	0	1	SRS-018 failed to amplify. Ref in WGS
CYP3A4	rs67666821	68	0	0	
DRD2	rs1799732	68	0	0	
DRD2	rs1799978	68	0	0	
F2	rs1799963	68	0	0	
FKBP5	rs4713916	68	0	0	
GRIK4	rs1954787	68	0	0	
GRIN2B	rs2058878	68	0	0	
GRK5	rs2230345	68	0	0	
HPSE2	rs10883099	68	0	0	
HTR2A	rs6311	68	0	0	
HTR2A	rs7997012	68	0	0	
HTR2A	rs6313	68	0	0	
HTR2C	rs1414334	67	0	1	SRS-006 failed to amplify. Hom in WGS

HTR2C	rs3813929	68	0	0	
ITGB3	rs5918	68	0	0	
KCNJ6	rs2070995	68	0	0	
KIF6	rs20455	68	0	0	
LILRB5	rs12975366	67	0	1	SRS-008 failed to amplify. Het in WGS
LPA	rs10455872	67	0	1	SRS-002 failed to amplify. Ref in WGS
LPA	rs3798220	68	0	0	
MTHFR	rs1801131	68	0	0	
MTHFR	rs1801133	68	0	0	
MTR	rs1805087	67	0	1	IBD-017 failed to amplify. Het in WGS
NOS1AP	rs10494366	68	0	0	
NOS1AP	rs10800397	68	0	0	
NOS1AP	rs10919035	68	0	0	
NOS3	rs1799983			1	JDM-010 failed to amplify. Het in WGS
OPRD1	rs678849	68	0	0	
OPRM1	rs1799971	68	0	0	
OPRM1	rs510769	68	0	0	
POLG	rs113994095	68	0	0	
POLG	rs113994097	68	0	0	
POLG	rs113994098	68	0	0	
POR	rs1057868	68	0	0	
PRSS53	rs7294	68	0	0	
SLC6A4	rs1042173	68	0	0	
SLC6A4	rs25531	68	0	0	
UGT1A4	rs2011425	68	0	0	
VCAN	rs309557	68	0	0	
YEATS4	rs7297610	67	0	1	SRS-020 failed to amplify. Ref in WGS
No gene	rs2952768	68	0	0	

Table 6.16 Validation of SNPs in additional pharmacogenes

In the case of the assay for rs2032582 (*ABCB1*), two samples, JDM-010 and SRS-14, clustered ambiguously between het and ref clusters and could not be called (Figure 6.4). Both were heterozygous in WGS data. The WGS data were rechecked, but no SNPs could be seen that might have affected results. In the case of the assay for rs2279343 (*CYP2B6*), six samples clustered ambiguously and could not be called, although the software called four as heterozygotes and two as unknowns (Figure 6.5). All samples were heterozygous in WGS. The WGS data were rechecked, but no SNPs could be seen that might have affected the results. Of note, most results were from an early run (section 6.3.1.5.4).

All samples in all other assays amplified normally and results were the same as the WGS calls.

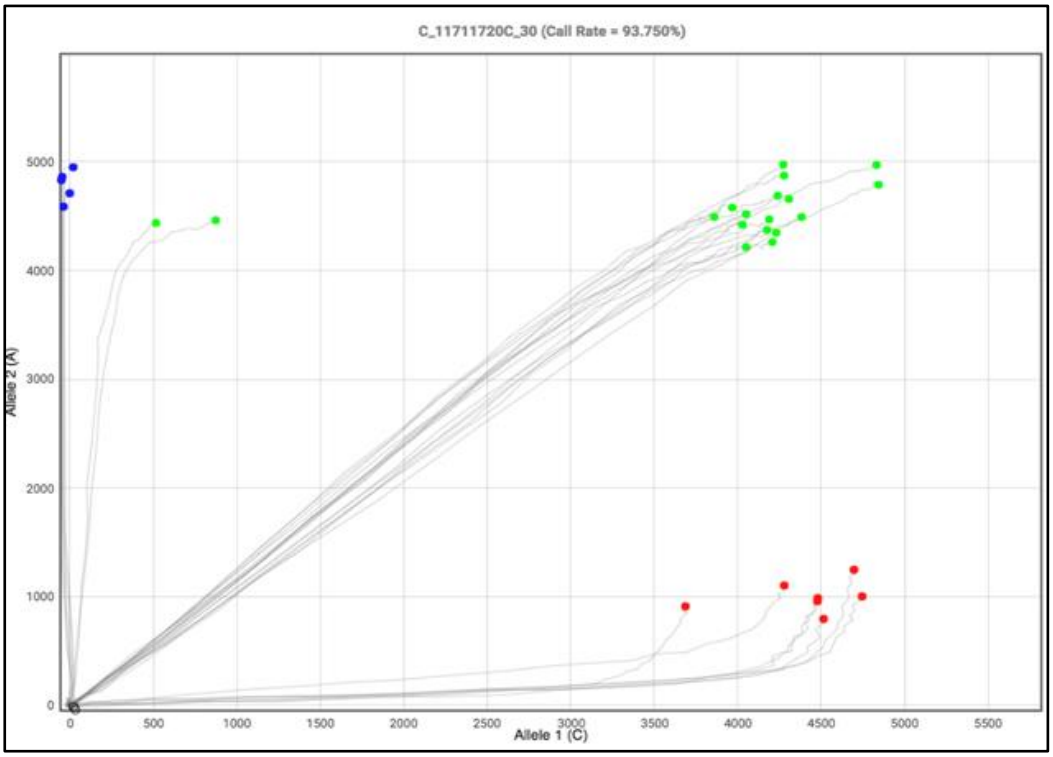

Figure 6.4 ABCB1 rs2032582: clustering of samples JDM-010 and SRS-014 (circled)

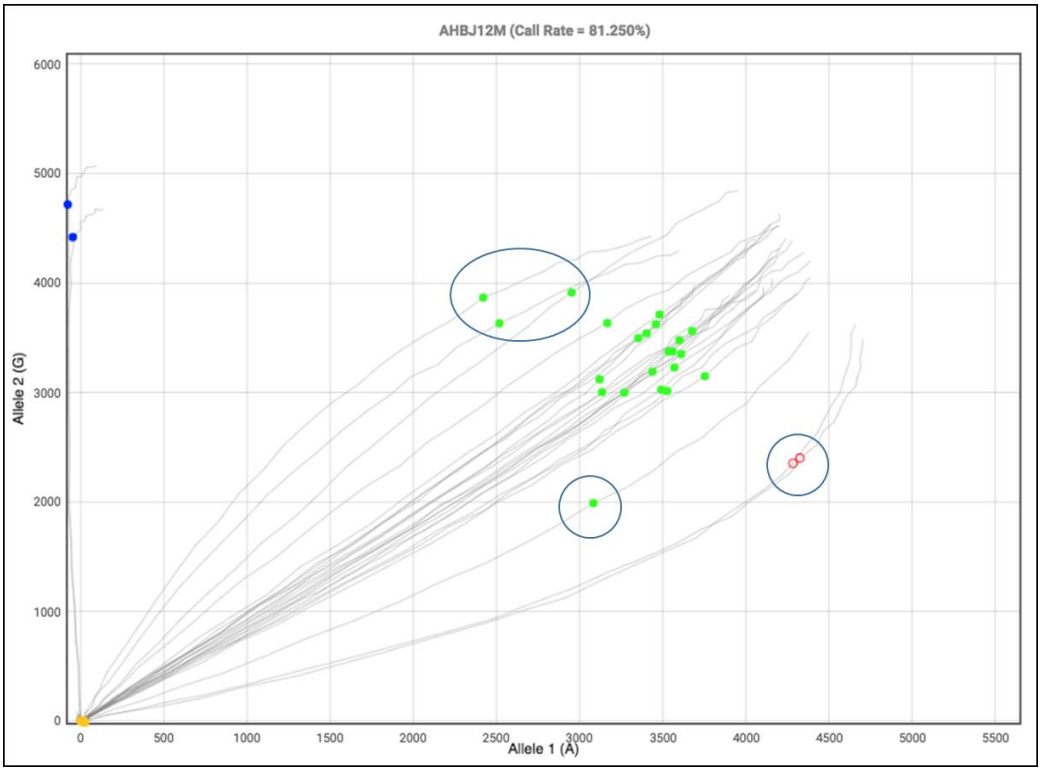

Figure 6.5 CYP2B6 rs2279343: clustering of samples BBS-002, BBS-012, BBS-014 (circled left), IBD-003 (circled centre), IBD-005, IBD-006 (circled right)

6.2.3 Comparison of SNP and WGS data for tracking SNPs

In four of the 24 tracking samples (16%), one or two samples failed to amplify (Table 6.17). Two assays, those for *DMRT2* (Figure 6.6) and *SNTG2* (Figure 6.7), had one or more samples that clustered ambiguously. In the case of two assays, *GRIN3B* and *NKD2*, there was disagreement between SNP and WGS data for a single sample, SRS-012 (Figure 6.8). In each case SNP data called the sample as a homozygote non-reference, but WGS data showed a clear heterozygote. There were no nearby SNPs that would account for the discrepancy in either case. In a final assay, that for *SOX6*, the sample did not amplify normally (Figure 6.9).

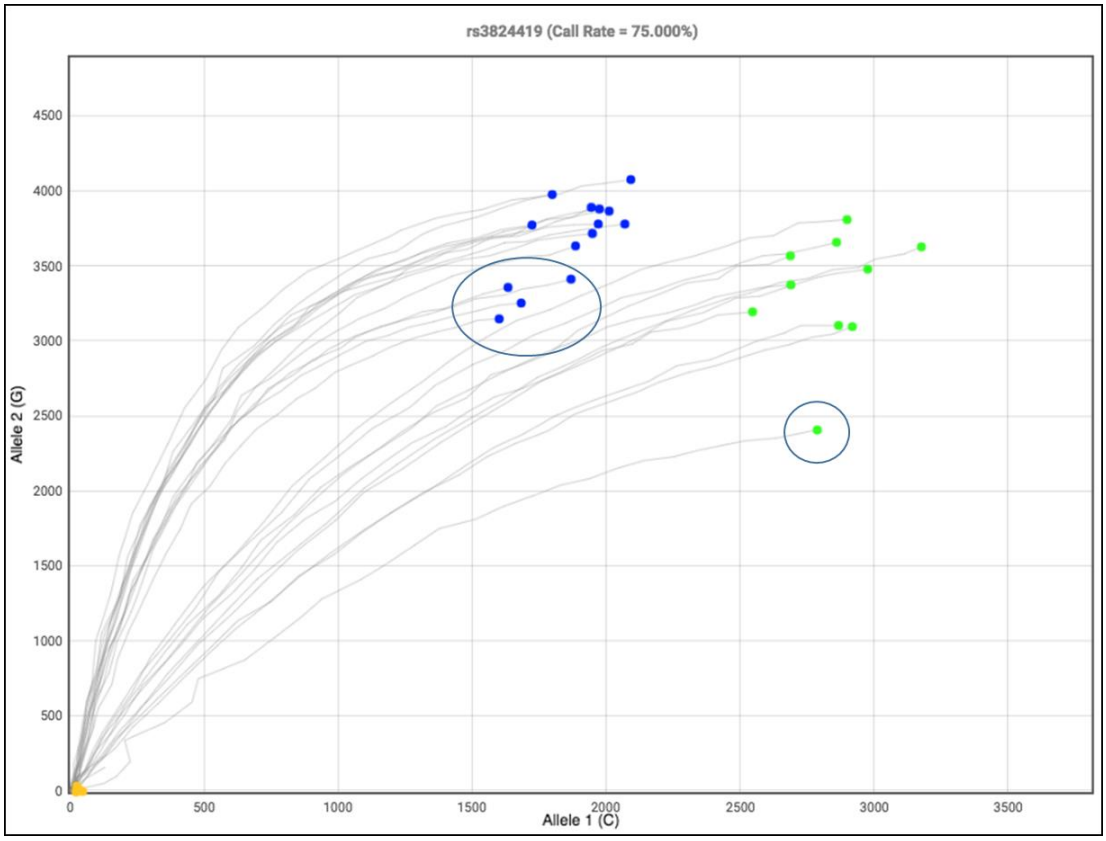

Figure 6.6 DMRT2 rs23824419: clustering of samples BBS-002, BBS-009, BBS-014, IBD-019 (circled left) and IBD-018 (circled right). SRS-021 not included in image

Gene	rs number	Samples concordant	Samples discordant	Samples failing to amplify	Comment
B9D2	rs2241714	66	0	2	JDM-012 and BBS-018 failed to amplify
COG1	rs1037256	68	0	0	
COL4A4	rs10203363	68	0	0	
DMRT2	rs3824419	62	6	0	Ambiguous clustering BBS-002, BBS-009, BBS-014, IBD-007, IBD-018, SRS-021
DNMT1	rs2228611	68	0	0	
EVC	rs4688963	67	0	1	SRS-020 failed to amplify
FERMT1	rs10373	68	0	0	
FREM2	rs9532292	68	0	0	
GRIN3B	rs4807399	67	1	0	SRS-012 had shorter trajectory, called as hom, het in WGS data, but low concentration sample
L2HGDH	rs2297995	68	0	0	
LAMA3	rs9962023	68	0	0	
MUC6	rs7481521	68	0	0	
NDUFV3	rs4148973	68	0	0	
NKD2	rs60180971	68	0	0	SRS-012 had shorter trajectory, called as hom, het in WGS data, but low concentration sample
NPHS2	rs1410592	67	1	0	
PDP1	rs4735258	68	0	0	
PLCG1	rs753381	68	0	0	
SLC12A6	rs4577050	68	0	0	
SNTG2	rs4971432	66	1	1	SRS-020 failed to amplify. IBD-008 clustered ambiguously between het and hom. Het in WGS
SOX6	rs4617548	67	1	0	SRS-020 poor amplification. Het in WGS
SUMF1	rs2819561	68	0	0	
TDRD7	rs1381532	68	0	0	
TNFRSF4	rs17568	68	0	0	
WNK1	rs7300444	67	0	1	SRS-020 failed to amplify.

Table 6.17 Validation of SNPs in tracking SNP

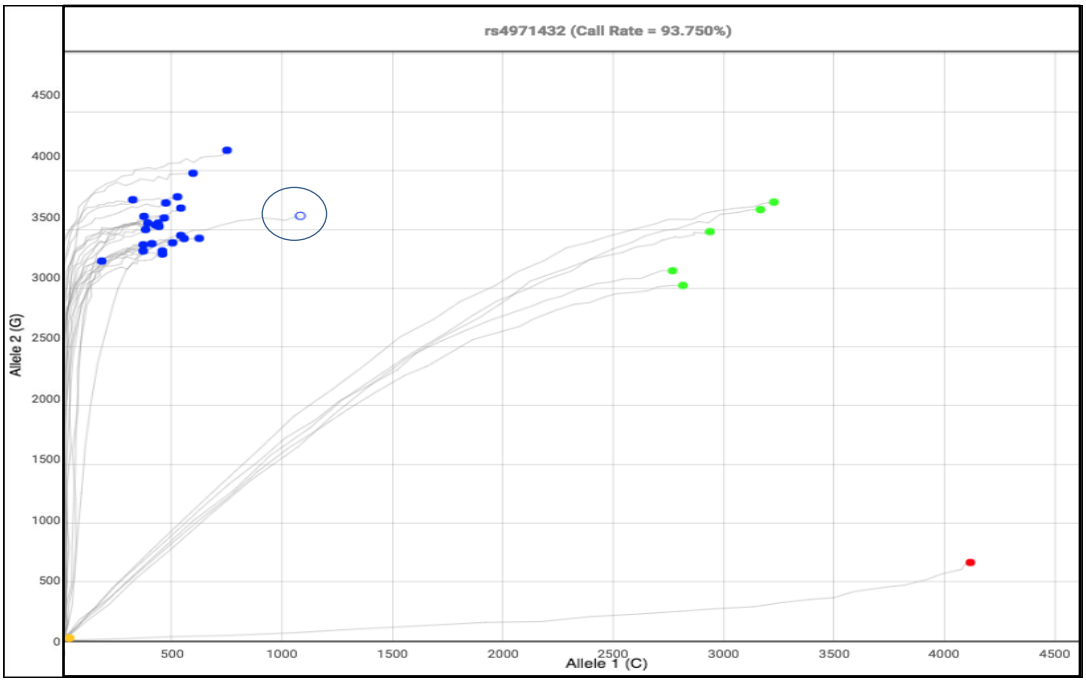

Figure 6.7 SNTG2 rs4971432: clustering of sample IBD-008 (circled) between wild type and heterozygotes

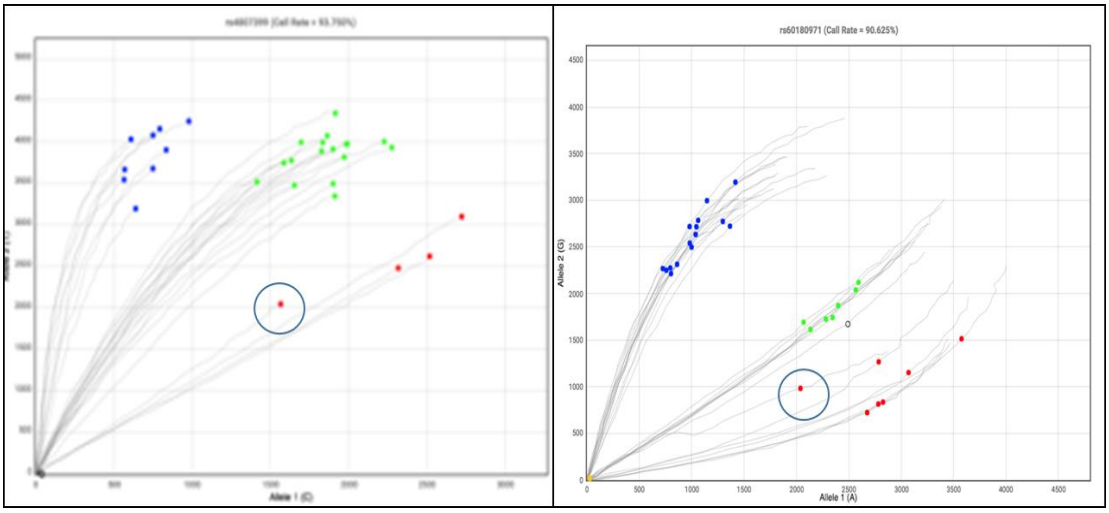

Figure 6.8 GRIN3B rs4807399 and NKD2 rs60180971: clustering of sample SRS-012 (circled): shortened trajectory and calling as non-reference homozygote for GRIN3B rs4807399 (left) and NKD2 rs60180971 (right)

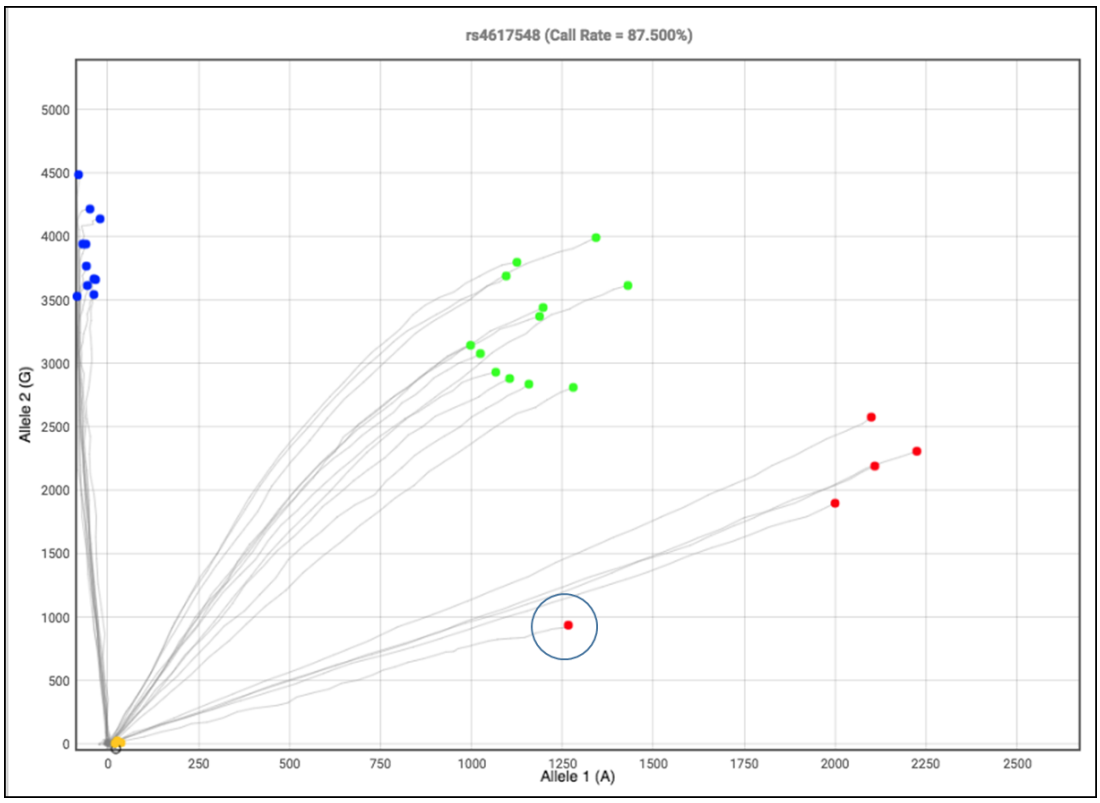

Figure 6.9 SOX6 rs4617548: clustering of sample SRS-020: shortened trajectory and calling as non-reference homozygote (circled)

6.2.4 Comparison of copy number variants

Copy number variants could only be validated for IBD-007 (duplication of *CYP2D6*) and IBD-013 (deletion of *CYP2D6*) and this was done by Congenica Ltd. The copy number duplication detected by Astrolabe in IBD-007 and the copy number deletion in IBD-013 were confirmed (Figure 6.10).

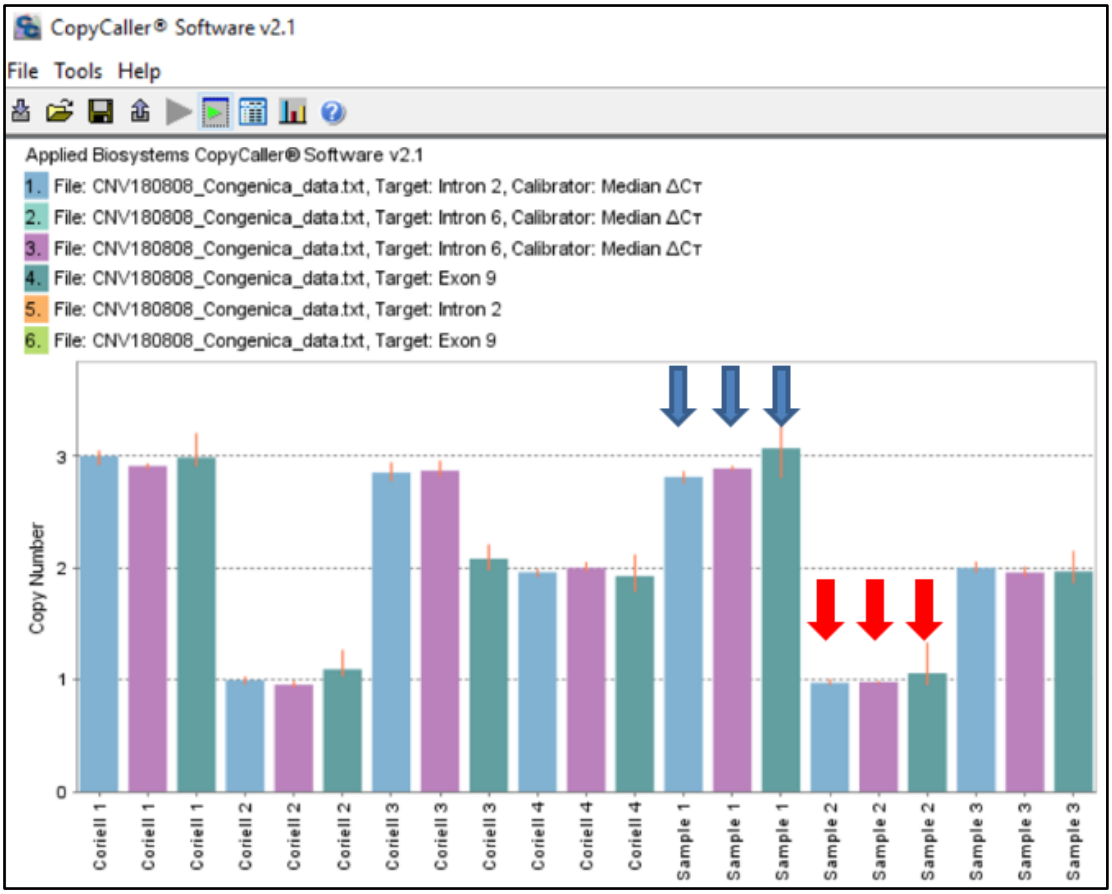

Figure 6.10 Copy number variants validation for IBD-007 (blue arrows, copy number gain) and IBD-005 (red arrows, copy number loss)shown in CopyCaller software. Image courtesy of Congenica Ltd

6.3 Discussion

6.3.1 Performance of whole genome sequencing compared to SNP genotyping

6.3.1.1 Pharmacogenes with CPIC or DPWG prescribing guidelines

6.3.1.1.1 SNPs in pharmacogenes with CPIC or DPWG prescribing guidelines

Overall there were 75 individual assays for SNPs pertaining to the actionable pharmacogenes discussed in Chapter 5. Each one was run for each of 68 samples, making a total of 5100 individual assays.

In total, there were 18 failures to amplify (<0.5% of assays run). This consisted of 12 assays where a single sample failed and three assays where two samples failed. Possible reasons for this are discussed in section 6.3.1.4.3.

There were five issues when comparing WGS and SNP data affecting < 0.1% of assays run. Two were in *HLA-A* rs1061235 where samples JDM-007 and SRS-017 did not cluster with other heterozygotes, neither did they cluster with the single homozygote (Figure 6.1). This may be because there was only a single homozygote for rs1061235 in the entire cohort. The variant calling software operates by comparing amplification curves between samples, so the fewer of one particular type is seen, the more difficult the software finds it to call the sample. When the WGS data were re-examined, additional SNPs were seen in the surrounding 20 base pairs. This may have affected sample amplification and is the likeliest reason for the discrepancy. This could be resolved by the use of Sanger sequencing to determine the sequence at this nucleotide.

A third disagreement was in sample IBD-007 for *SLCO1B1* rs4149056, where the sample clustered as a heterozygote (Figure 6.2) but had been originally called as ref. When the WGS data were re-examined, it was seen that is was a heterozygote, with 6/22 reads (27%) being non-reference (Figure 6.3). This was missed originally. It is unclear whether the individual is mosaic for the non-reference allele or that due to the relatively low read depth, the ratio was not 50:50. Sequencing at a greater read-depth might help to resolve this.

The remaining two issues were with *HLA-B*, where IBD-012 was heterozygous for both SNPs tested and JDM-001 was heterozygous for one and homozygous for the other. While this agreed with WGS data, none of the samples were called as *HLA-B* positive based on the original SNPs chosen. As discussed in sections 5.3.2.3.4 and 6.1.2, the SNPs originally selected were not included in the Congenica assay but there were two alternative *HLA-B* SNPs. Selection of SNPs for *HLA-B*15:02* is difficult, and there is little agreement as to which SNPs are best. There appears to be ethnic variation also. When comparing the Congenica SNPs for these samples with the SNPs originally examined in WGS data, neither of these individuals has any positive SNPs for *HLA-B* *15:02 in the WGS data, while other individuals with negative SNPs in Congenica data have a single positive SNP in WGS data. When the ethnicities of the patients were considered, patient IBD-012 is described as white European, making HLA-B *15:02 positivity unlikely for her. Patient JDM-001's ethnicity is described as other, so it remains possible that they are from an ethnic group where *HLA-B* *15:02 is seen. It may be that SNP testing is not a good method of

testing for *HLA-B* *15:02, and this is certainly the case for other *HLA-B* haplotypes such as *HLA-B* *44. Sanger sequencing would not help with resolution of this issue. Serological testing of HLA alleles is done routinely as part of a bone marrow transplant workup and remains the gold standard(793).

Of note, as mentioned in section 6.2.1.9, an additional SNP, rs4149015, was used in the Congenica panel compared to in the original WGS analysis of *SLCO1B1*. This SNP distinguishes the *17 and *21 haplotypes. The function of the *21 haplotype is undetermined and no prescribing advice exists for it. Several samples were identified that were positive for rs4149015, but negative for rs4149056, identifying them as having the *1/*21 diplotype. The samples in question are JDM-004, SRS-002 and SRS-018. The population frequency of the *21 haplotype is unknown. From these data, the frequency in the cohort is 4%. However, it does not change prescribing advice for any of these individuals, and they would be given the same prescribing advice as individuals with the *1/*1 haplotype, as they were originally.

6.3.1.1.2 CNVs in pharmacogenes with CPIC or DPWG prescribing guidelines

The sole deletion and one of the duplications detected by Astrolabe were confirmed. Lack of DNA meant the others could not be confirmed. These results suggest that Astrolabe can detect CNVs. However, in several samples, Astrolabe detected a duplication that did not seem to be present when WGS data were reviewed. As no DNA remained, these samples could not have their copy number status clarified, leaving doubt over the ability of Astrolabe to confirm CNVs.

6.3.1.2 Additional pharmacogenes without CPIC or DPWG prescribing guidelines

Overall there were 78 individual assays for SNPs pertaining to the pharmacogenes without CPIC or DPWG guidelines, that is, those pharmacogenes not originally considered when examining the WGS data. Each one was run for each of 68 samples, making a total of 5304 individual assays.

In total, there were 17 failures to amplify (< 0.5% of assays run). This consisted of 12 assays where a single sample failed and one assay where five samples failed. Possible reasons for amplification failure are discussed in section 6.3.1.4.3.

There were 8 disagreements (< 0.15% of total assays run) between SNP and WGS data, although these were all found in just 2 different SNPs (2.5% of SNPs).

The first of these was *ABCB1*, rs2032582, where JDM-010 and SRS-14 clustered ambiguously between het and ref. In both cases, the samples were clearly ref in WGS, and there were no SNPs in the surrounding areas to explain this discrepancy, although SNPs that might have affected primer binding have not been excluded. Therefore, the genotypes at this SNP remain undetermined. This could be resolved by additional testing, for example by Sanger sequencing of this variant.

The second issue was with *CYP2B6*, rs2279343, where multiple samples clustered ambiguously between ref and het, meaning that they could not be called. This affected samples BBS-002, BBS-

012, BBS-014, IBD-003, IBD-005 and IBD-006. All the samples were ref in WGS data. No SNPs were identified in the area immediately surrounding this nucleotide, but more distant SNPs have not been excluded. All of these samples were run together, although no quality control issue was identified that might explain it. Sanger sequencing would confirm the correct genotype.

6.3.1.3 Tracking SNPs

While not relevant to whether pharmacogenomic data can be extracted successfully from WGS data, issues with tracking SNPs have been included for comparison. There were 24 individual tracking SNP assays. Each was run for 68 samples, making a total of 1632 individual assays.

In total, there were six failures to amplify (< 0.5% of assays run). This consisted of 4 assays where a single sample failed and one assay where two samples failed. Possible reasons for amplification failure are discussed in section 6.3.1.4.3. There were five assays where there were disagreements between DNP and WGS data affecting ten samples (< 1% of assays run).

The first of these was *DMRT2*, rs3824419, where multiple samples, BBS-002, BBS-009, BBS-014, IBD-007, IBD-018 and SRS-021, clustered between het and ref, while all were ref in WGS. The same issue was seen for SRS-012 for *GRIN3B* rs4807399 and *NPHS2* rs1410592. No SNPs were identified in the surrounding nucleotides and no issue was identified in QC data, although SNPs further upstream or downstream have not been ruled out as a cause. Sanger sequencing would resolve this issue. The same is true of *SNTG2* rs4971432, in sample IBD-008, although in this case it was called as het in WGS.

The final disagreement was in *SOX6* rs4617548, where the sample appeared to be following the homozygote trajectory, but did not cluster with the other homozygote samples. It was a heterozygote in WGS data. This appears to be an amplification problem, and SRS-020 is a sample that failed to amplify on more than one occasion (section 6.3.1.4.3). Resolution of this discrepancy would require Sanger sequencing.

6.3.1.4 Overall performance for all pharmacogenes

6.3.1.4.1 Clustering failures

In the vast majority of samples, WGS sequence data were confirmed by SNP data. In those cases where there was disagreement, it was generally when the SNP data did not cluster sufficiently well to make a call, meaning that the SNP data would have been ambiguous regardless of WGS data. This was the case for the *HLA-A*, *ABCB1*, *ATM* and *CYP2B6* SNPs described in sections 6.3.1.1 and 6.3.1.2. With the exception of *HLA-A* no causes for ambiguous clustering were found, and Sanger sequencing would be the best way of resolving this. However, DNA shortage meant that this was not an option. Issues with clustering are further discussed in section 6.3.1.5.4. An additional issue was the fact that several samples were called as *HLA-B* *15:02 positive by the Congenica Ltd assay but not by the WGS assay where different SNPs were used. A summary can be seen in Table 6.18. When tracking SNPs were excluded and clustering issues ignored, the overall disagreement rate for pharmacogene assays was < 0.01.

Type of SNP	No. of SNPs per sample	Total no. of assays	No. of concordant assays	No. of assays not amplified	No. of assays with clustering issues	No. of assays truly discordant	% truly discordant
Actionable pharmaco-genes	75	5100	5079	18	2	1	**0.02**
Other pharmaco-genes	78	5304	5279	17	8	0	**0**
Tracking SNPS	24	1632	1617	6	7	2	**0.1**
Total	177	12036	11975	41	17	3	**0.02**
Total pharmaco-genes	153	10404	10358	35	10	1	**<0.01**

Table 6.18 Comparison of WGS data with SNP data, showing percentage of true disagreements

6.3.1.4.2 Disagreements between WGS and SNP data

Only in *HLA-B* and *SLCO1B1* were there actual disagreements. In *HLA-B* it appears that using SNP data is not a good method of identifying *HLA-B* *15:02 positive individuals at present. Until SNPs are validated in large numbers of individuals of different ethnicities whose *HLA-B* haplotypes have been confirmed serologically, this will continue to be the case. With *SLCO1B1*, it appears that the SNP data identified an individual who is positive, or perhaps mosaic, for the rs4149056 variant. This was originally missed in WGS data. If this individual were a mosaic for this variant it is unclear what the clinical effect of this would be.

6.3.1.4.3 Failures of amplification

The failures of amplification in general did not affect confirmation of WGS data in the actionable pharmacogene because, in almost every case, it was not the SNP seen in WGS that failed to amplify. However, the possibility remains that other, rarer haplotypes might be missed in this way. There were only two cases where an important SNP failed to amplify. In the first the haplotype could be confirmed without the SNP (and was confirmed in an identical twin) and in the second, a SNP in very tight linkage disequilibrium was also included in the panel and so the genotype could be confirmed. There were more issues for the non-actionable pharmacogenes, as many had only a single SNP included.

Failures of amplification have several possible underlying causes. Possibilities include primer non-binding, for example, due to a SNP underlying the binding site, poor loading of the plate so that DNA did not enter that particular assay well, a problem with the plate such as poor spotting of the probe onto the plate or a homozygous deletion of that particular section of DNA, which is unlikely if the area is well covered by WGS.

Interestingly while most samples had one or no failed amplifications overall, several samples had failed amplifications over more than one assay. This is shown in Table 6.19. However, all of these samples amplified well in other assays. A possible explanation is plate loading, which is operator dependent. While the spotting of the samples on the plate is done by machine, the plate then has to be manually filled with oil and sealed. This must be done very rapidly as samples will evaporate until sealing is completed. The introduction of bubbles when the plate is being filled with oil affects how the assays perform, with areas under a bubble not amplifying correctly. Therefore, if a small bubble is present, one or more of the assays for a particular sample may not work. However, in the case of all SRS samples and some other samples, small volumes of DNA remained for performing this experiment. Dilution was required to increase the volume to the minimum required for the assay. While the concentrations of the samples remained above or at the minimum recommended for the assay, it is possible that they were insufficiently mixed and that the concentration was not even throughout the sample. In addition, the SRS samples were older than the other samples in the study.

Sample	Number of failures	Possible reason
IBD-014	2	No reason identified. Possible loading error
IBD-017	2	Very small volume of DNA available for assay. Required dilution. Possible insufficient mixing or loading error
SRS-006	2	Very small volume of DNA available for assay. Required dilution. Possible insufficient mixing or loading error
SRS-008	5	Very small volume of DNA available for assay. Required dilution. Possible insufficient mixing or loading error
SRS-020	5	Very small volume of DNA available for assay. Required dilution. Possible low concentration, insufficient mixing or loading error

Table 6.19 Samples with failures in more than one assay

6.3.1.5 Issues identified

Overall, the WGS data were in almost complete agreement with the SNP data, making WGS a feasible method for the extraction of pharmacogenomic data. This is in agreement with previously published data and the recent large study by Reisberg(636, 794). However, some issues were identified in this experiment.

6.3.1.5.1 Confirming copy number variants

The use of a SNP-based assay did not allow for confirmation of copy number variants in *CYP2D6* (section 5.2.4.4). This is a limitation of SNP-based testing, and other assays exist for the confirmation of CNVs(793, 795, 796). A qPCR-based assay, the TaqMan® *CYP2D6* CNV assay, was used to confirm two of the identified CNVs. Unfortunately, in this cohort, shortage of DNA meant that this confirmation could not be performed for the remaining samples resulting in the copy number variants identified by Astrolabe remaining unconfirmed. The use of long read WGS

is likely to mean that extracting copy number information from WGS data will be more straightforward in future(797).

6.3.1.5.2 Clarification of disagreements

Again, due to a lack of DNA, disagreements between SNP and WGS data could not be clarified and neither could SNP samples that clustered ambiguously. Ideally, these discrepancies could be resolved by Sanger sequencing of the variants in question.

6.3.1.5.3 Choice of SNPs

In the initial WGS analysis, the analysed SNPs were restricted to those that would allow alteration in prescribing. However, this meant that rare SNPs were missed, as was shown by Astrolabe (see section 5.2.4). However, Astrolabe also called haplotypes that did not appear to be present, apparently calling them on the basis of one of a number of SNPs or minor SNPs when the major SNP for that haplotype was not present. This is a limitation of computational calling. However, in order to analyse large numbers of SNPs, computational support is necessary. Analysis was done manually in this project, but that meant that it was extremely difficult to expand the number of SNPs looked at (section 5.3.1.2). The restricted SNP set is also a limitation of SNP genotyping, but unlike WGS, there is no option to check additional SNPs or to expand the set of SNPs examined. For example, in the case of *UGT1A1*, the inability to interrogate rs8175347 (a triplet repeat variant) in the SNP genotyping data meant that samples could be divided only into increased and decreased function, rather than calling exact diplotypes. While this is certainly good enough to allow use of current prescribing guidelines, it may not be if guidelines are refined to take account of the exact diplotype seen. It is likely that as pharmacogenomics knowledge increases, particularly in relation to exactly which SNPs are required for a haplotype to be present in a given ethnicity, the most accurate way to obtain pharmacogenomic data will be to use WGS and interrogate it with automated bioinformatics methods.

6.3.1.5.4 Resolving issues with clustering

In the Congenica Ltd custom pharmacogenomics assay samples were called on the basis of the detection of two fluorescent dyes. In heterozygotes, approximately equal amounts of both dyes should be detected, while in reference or non-reference homozygotes only one dye should be detected (Chapter 2). In reality, small amounts of the other dye are detectable even in homozygote samples and the software calls samples in part by looking at how they behave with respect to other samples that have been tested. This means that the greater the number of samples run on a particular machine the more the calling improves. In the earliest runs, many samples were ambiguous, but as more samples were run, the software's ability to call samples improved and many of the early clustering issues were resolved. However, in assays where few reference or non-reference homozygotes were seen, the software appeared to have difficulty in telling the difference between homozygotes (either reference or non-reference) and heterozygotes that did not amplify exactly like the other heterozygotes.

It is possible that the observed clustering issues will resolve as more assays are performed using the platform. They might also be improved by the purchase of commercial control samples, but

this is expensive, both in terms of the initial expenditure, and that fact that a control then replaces a normal sample, albeit in a single run. While it is possible to manufacture oligonucleotides that contain the reference and control sequences for all assays that are then ligated together and can be run as a single sample, this is expensive and caution would be needed to avoid contamination of other samples, equipment and machinery.

6.3.1.5.5 Poor coverage in WGS data

While not an issue for the pharmacogenes looked at in the original WGS analysis, poor coverage (defined as a read depth of less than 10x in WGS data) was an issue for some SNPs in the additional pharmacogenes, namely the SNPs for *ADR2A* (rs1800545), *ADRB*1, *DRD2* (rs1799732), *OPRM1* (rs10769) and *SLC6A4* (rs25531). This was particularly the case for *ADRAB1*, *DRD2* and *SLC6A4*, where up to 25% of samples had a read depth of less than ten. While this did not appear to affect agreement between SNP and WGS data and there were no discrepancies noted in any of them, poor coverage might mean that these SNPs would be difficult to call in WGS as they would not meet the pre-set quality control standards. Recent EMA Good Pharmacogenomics guidance has suggested a that a read depth of 30x is desirable(798).

6.3.1.5.6 Prescribing advice

Using a SNP based assay allowed confirmation of SNPs which were manually converted to haplotypes and diplotypes as was done with WGS data. Therefore, the diplotypes called from the WGS sequence were not directly confirmed and neither was the prescribing advice that was manually collated. A solution would be to use a commercial company that converts SNP data to prescribing data, for example Translational Software Ltd (www.translationalsoftware.com).

6.3.1.5.7 Phasing

As discussed in section 5.3.2.2.1, it is not possible to detect phase using current short read WGS data. Most SNP based testing cannot detect phase either, so there is the possibility that variants that are assumed to be in *trans* are in *cis*, and vice versa. This was not possible to resolve in this study and it is an issue with most commercial pharmacogenomic tests. As in the case of CNVs, long-read WGS technologies have the potential to resolve this(363, 608).

6.3.1.6 Limitations and improvements

There were a number of limitations to this study. Firstly, not all samples had sufficient DNA remaining for confirmation. Had more been available, it would have allowed confirmation of copy number variants. It also would have been possible to rerun any assays where there were issues with clustering and to confirm any disagreements using Sanger sequencing.

Secondly, a relatively limited SNP set was examined initially in WGS data. Results from Astrolabe and the extended set of SNPs for some genes in the Congenica Ltd assay suggest that looking at a larger number of SNPs might be beneficial. However, calling SNPs manually was time consuming, and also allowed the possibility of operator error. Automating this step, possibly by building a list of SNPs into software such as Qiagen Variant Analysis™, Congenica Sapientia™ or Fabric Genomics Omicia™, would allow this to be done for larger cohorts in a shorter time. It

would also be possible to write an algorithm that would call SNPs and haplotypes from a vcf file. This would be particularly beneficial for genes with large numbers of haplotypes such as *CYP2D6* and was done recently by Reisberg et al.*(636).* However, as was seen in the Astrolabe data, automated calling has pitfalls, such as when a haplotype was imputed from the presence of minor SNPs only, without the major or "tag" SNP for a haplotype being present. For this reason, using a relatively small SNP set and checking manually is good for validation purposes, even though it does not exclude the possibility of missing rare haplotypes.

Thirdly, it was not possible to determine whether individuals were positive for *HLA-B* *44 or *HLA-B* *58:01 from WGS sequence by manual comparison. This was mainly because there is extremely limited data as to what constitutes these haplotypes and which SNPs are allowable and which are sufficient to call a sample as *44 or *58:01. Although this was attempted, it became clear that further work would be required to determine the consensus sequences to be used, especially in mixed populations. Clinical testing of *HLA* haplotypes is also possible.

The use of long-read sequencing data would allow calling of phase and also increase the ability to detect copy number variants, which would be an improvement on current methods. This would be the preferred method going forward, especially as costs come down and the technology is more readily available.

An aim is to bring a study such as this into the clinical setting and see how it could influence prescribing and ascertain the most beneficial time and setting of testing. This is discussed in section 6.3.4.

6.3.2 Advantages and disadvantages of WGS compared to other methods

As shown in section 6.3.1, WGS performed well when compared to SNP genotyping, the method most commonly used in commercial pharmacogenomic tests. However, both methods have strengths and weaknesses and may therefore be appropriate for different clinical circumstances.

6.3.2.1 Advantages of WGS

6.3.2.1.1 Potential to review data

WGS has a major advantage over any current commercially available tests in allowing the data to be accessible for review. As WGS sequences all coding and non-coding areas of the genome, further pharmacogenomic data can be extracted at any point. This might be important when pharmacogenomic guidelines change to include additional pharmacogenes or haplotypes, when novel pharmacogenes are discovered or when additional information about haplotypes becomes available, for example in understanding the clinical consequence of a rare haplotype or the discovery of a new haplotype. This is likely to be important as we learn more about pharmacogenomics and in particular about modifiers. Data from SNP assays are fixed and cannot be reviewed to add additional genes or haplotypes, meaning that the test might have to be modified and repeated for up-to-date prescribing information. This may be even more relevant to individuals from minority ethnic groups, who may not be well served by current tests. Much of the

data on which we base our current genomic knowledge do not accurately reflect human diversity(367).

6.3.2.1.2 Prospective data

As data can be reviewed and revised at any time, WGS-based testing is ideal for prospective data use, meaning that testing can be initiated at any point and the data interrogated when relevant. While this would require computational solutions and linked prescribing, which is discussed in section 6.3.4.3, it means that data are available at the point of prescribing and eliminates the wait for test results. This is currently a limitation of pharmacogenomic prescribing, as physicians may be reluctant to delay prescribing while awaiting results. Although this could be alleviated by point of care testing, this is not cheap, nor is it available for all pharmacogenes. Rapid initiation of treatment may be particularly important for chemotherapeutic agents. The earlier in life that testing is done, the more prescribing episodes the data will be available for. Prospective testing has been implemented in St Jude's Children's Hospital as part of the PG4KDS study (https://www.stjude.org/research/clinical-trials/pg4kds-pharmaceutical-science.html).

6.3.2.1.3 Increased accuracy

In addition to the ability to cover all pharmacogenes and haplotypes, the introduction of long read sequencing technology will allow a single test to look at copy number variants in addition to SNPs. Also unlike SNP-based testing, it will allow the identification of phase for more accurate haplotyping, and will reduce the possibility that pseudogenes are being sequenced or genotyped instead of the pharmacogene, thereby increasing the accuracy and reliability of results.

6.3.2.1.4 Adding value to WGS sequencing

While the cost of WGS is reducing all the time, it is still an expensive test compared to WES, single gene tests, or SNP-based PGx testing. However, more people are having WGS sequencing done for diagnostic reasons, and WGS will be an important test in the new NHS test directory when it is introduced in 2019. Extracting pharmacogenomic data will add value to WGS. In particular, patients who are eligible for WGS are likely to have complex disease, and therefore may stand to benefit from pharmacogenomic testing at a younger age. It is likely that at some point WGS will become the test method of choice for all genetic testing, especially when costs reduce to a point where it is not much more expensive to perform WGS testing than to do a single gene test. Being able to obtain pharmacogenomic data for little extra cost would be a benefit of this. WES has additional limitations in that a proportion of pharmacogenomic SNPs are intronic and so would not be covered by WES, limiting the utility of WES for pharmacogenomic testing(636).

6.3.2.1.5 Gene or haplotype discovery

As with gene discovery in diagnostics, WGS can be used in pharmacogenomic gene or haplotype discovery. However, caution needs to be exercised in the prediction of functionality from variants, and may need phenotypic testing, such as enzyme assays, to back it up(798).

6.3.2.2 Disadvantages of WGS

6.3.2.2.1 Cost

Currently, WGS is more expensive than other pharmacogenomic testing methods, and unless it is being done for another reason, such as diagnostics, it would not be cost effective to carry out pharmacogenomic testing by this method.

6.3.2.2.2 Requirement to review data

While, as discussed in section 6.3.2.1.1, the potential to review data is a huge advantage, it may also bring with it responsibilities, such as ensuring that patients have access to the most up-to-date guidance, not just that which was available at the point when WGS was done. If WGS were to be routinely used for pharmacogenomics, it is probable that there would need to be a robust method for data review and for allowing patients and their doctors to access this. This obligation is less likely to exist with a SNP-based test, where the number of pharmacogenes and haplotypes checked cannot change, although there might still be implications should prescribing guidance change.

6.3.2.2.3 Secondary findings and variants of unknown significance

The utilisation of WGS in the clinical setting raises the possibility of identifying both variants of unknown significance that cannot be interpreted and also of secondary findings. Certain secondary findings are a particular risk in the pharmacogenomics setting. As discussed in Chapter 5, if a virtual panel of pharmacogenes were to be drawn up, it might include *CFTR,* which, while not relevant to anyone without a diagnosis of cystic fibrosis (CF), would constitute a carrier test in individuals without CF should a pathogenic variant be identified. Also relevant is *G6PD.* Should heterozygous variants be identified in a male individual, a diagnosis of Kleinfelter syndrome or other sex chromosome abnormality would be possible, although there are other explanations such as mosaicism. A further example is the *BRCA* gene, where patients may be more responsive to alternative chemotherapeutic agents such as PARP inhibitors if they carry a pathogenic *BRCA* variant. A decision would be required in advance regarding which genes would be looked at and which, if any, incidental findings would be fed back to patients. Patients must be informed of this and have the ability to opt into, or out of, secondary findings.

6.3.2.2.4 Ability to interpret WGS data

Currently the ability to extract pharmacogenomic data from WGS sequence is far in advance of the ability to interpret it. It has been shown that rare variants, some of which will be clinically important, will be identified in a majority of patients(799). Determining clinical relevance and how to alter prescribing to take account of them will improve but at present, particularly in individuals from minority ethnic populations, we cannot always interpret the data we find. This means that the current approach is to utilise WGS data in exactly the manner of a SNP-based test, looking only at SNPs we can interpret confidently, and therefore providing no benefit over current, cheaper testing methods(800).

6.3.3 Benefits, limitations and ethics of pharmacogenomic testing

6.3.3.1 Benefits of pharmacogenomics

6.3.3.1.1 Patient safety and drug efficacy

Pharmacogenomics is often described as giving the ability to give the right dose of the right drug to the right patient at the right time, and this has been a fundamental aim of medical professionals from long before pharmacogenomics was a concept. However, the study of pharmacogenomics has helped to refine the risk to patients and allow safer prescribing based on genetic background. The prescribing guidelines discussed in Chapter 5 can be divided into those affecting drug safety and those affecting drug efficacy, although of course, these may go together, especially if compliance is affected by adverse effects. Most of the current guidelines relate to toxicity rather than efficacy. *HLA-A* *31:01 and *HLA-B* *15:02 haplotypes and the development of severe cutaneous adverse effects is an excellent example of how pharmacogenomics can influence patient safety, where anti-epileptics such as carbamazepine, while working perfectly to control seizures, can cause life-threatening adverse effects(555, 557). Pre-emptive testing has been shown to reduce the risk of adverse outcomes(801). Similarly, codeine is not efficacious for poor metabolisers, but ultra-rapid metabolisers are at risk of overdose(526, 527, 645, 802). In the case of *HLA-B* *44 and ribavirin, the pharmacogenomic background mainly affects efficacy, increasing or decreasing the likelihood of viral clearance.

The benefits may not always be clear. For example the EU-PACT and US-based COAG trials looked at the benefits of pharmacogenomic testing for warfarin prescribing. While the former showed a benefit, the latter did not(803, 804). Possible explanations include the more racially diverse population of the COAG trial and differences in trial methodology(805). A more recent study, the GIFT trial, has confirmed the findings of the EU-PACT study, that genotype guided prescribing is beneficial(806).

Recently, there has been some research into patient perception of pharmacogenomics. In general, patients respond positively to pharmacogenomics in theory, and patient satisfaction appears to be higher when pharmacogenomic data are used(113, 807-809). However, patients did have concerns about insurance implications and the possibility of medical discrimination, highlighting the importance of the consent process where these concerns can be discussed and addressed, as is the case with all genetic testing.

6.3.3.1.2 Cost benefit

There is a potential economic argument for the introduction of pharmacogenomic testing due to the high cost of adverse drug reactions. The percentage of hospital admissions caused by ADEs in Europe has been estimated at between 2-10%(810). One US study put the cost at $666,159,537 in the year 2007 in the US for patients over the age of 65 only(811). Genetic drug reactions are often considered unpreventable adverse events, but may not be if the genotype is known in advance. A 2004 study by Pirmohamed et al. estimates that 6.5% of emergency admissions are related to ADEs and that each ADE event resulted in eight bed days, at a total yearly cost of £466,000,000(599). The cost of an NHS bed day is estimated at £500-£600 for a

short stay emergency admission, but a paediatric ICU bed can cost up to £4500 per day(812). A 2001 study by Impicciatore et al. found that 9.3% of paediatric in-patients have ADEs, approximately 12% of them severe. The rate of hospital admission for ADEs was 2.09% and 39.3% of these were considered life-threatening(813). Furthermore, recording of ADEs is not consistent, raising the possibility of patients being exposed to the same or a related drug again(412, 814).

However, due to the current high cost of testing and the fact that variants causing adverse outcomes are generally rare, it has been difficult consistently to show an economic benefit. This may also be due to the difficult in estimating total costs of adverse drug outcomes, which may include loss of earnings or productivity. A 2016 review by Berm et al. showed that many studies found an economic benefit, but interestingly, these studies were more likely to be funded by pharmaceutical companies than those showing no benefit(815). Another review also found that the majority of studies showed economic benefit, with the cost of testing being a major factor(816). This is an argument for the extraction of data from existing WGS data as it reduces the cost of the testing. Proof that single gene single drug studies are cost effective may not be enough to drive implementation of testing, especially if testing is not available at the point of care with a rapid turnaround time(817-819).

It is easier to analyse cost-benefit in defined groups of patients in relation to a single drug, but harder to extrapolate to a population level and especially when pre-emptive testing of multiple genes will have effects over a patient's lifetime(820). However, as the cost of genetic testing goes down and the cost of healthcare increases, is likely that pharmacogenomic testing will become more cost effective. There is significant scope for harm avoidance, in that there is usually an alternative drug or dose reduction to reduce the risk of the ADE. A study by the Vanderbilt University Medical Centre estimated that 65% of their patients were exposed to one or more drugs with a known pharmacogenomic association in a five year period(821). Being aware of a potential risk, while it might increase anxiety, has few ill-effects(822, 823). In addition, cost benefit is not the only driver of test implementation. Additional costs, such as those of litigation may increase demand, particularly from insurance companies or alternatively force pharmaceutical companies to recommend pharmacogenomic testing in advance of prescribing a drug. A case in point is the lawsuit taken by the state of Hawaii against the makers of clopidogrel (Plavix)(824). Clopidogrel is a prodrug that needs to be converted to an active metabolite. Individuals with the *CYP2C19* *2 or *3 haplotypes may not convert clopidogrel to the active metabolite efficiently, leaving them at risk of thrombotic events, such as restenosis post angioplasty. The original trials used mostly Caucasian patients, where the frequency of the *2 and *3 alleles were in the order of 10-20%. However, in Pacific Islanders, the rates were significantly higher, with up to 77% of people carrying one of these alleles, meaning that many Hawaiians do not benefit from standard dosing of clopidogrel, and increased rates of death post-MI were seen in these individuals(825). The manufacturers introduced a black-box warning in 2010, saying that clopidogrel may be unsuitable for individuals with *CYP2C19* *2 and *3 haplotypes, but continued to market it in Hawaii. The case became part of a class action but did not go to trial because of a jurisdiction issue. However, it

raises interesting issues about the recommendation of pharmacogenomic testing and prescribing by drug companies to prevent lawsuits such as this one.

6.3.3.1.3 Benefits in research and drug discovery

Pharmacogenomics is now being used as a strategy in the pharmaceutical industry, often to refine which patients will benefit from which drugs, most usually using a GWAS approach(826, 827). Pharmacogenomics is also being used to look at drugs that have failed previous clinical trials to see if they may be suitable for a proportion of patients. Drugs that have been developed more recently, such as eliglustat, a substrate reduction therapy for Gaucher disease, have come to the market with pharmacogenomic-related prescribing as part of their EMA and FDA licences(615). *CYP2D6* normal and intermediate metabolisers get two 84mg doses daily, while poor metabolisers get only a single dose. The EMA states that it should not be used in ultra-rapid or undetermined metabolisers. There is also advice around reducing dose where CYP2D6 inhibitors are being taken also. The strategy of screening candidate drugs using newer technologies such as pharmacoproteomics and pharmacotranscriptomics is being more widely used(828, 829).

With the cost of bringing a drug to the market estimated by the Tuft's Centre for the Study of Drug Development at more than $2.5 billion, strategies to ensure correct targeting of drugs are important(830, 831). Cancer immunotherapy and ivacaftor for cystic fibrosis are excellent examples of where good patient stratification is key to bringing a drug to market(397, 832, 833). Once a drug has been shown to work in one set of patients and gained a license, it is easier to extend the indications(834). Given the enormous expense of newer drugs and the restricted funds available to the NHS, targeting drugs at patients who are most likely to benefit reduces the number needed to treat and the cost per quality adjusted life year (QALY). This makes it easier to justify funding(835). Furthermore, the ability to predict likely ADEs means that drug companies may be able to develop dosing strategies that will improve the safety profile of their drug, again increasing their ability to get drugs to market.

6.3.3.2 Limitations of pharmacogenomics

6.3.3.2.1 Cost

Discussed in section 6.3.3.1.2, the high cost of testing is one reason that pharmacogenomics is not considered cost effective at present. Prices vary worldwide and are generally higher in North America, with single gene tests being, in general, cheaper than panel-based tests(816). As well as the potential savings discussed previously, there may be potential costs associated with pharmacogenomic testing, for example if the alternative treatment prescribed is more costly than the original. There is also a question of who pays. In the US, where healthcare is insurance based, the insurer will often pay for testing, with or without contributions from the patient or patient's employer (www.genesight.com). In the UK, where few people have private health cover and most people rely on the state-funded NHS, it is likely that pharmacogenomic testing would be mainly provided by the NHS, which means that a cost-effectiveness benefit would be required.

6.3.3.2.2 Potential for incorrect results

As discussed previously, it is difficult to identify rare haplotypes, whether a WGS or SNP genotyping strategy is used. If patients are called as normal metabolisers in the event of a rare haplotype being missed, they are no worse off than if they had never had pharmacogenomic testing, as they are prescribed standard doses of a drug as they would have been without testing. However, the problem arises if they have their drug dose changed in response to pharmacogenomic testing, but a rare haplotype has been missed which would make the change in prescription incorrect and might cause harm to the patient, by giving them a lower or higher dose than necessary. This may be a particular risk for non-Caucasians, for whom there is less understanding about which SNPs make up which haplotype. An example of when a rare haplotype might be missed can be seen in *CYP2D6*, where rs1080985 is seen in haplotype *2 and *11B. *2 is considered normal function while *11B is considered non-functional. It is the presence or absence of additional SNPs that allows the two to be distinguished, and this is something that must be carefully considered when deciding which SNPS to analyse or to include in a test. In this case, the patient would continue on the normal drug dose and so there would be no additional harm from pharmacogenomic testing, although they may still suffer from adverse reactions. However, suboptimal pharmacogenomic testing would provide false reassurance. Commercial pharmacogenomic tests have disclaimers warning patients and clinicians of this possibility. The risk has not been well quantified but, even if low, is worthy of attention(629).

6.3.3.2.3 Obsolete prescribing advice

One-off tests, either SNP or WGS-based, may provide false reassurance. Guidelines are regularly updated, with advice for additional drugs being given and guidance for drugs withdrawn, as in the case of the CPIC guidelines for *DPYD* and Tegafur, which are no longer in use. Occasionally, as mentioned in section 5.3.5, haplotype definitions are changed. Ways of dealing with this are discussed in section 6.3.4.3.4 but it is important that the patient and their prescribers are aware of the limitations of the advice they have been given and where to go to get up-to-date information.

6.3.3.2.4 Lack of clear regulatory structure

Currently, with the UK in a state of regulatory transition pre-Brexit, it is unclear which set of prescribing guidelines should be followed. The DPWG guidelines are European, while the CPIC guidelines are developed by a multinational consortium. In the absence of any clear national decision, and with people accessing private pharmacogenomic testing mainly based in the US, it is likely that implementing pharmacogenomic-based prescribing in the NHS will be challenging. This is discussed further in section 6.3.4. The majority of pharmacogenomic guidelines are recommendations rather than requirements, even with FDA or EMA labels, and so without a nationally agreed strategy, implementation will be in the hands of individual clinicians.

6.3.3.2.5 Impact on patient and need for consent and counselling

With any genetic test, but especially one involving some uncertainty or a risk of secondary findings, a formal consent process is required. As part of this the patient will need to be counselled as to the risks, benefits and limitations of the test(836). Currently, companies such as 23andme

offer limited pharmacogenomic testing, as well as some clinical testing, as part of their SNP genotyping test package with no face-to-face consent or counselling process(837). NHS England has suggested that the new test directory will expand to include pharmacogenomic testing. However it is not yet clear how this might work, and whether pharmacogenomic results would be considered as secondary findings of clinical significance when WGS is done, as CF carrier test results would be, or whether they might be considered separately. It is already challenging explaining the risks and benefits of panel-based or exome testing in a time-constrained clinical appointment without the need to discuss the pros, cons and possible risks of pharmacogenomic testing. Of particular note are some pharmacogenomic scenarios where test results might have implications for other members of the family, with the m.1555A>G mitochondrial hearing loss variant and SNPs associated with drug induced long-QT syndrome highlighted as examples where cascade screening may be necessary(838). In addition, if results will be updated to take account of new discoveries, patients need to understand this and know that they are consenting for tests, the potential significance of which is unknown.

6.3.3.2.6 Extent of current pharmacogenomics knowledge

Current pharmacogenomic guidelines cover only a small proportion of drugs prescribed, and only for the most common haplotypes. There are few guidelines that take account of drug-drug interactions, although polypharmacy is common. In addition to this, for some drugs, while it is clear that their metabolism depends largely on genetic factors, the exact genes and SNPs involved are not yet well understood(839). Further research is required to elucidate additional pharmacogenomic factors to enable the production of more comprehensive guidelines. There are also limitations around prescribing alternative drugs, such as tamoxifen to prevent breast cancer recurrence in *CYP2D6* poor metabolisers, where it is unclear what alternative treatments are best, particularly in premenopausal women(840).

6.3.3.2.7 Factors other than pharmacogenomics

While pharmacogenomics affects how individuals respond to drugs, it is a small part of a much more complex picture, with physiological factors (age, sex, pregnancy), pathological factors (hepatic and renal function), drug interactions and the presence of enzyme inducers and inhibitors all playing important roles in drug efficacy and toxicity(841). Pharmacogenomic guidelines do not take these other factors into account, but they are important to consider alongside pharmacogenomics when prescribing medication.

6.3.3.3 Ethics of pharmacogenomic testing

As with all genetic testing, ethical considerations must be taken into account.

6.3.3.3.1 Access to testing

The cost of pharmacogenomic testing runs the risk of creating an inequitable system, where only the wealthy or those with insurance can access testing. In the UK the NHS aims to avoid health inequality but, as can be seen in cases such as IVF funding, there are discrepancies. The introduction of pharmacogenomic testing by NHS England would be the best way of avoiding this. However, it is likely that, at least initially, pharmacogenomic information will be provided only for

those already having WGS for diagnostic reasons, rather than being widely available. Since those with insurance or independent means can access private testing, inequality is possible.

6.3.3.3.2 Acceptability of testing

Research has shown that there is significant variation in knowledge and acceptance of genetic testing among different ethnic and socioeconomic groups and that this may affect willingness to undergo testing(842-845). Even where access to testing is not limited, uptake by certain groups may be.

6.3.3.3.3 Access to treatment

As previously discussed, sometimes there may not be good alternative treatments available when pharmacogenomic testing reveals that standard treatment is less suitable for particular groups of patients such as in the case of *IFNL3* testing and the antivirals peg-interferon and ribavirin(565). Having a favourable genotype (CC) increases the chances of SVR and allows for shortened therapy regimes, making treatment more cost-effective, while having an unfavourable genotype reduces the chance of SVR, even with longer treatment regimes. However, there are currently few alternative treatments available for hepatitis C, and the consequences of untreated infection are serious, including the risk of chronic liver disease and hepatocellular carcinoma. Denying the patient treatment on the basis of a pharmacogenomic test is problematic, not least because a proportion of patients with an unfavourable genotype will have a good response to treatment. A further issue arises in the case of children as although they will experience similar effects to adults and are at risk of serious adverse events, many pharmacogenomic prescribing guidelines make recommendations for adult patients only(592, 802).

Another consideration is whether or not clinicians are willing to utilise pharmacogenomic prescribing guidelines, and whether their places of work support them in this. In the US some high profile hospitals, such as the Mayo clinic have introduced pharmacogenomic testing on a research basis, with results being included in the medical record and used to inform prescribing (http://mayoresearch.mayo.edu/center-for-individualized-medicine/pharmacogenomics-study.asp). Again, this risks inequality if some research-active hospitals introduce testing while others do not. This could be a particular problem for patients who are mainly treated in the primary care setting.

6.3.3.3.4 Discrimination

As with many types of genetic testing, patients have concerns about the possibility of discrimination by insurers, employers or healthcare providers because of their genetic background. In some cases, there may be concerns that this will affect other family members beyond the original patient. While there is currently a moratorium on insurance companies asking for genetic test results (with some exceptions), there is no guarantee that this will continue to be the case.

6.3.3.3.5 Ethnicity-based disadvantage and discrimination

As discussed in section 5.3.3.2, particular ethnic groups may respond differently to particular drugs, due to a higher rate of haplotypes associated with ultra-rapid or poor metabolism. This could be a disadvantage if pharmacogenomics is used during drug development and trial stages where, if particular haplotypes were excluded, higher numbers of certain ethnic groups could find themselves ineligible for treatment or at greater risk of side effects(846).

6.3.3.3.6 Data storage

A corollary of many of the points above is that patients have concerns about data storage and who might be able to access their genetic data. This is of particular concern when it comes to storing WGS data, where a significant amount of medical information might be accessible in the case of a breach, not just affecting the patient but also their family members(847). Research has shown that it is relatively simple to deanonymise DNA data(848).

6.3.3.3.7 Consent

Consenting for genetic testing is complex. Thought needs to be given to the potential for finding variants of unknown significance and medically relevant secondary findings and how this will be explained to patients. In addition, as mentioned in section 6.3.3.2.5, if results are to be continually updated in line with new pharmacogenomic discoveries, the patient will, in effect, be consenting to tests they cannot know about or understand. The consent process for commercial testing, which may include carrier testing, such as through 23andme and similar companies, is not particularly detailed, and may fail to explain how relevant the results are. For example, it reports only Ashkenazi Jewish *BRCA1* and *BRCA2* mutations, which is not reassuring for anyone without Jewish ancestry, and frequently reports only a small number of common mutations for other syndromes(849). Current commercial consenting processes would not be considered acceptable within the NHS, where pharmacogenomic testing may take place in future. Pharmacogenomic tests have been relatively non-controversial, although, should they be used to decide treatment eligibility, this may change. As part of the 100,000 Genomes project, which had a detailed consent process with participants being enrolled by people with a good knowledge of genetics, consenting to WGS was still challenging for some individuals. The addition of a complex issue such as pharmacogenomics may prove difficult, especially as it is currently not well understood by many clinicians.

6.3.3.3.8 Commercial testing

At present, the main options for testing are from US-based commercial providers. Not only are they not bound by EU data protection laws, but they are often unclear as to what exactly they test for. Many of the companies offer tests tailored to particular medication types, such as a “psychiatric panel”. While they state what genes are included, they often do not state what SNPs and therefore haplotypes are covered by their testing, meaning that it is difficult to know the reliability or accuracy of the test results. In addition, close reading of terms and conditions shows that many companies have ownership of genetic data after testing. Even if current terms and conditions limit data-sharing or exploitation for commercial use, there is nothing to stop them from

changing, especially if a company is sold or subject to takeover. The marketing of some direct-to-consumer genetic tests as a recreational activity, giving information about ancestry and harmless traits such as coriander aversion, may mean that consumers do not give due consideration to medically relevant genetic testing, including pharmacogenomic testing. There is evidence to show that direct-to-consumer providers may return inaccurate results and if companies such as these are providing pharmacogenomic testing, such as in the case of 23andMe and *G6PD* testing, there is a risk that consumers are getting inaccurate information(849). Currently, there is a safety net, at least in the UK, where clinically actionable results are confirmed in a certified laboratory but this would not be possible in the event of large-scale uptake of commercial pharmacogenomic testing.

6.3.4 Challenges of implementing pharmacogenomics in the NHS

6.3.4.1 Cost and economics

The NHS is under considerable funding pressure at present due to multiple factors, among them an aging population, the economic climate, staffing issues and funding. In addition, drug costs are increasing and are likely to continue to do so(850). The introduction of novel technologies and therapies occurs only when a benefit, either financial or in quality of life, is shown. As discussed in section 6.3.3.1.2, there is a potential cost benefit in the introduction of pharmacogenomic testing, but it has been difficult to prove. Utilising the strategy employed here in extracting pharmacogenomic information from WGS data where the data has been produced for other reasons might be a cost effective way of introducing pharmacogenomics, especially when NHS England introduces WGS sequencing to its diagnostic directory. The additional cost of extracting and analysing data would lie in the development or purchase of software that could do this automatically and technology to access results at the prescriber's end. Development and updating of guidelines and the education and training of prescribers would also have associated costs.

6.3.4.2 Patient consent and willingness to participate

This is discussed in section 6.3.3.3.7.

6.3.4.3 Utilising results

A major issue will be how pharmacogenomic data could be utilised in the NHS setting. In order for the results to have any utility at all, prescribing would have to be changed as a result. This has multiple barriers associated with it.

6.3.4.3.1 Feedback and explanation of results to patients

Communication is essential, especially if results are used to determine treatment eligibility or alter medication dose. It is unclear who would take responsibility for this and it is difficult to see how any medical professional could find time for this in their current practice. A public education programme might assist with this although for these campaigns to succeed, the message must be very simple, and pharmacogenomics is a complex concept.

6.3.4.3.2. Guidelines

Currently, some drugs have multiple sets of guidelines associated with them, others have none, while still more have guidelines involving knowing the haplotypes or genotypes of multiple genes or SNPs. For the NHS to adopt pharmacogenomic testing, a single set of NHS-approved guidelines would be required. These could either be developed by the NHS or external guidelines could be adopted but, either way, they would need to be accessible and updatable, possibly as part of the British National Formulary (BNF). Although the future status of the UK with respect to the European Medicines Agency (EMA) is unclear, it is possible that UK guidelines will have to reflect EMA labelling(496).

6.3.4.3.3 Willingness of prescribers to change practice

A WGS approach, meaning that pharmacogenomic data would be available prospectively, would avoid the situation where a prescriber has to await the results of a genetic test. However, the prescriber still has to be aware of the pharmacogenomic results, and understand how drug dose should be altered and why, in order to ensure appropriate implementation. This is something that could be done during medical education, but training would also have to be available to all qualified prescribers. Unequal uptake in certain specialties or geographical regions could lead to inequality of treatment, especially if initially, pharmacogenomic data are only available for the small proportion of patients who require diagnostic WGS. The use of reactive pharmacogenomics, where patients are only tested when being started on a drug, is even more challenging as there is a wait for test results before the drug can be started. There is an intermediate possibility, where subsets of the population could have pharmacogenomic testing at a particular time, for example, at the age of 60 following which drug use increases and there are increased risks of polypharmacy, or when a particular diagnosis, such as that of cancer or psychiatric illness, is made. This approach would require additional research to determine the optimum test time and conditions and to model the potential benefits and risks. Trial before implementation would also be required.

6.3.4.3.4 Access to and update of pharmacogenomic data

In order for data to be available at the point of prescription, it will need either to be widely accessible to all medical professionals involved in the patient's care or be carried by the patient. The EU based UPGx project provides patients with a QR code on a card and doctors and pharmacists with scanners so that data can be accessed as required(639). Additional solutions might include storing data in the cloud or smartphone app that can be regularly updated. All of these systems would have the advantage of being able to update pharmacogenomic guidelines without contacting every patient and clinician, but there are caveats about data protection.

6.3.4.3.5 Computational infrastructure

The NHS has tried and thus far failed to implement a universal medical records system. Centrally stored patient records, accessible to all treating clinicians, would be of enormous assistance in the implementation of pharmacogenomic testing. Alerts could be stored as part of this and be flagged up whenever a patient is being prescribed a drug with a pharmacogenomic guideline

relevant to them. This was seen to be a major advantage when piloting the introduction of pharmacogenomic testing in primary care in the Netherlands, where a single prescribing system is integrated into all GP surgeries and pharmacies(851). Currently in the UK, there is a mixture of paper and various online prescribing software in use, making the implementation of pharmacogenomics challenging.

6.3.4.4 Additional challenges

As discussed in section 6.3.3, there are additional factors to be considered before pharmacogenomics implementation in the NHS. These include ethnic differences in populations, SNPs and haplotypes that are relevant to local populations, the SNPs and haplotypes to be analysed or excluded and relevant non-genetic factors.

6.4 Conclusions and future directions

Pharmacogenomic testing is challenging regardless of the method used. Both SNP-based testing and the use of WGS to extract pharmacogenomic data have their challenges and both have advantages and disadvantages. WGS excels in having the potential to review data regularly and extract additional data as the evidence base supports it. However, optimal use of WGS will require considerable computational resources and additional research to determine exactly how haplotypes should be called. This may require additional experimental data to assess which SNPs relate to which haplotype, which SNPs are always present and which are indicators in some ethnic groups but not others and the exact nature of any phenotypic effect. It is possible that in future SNPs rather than haplotypes will be used to define function.

WGS is an accurate way of extracting pharmacogenomic data and appears to be, for the currently actionable pharmacogenes and some additional pharmacogenes, at least as accurate as SNP-based testing. This agrees with the recently published findings of Reisberg et al. who found that WGS was as accurate as SNP based testing and more accurate than WES(636). Some of the shortcomings of WGS, including the inability to detect phase and the difficulty in determining CNVs, are likely to be overcome by the introduction of long-read sequencing technologies.

Future directions include validating genes that could not be validated, although current stocks of DNA do not allow this. Obtaining further DNA would allow validation of missing samples and further CNV validation. This could be done by Sanger sequencing. Further work on extracting HLA haplotypes is also required and it is possible that bioinformatics tools will help with this(852). A small cohort was tested in this study so replicating these findings in a larger cohort would be important, especially the experimental confirmation of CNVs. Recently published data have confirmed the findings of this study however(636). The development of software that will extract this data automatically and also perhaps provide prescribing guidance would make this a much less laborious task.

The introduction of pharmacogenomics to the NHS is likely to happen in the future, but will come with significant challenges and logistical problems. Resolving these will require decision making

at a national level and significant investment in infrastructure and education, as well as an ongoing commitment to update data and adopt new guidelines. Small studies to assess feasibility in defined groups or clinical areas will clarify requirements and determine how obstacles could be overcome. It is important that testing is introduced because it is believed to be beneficial and not just because it is possible with existing data.

Chapter 7 Conclusions and further work

7.1 Summary of findings

The overall aim of the project was to explore how WGS can be used to increase personalisation of diagnosis and treatments in the clinical setting, and how techniques such as deep phenotyping can be used to assist with this. It looked at the challenges of diagnostic testing using WGS in the investigation of patients with undiagnosed ciliopathies. It outlined the logistics of deep phenotyping and included the development of a flexible deep phenotyping database. It also aimed to show that it is possible to extract pharmacogenomic information from WGS data in a clinically useful manner. The main findings are summarised below.

7.1.1 Whole genome sequencing for genetic disorders is challenging and may not lead to a diagnosis even in cases that appear to be genetic in origin, with variants of unknown significance and secondary findings complicating analysis

7.1.1.1 Whole genome sequencing in a singleton with a clinical diagnosis of BBS

A child with a clinical diagnosis of Bardet-Biedl syndrome had WGS carried out. Following filtering and careful analysis of results, no clearly pathogenic mutations were found in any known BBS genes. Several possible candidate genes were identified.

A heterozygous missense variant and a heterozygous possible promotor variant were identified in *CEP290*, a good candidate gene. However, both were classified as variants of unknown significance and functional work would be required to confirm pathogenicity. In addition, with only maternal samples available, it was not possible to prove that the variants were in *trans* and one would have to be mosaic or *de novo*, given that they were absent in the mother.

A heterozygous *NPHP4* variant, previously reported as pathogenic in nephronophthisis in *trans* with a frameshift variant, was identified. However, the only other variant found in this patient is a heterozygous one that has been evaluated as a benign polymorphism. Other variants such as a heterozygous frameshift variant in *CDHR1* and a heterozygous splice-site variant in *WDR19*, while interesting, were not pursued as no second variant or possible CNV could be identified.

No definitive diagnosis was identified and further work is required to understand the cause of BBS in this patient.

7.1.1.2 Whole genome sequencing in monozygotic twins with a clinical diagnosis of BBS

A pair of monozygotic twins with a clinical diagnosis of Bardet-Biedl syndrome had WGS carried out. Following filtering and careful analysis of results, no clearly pathogenic mutations were found in any known BBS genes. Several possible candidate genes were identified.

Although two heterozygous variants, previously recognised as pathogenic, were seen in *ABCA4*, the patients' phenotype is not in keeping with *ABCA4*-related disease, which, thus far, has been known to consist only of ocular phenotypes. As *ABCA4* is expressed only in the retina, it is difficult to see how it could be implicated in a multisystem disorder. In addition, unlike the majority of patients with a diagnosis of BBS, no eye phenotype has been seen in these patients. One of the variants seen is common, but has been shown to be associated with adult onset macular degeneration. This is a secondary finding, with implications for the patients as they get older, and illustrates the importance of properly informed consent and prior decisions about which variants will be reported back.

A heterozygous missense variant of unknown significance was seen in *CEP164*, a gene known to cause a wide range of ciliopathies. However, no second variant could be identified. There was an area of approximately 30 base pairs with reduced coverage in exon 29, which might indicate a partial exon deletion. Small CNVs are difficult to confirm as they will not be picked up by array CGH where multiple missing probes are required to call a CNV. Equally, if in an area without benign variation, they will not be seen on a SNP array. Deep resequencing of the exon would be a possible way of investigating further, as would functional work, such as RTPCR or immunohistochemistry.

A further heterozygous variant of unknown significance was seen in *INPP5E*, a gene implicated in MORM and Joubert syndromes. A drop in coverage near the end of exon 7 was also seen, raising the possibility that the second variant could be partial exon deletion, but when the adjacent intron was considered, this looked like poor coverage. This was supported when other samples were reviewed.

A heterozygous variant in *TRIP12* that has previously been reported as pathogenic was seen. Mutations in *TRIP12* are believed to cause intellectual disability and have also been associated with obesity. In addition, the protein product of *TRIP12* appears to interact with genes involved in ciliogenesis. This would explain part of the observed phenotype and would fit with the fact that no retinitis pigmentosa or renal problems have been observed. It would not, however, explain the patients' polydactyly. Variants such as the one seen in *GDF5* might explain polydactyly, but again this could not be confirmed.

Functional work would be required to further investigate all of the ciliopathy gene variants. While no clear pathogenic variants were found in ciliopathy genes, the heterozygous variant in *TRIP12* is actionable, in that it has previously been reported as pathogenic. It is being validated in an NHS laboratory. It should also be confirmed that this is *de novo*, rather than inherited from a parent without learning difficulties. This constitutes an unexpected finding, as it is not a known ciliopathy gene. As it is inherited in an autosomal dominant manner, the recurrence risk will differ to the risk of one in four given for recurrence of BBS. It will either be higher (one in two if inherited from a parent and still believed to be pathogenic) or lower (if *de novo*). In view of the parents' normal learning there is also the possibility of parental mosaicism, which would make the risk harder to quantify. It also has implications for the twins should they start their own families, as the offspring

recurrence risk would be one in two if this gene were causing learning difficulties and obesity. In addition, it would remove the diagnosis of BBS which the family had previously been given, which is always difficult, particularly for families active in a support group or seen in a specialist clinic.

While the heterozygous *TRIP12* variant may be a cause for some aspects of the patients' phenotype, it does not explain all the features seen and further work is required.

7.1.1.3 Conclusions about whole genome diagnostics

Despite extensive review of the whole genome sequences of both the singleton and monozygotic twins, no clearly pathogenic variants were identified that would explain their phenotypes. Many VUS were identified, and in the case of the twins, a variant in a known intellectual disability gene was identified instead. This highlights some of the difficulties in using WGS for diagnostics. Firstly, the fact that no definitely pathogenic variants were identified means that either variants are being missed for some reason, for example a small CNV, poor coverage of the genes of interest, filtering and removal of a variant of interest due to its not passing quality control measures or a non-coding variant, the significance of which is not known. In this study, despite having WGS available, non-coding variants were only considered if a coding variant was identified in the gene first. This is because our ability to interpret deep intronic variants is very limited. However, it is certainly possible that remaining undiagnosed cases of rare disease with known causes may be due to non-coding variants such as in distant enhancers and this has been shown to be the case for ciliopathy genes(853). Secondly, if any of the VUS identified are responsible, functional work, which is time-consuming and expensive would be required to prove it. Thirdly, the fact that the twins' intellectual disability may be caused by a variant in an intellectual disability gene rather than by variants in a ciliopathy gene shows the importance of good consent including covering the possibility of unexpected findings. No secondary findings of clinical import were identified in these cases, but when WGS is being done as a diagnostic test, consent should include a discussion of what should or should not be reported back. The risk is especially high if filters such as all known OMIM morbid genes are being applied.

Advantages include the ability to review these genomes regularly as new genes are identified or our understanding of non-coding variation increases. In addition, there are other advantages, such as better quality data and the ability to use the data for other purposes, such as screening for other disease risk factors or pharmacogenomic variation. Disadvantages include cost, difficulties in analysis and the cost of data storage. There are additional issues which could be argued as either advantages or disadvantages, including the ability to identify secondary variants of clinical significance and the ability to review data as new genes or disease mechanisms are identified. One of the most concerning issues with NGS and WGS in particular is the very large number of variants being discovered. Most diagnostic laboratories have no capacity to perform functional studies and it is very likely that many pathogenic variants are being reported as VUS.

The conclusion is that WGS has both advantages and disadvantages, and while there are technical reasons for choosing it over WES, at present most of the deep intronic variants are uninterpretable.

7.1.2 Deep phenotyping and development of phenotyping database

Deep phenotyping is vital for accurate diagnostics, patient stratification, analysis of clinical trial data and more. Even in the age of agnostic genetic testing, phenotyping is important in determining pathogenicity of variants and whether a mutation is a good fit for the patient in question. In this study a custom database was built for the storage of deep phenotyping data, patients were phenotyped, and their information stored in the database.

7.1.2.1. Building of a custom database

A bespoke database was developed as the available databases were either too simplistic, too inflexible or too expensive for use in the project in question. MS Access database software was chosen because it is freely available, relatively easy to customise and possible to store in secure locations such as the UCL Data Safe Haven. In addition, it is easy to train individuals to enter or access data when appropriate and easy to output data in anonymised form, including as text files.

There were advantages to building a custom database. It could be modified to accept any data type as that data type was encountered. Cross-linking of data sheets reduced the laborious task of data entry and minimised the chance of error. The database proved useful in multiomics studies.

A major challenge was the multiple ways in which clinical data are recorded and stored in the NHS. Even hospitals which have electronic records use multiple databases, for example pathology, radiology, genetics and prescribing, and many still incorporate scanned copies of paper notes. This is something that is likely to improve in future and datamining software may allow data to be extracted directly form medical records. Storage of data is an ongoing issue and is costly. While eventually, it may become cheaper to repeat genome sequencing than to store data, this will not be the case for clinical information that has been collated manually, and is not the case at present. The development of new and cheaper methods of data storage or the ability to extract data as required is vital. Consideration was given to compliance with relevant data protection legislation and information governance regulations. Loss of data exposed a weakness in the UCL Data Safe Haven that has now been rectified.

Custom deep phenotyping databases are relatively easy to build, maintain and manipulate. They may provide a good solution when commercial databases do not meet requirements or are too expensive. However, data entry is labour intensive.

7.1.2.2 Use of PhenoTips® database

Storing clinical descriptors is best done using a medical ontology and following review of the options, HPO terms were chosen as the most comprehensive ontology for the purpose. It was determined that an HPO database such as PhenoTips®, which was linked to the MS Access database by patient identification number was a better way to store the data than trying to integrate it with the MS Access database, both because the number of HPO terms made the MS Access database unwieldy, but also because PhenoTips® prompts for the entry of the most exact

HPO term. It also prompts for the consideration of additional HPO terms. The two datasets were combined as text files for integration with multiomics data.

7.1.2.3 Utilisation of data

One of the outcomes of the project was the realisation that for multiomics studies, longitudinal data are extremely important. However, longitudinal data that cannot be linked, either to a multiomics sample or a clinical report or outcome, are extremely difficult to use. Data were used for burden analysis studies, patient stratification for omics analysis, in machine learning and in the development of an integrated clinical and multiomics database as well as in informing pharmacogenomic analysis. In particular, in pharmacogenomic analysis, it allowed the identification of patients who had already been prescribed a drug with pharmacogenomic prescribing guidelines, and in particular, two patients who had variants in the relevant genes.

7.1.2.4 Conclusions about deep phenotyping and development of phenotyping database

Deep phenotyping is complex and time consuming, especially because of how NHS records are held. Longitudinal data are useful for whole genome studies, but difficult to extract. HPO terms are good for describing physical features, but may be less helpful for other more dynamic disease processes. A database such as PhenoTips® is helpful because it pushes clinicians to choose the most detailed descriptors. Building a custom database is a cost effective way to meet requirements although care needs to be taken with how and where data are stored. Phenotypic data can be utilised in many ways, but considering how data will be outputted for integration is important.

7.1.3 Use of whole genome sequences for extraction of pharmacogenomic data

7.1.3.1 Extraction of data and generation of prescribing reports

Pharmacogenomic data were initially extracted for 17 actionable pharmacogenes. The SNPs selected for analysis were those associated with peer reviewed prescribing guidance, generally from CPIC or DPWG. SNP data were extracted and combined to give haplotype and diplotype data for each of 84 patients, a cohort which included two sets of monozygotic twins and ten parent-child trios. Prescribing advice was then generated for each patient using a traffic-light system; red for drugs to be avoided, orange for drugs to be used with caution and green for drugs to be used as directed. All patients were found to have some variants that would lead to changes in prescribing or monitoring should they be prescribed a relevant drug. The median number of genes with variants that would alter prescribing was four. Even within this small cohort, individuals were identified with variants in a gene that would affect the metabolism of drugs they had already been prescribed.

7.1.3.2. Validation of data

Haplotype or genotype frequency data were compared to published frequencies and found to be similar. Validation was performed in 72 of 84 patients using a pharmacogenomics SNP assay developed by Congenica Ltd. This assay covered most but not all of the pharmacogenomic genes selected originally and a number of additional pharmacogenes. These additional SNPs were also

extracted from WGS data for comparison. WGS data extraction and SNP genotyping performed very similarly for all of the SNPs checked. Most failures were due to non-amplification, and these were generally due to low DNA concentration. Some samples clustered ambiguously between groups of SNPs in the validation data. Occasionally, this appeared to be due to additional nearby SNPs. Only three true disagreements were seen, two of which were in tracking SNPs. A final disagreement was seen in a case where the allele fraction was low and the SNP was not called in WGS. The overall true disagreement rate when tracking SNPs and clustering issues were ignored was <0.01%. Of the five copy number variants seen in *CYP2D6*, confirmation was only possible for two, both of which were confirmed. The conclusion was that WGS can be used for the successful extraction of pharmacogenomic data.

7.1.3.3 Challenges of pharmacogenomics implementation

Although the economic benefits of pharmacogenomic testing have yet to be conclusively proven, it is likely that if pharmacogenomic data can be successfully extracted from WGS data, then it is something that will increasingly be done as we move towards using WGS for diagnostics. While it is possible to extract pharmacogenomic data from WGS, there are still some genes and haplotypes for which this is challenging. Assessment of techniques such as long-read sequencing is required. At present this would add significantly to costs, but would solve issues such as phasing and the inability to call HLA haplotypes. In addition, significant investment will have to be made in infrastructure, training and staffing to implement this in the clinical setting. While there are advantages to using WGS, such as the ability to review data, there are also disadvantages including the possibility of identifying VUS and the obligation to review data. Issues of equity and ethics must be considered.

7.2 Future Directions

7.2.1 Diagnosis in patients

7.2.1.1 Patient BBS-018

In order to elucidate the cause of BBS in this patient further work is required. Trio WES or WGS sequencing is not an option for this patient as paternal DNA samples are unavailable. At present, the two best candidate genes are *CEP290* and *NPHP4*. In order to look at this further, functional work would be required. This could be done by looking at levels of RNA with RTPCR, and looking at protein expression with immunohistochemistry or proteomics studies. However, this might not help if the protein was present but non-functional. Deep intronic variants could be explored for effects on splicing by doing RTPCR of each exon to confirm the presence of cDNA. However, it may be that the only way of confirming pathogenicity is to make an animal model. Cell culture is unlikely to be very helpful as BBS patient fibroblasts do not generally have a recognisable ciliary phenotype(854). Other possibilities might include an array to identify duplications that may have been missed during WGS. A standard array CGH could identify larger deletions, but to identify very small deletions a specialist SNP array, such as the Illumina® CytoSNP-12 array, would be required. In addition, WGS results should be reviewed at set intervals or whenever a novel ciliopathy gene is identified.

7.2.1.2 Patients BBS-016 and BBS017

The first priority would be to obtain parental DNA to confirm whether the *TRIP12* variant seen is inherited or *de novo*. If *de novo*, that would add weight to it being a cause of learning difficulties and obesity in the probands. However, if it was found to be inherited from an intellectually normal parent, it would be less likely to be causative, and furthermore, would have wider implications as it is recorded, in the literature and the *TRIP12* database, as pathogenic.

Additional work that could be done includes exploring pathogenicity of the variants seen, in particular *INPP5E* and *CEP164*. This could be done as suggested for variants seen in BBS-018. An array would have the potential to identify small deletions not picked up during WGS. Finally, should parental samples be obtained, there is always the possibility of performing trio WES or WGS.

7.2.2 Deep phenotyping

Additional patients are being enrolled in ongoing multiomics studies as part of an expansion of HIGH-5. Ongoing follow up of patients currently in the database will provide longitudinal phenotyping information which is invaluable in multiomics studies. Adding these individuals to the database would provide both further deep phenotyping data and also allow assessment of which additional tables would be useful to add. Exploring the possibility of integrating HPO terms directly into the database would simplify data storage and remove the pre-analysis integration step currently required. In addition, work is ongoing to integrate the phenotyping database with software for multiomics analysis.

Additional work might include looking at extracting data directly from electronic records rather than having a manual data entry process and looking at further ways in which the data could be utilised.

7.2.3 Pharmacogenomics

Ongoing work includes expanding the cohort as more WGS sequence becomes available and adding additional genes as more guidelines are developed. Further work is being done to look at whether it is possible to use WGS to extract data for additional *HLA-B* haplotypes and which SNPs best represent *HLA-B* *15-02. Owing to a lack of DNA, there are several things that need clarification including *DPYD* haplotypes for four individuals, SNPs that clustered ambiguously during validation and CNVs in *CYP2D6* that still require confirmation. Work is also required to look at automating SNP calling and verifying the accuracy of this.

In terms of future work, plans for a pharmacogenomics implementation study in a large NHS trust providing secondary and tertiary care are being developed. This will use the SNP genotyping platform initially, and act as proof of the concept that it is possible to utilise pharmacogenomics to direct patient care. It will allow thought to be given to patient consent and how this might be obtained. It may be that in future all prescribing will be pharmacogenomics based and testing will become so routine that a formal genetic consenting process will be considered unnecessary. One

group already identified as a possible cohort are stroke patients, who may benefit from alternative antiplatelet therapies to the generally prescribed clopidogrel if they are intermediate or poor *CYP2C19* metabolisers. Also being considered is a prospective implementation of untargeted pharmacogenomic testing in general practice patients over the age of 50. Work is beginning on a pharmacogenomics module that will be available to doctors though Health Education England as well as implementing undergraduate and postgraduate pharmacogenomics teaching in various sites (University College London, St Georges University London, UK clinical pharmacology trainees).

However, in order to implement pharmacogenomics more generally in the NHS, as well as an analysis of cost effectiveness, some important decisions need to be made at national level. Examples include which prescribing guidelines will be utilised and whether testing will be done nationally or at the discretion of hospital trusts. Pilot projects will be required, as was done with the 100,000 Genomes project prior to the introduction of the national test directory. Genomics England is in the process of developing this. Computational infrastructure, training, prescriber education and patient resources will also be required. More importantly, any implementation should be done in an equitable manner, ensuring that patients are not disadvantaged by their location, gender, ethnicity or other characteristics.

Appendices

Appendix 1- Causative genes

ARL6 (BBS3)	*BBS10*	*MKKS (BBS6)*
BBIP1 (BBS18)	*BBS12*	*MKS1 (BBS13)*
BBS1	*C8orf37 (BBS21)*	*PTHB1 (BBS9)*
BBS2	*CEP290 (BBS14)*	*SDCCAG8 (BBS16)*
BBS4	*IFT27 (BBS19)*	*TRIM32 (BBS11)*
BBS5	*IFT74 (BBS20)*	*TTC8 (BBS8)*
BBS7	*LZTFL1 (BBS17)*	*WDPCP (BBS15)*

Table A1.1 Genes causing Bardet-Biedl syndrome

ADGRV1	Usher syndrome, type 2C
CDH23	Usher syndrome, type 1D
CIB2	Usher syndrome, type IJ
CLRN1	Usher syndrome, type 3A
HARS	Usher syndrome, type 3B
MYO7A	Usher syndrome, type 1B
PCDH15	Usher syndrome, type 1F
PDZD7	Usher syndrome, type IIC
SANS	Usher syndrome, type 1G
USH1C	Usher syndrome, type 1C
USH1E	Usher syndrome, type 1E
USH1H	Usher syndrome, type 1H
USH1K	Usher syndrome, type IK
USH2A	Usher syndrome, type 2A
WHRN	Usher syndrome, type 2D

Table A1.2 Genes causing Usher syndrome

ADA	*DCLRE1C*	*HPS6*	*LIG4*	*PIK3R1*	*STAT1*
ADAM17	*DKC1*	*ICOS*	*LRBA*	*PLCG2*	*STAT3*
AICDA	*DOCK8*	*IKBKG*	*MASP2*	*PTEN*	*STXBP2*
BTK	*EPCAM*	*IL10*	*MEFV*	*RAG1*	*TGFBR1*
CD3G	*FERMT1*	*IL10RA*	*MVK*	*RAG2*	*TGFBR2*
CD40LG	*FOXP3*	*IL10RB*	*NCF1*	*RIPK2*	*TTC37*
COL7A1	*G6PC3*	*IL21*	*NCF2*	*RTEL1*	*TTC7A*
CTLA4	*GUCY2C*	*IL2RA*	*NCF4*	*SH2D1A*	*WAS*
CYBA	*HPS1*	*IL2RG*	*NPC1*	*SKIV2L*	*XIAP*
CYBB	*HPS4*	*ITGB2*	*PIK3CD*	*SLC37A4*	*ZAP70*

Table A1.3 Genes causing monogenic very early onset inflammatory bowel disease (VEOIBD)

GENE	ALMS	BBS	EVC	JATD	JBTS	LCA	MKKS	MKS	MORM	NPHP	OFD	SLS	USH	Other
ADGRV1														
AHI1														
AIPL1														RP
ALMS1														
ANKS6														
ARL13B														
ARL6														
ARMC9														
ATD														
B9D1														
B9D2														
BBIP10														
BBS1														
BBS2														
BBS4														
BBS5														
BBS7														
BBS9														
BBS10														
BBS12														
C2CD3														
C2orf71														RP
C5orf42														
C8orf37														
CC2D2A														COACH
CDH23														
CEP41														
CEP83														
CEP104														
CEP120														
CEP164														
CEP290														

GENE	ALMS	BBS	EVC	JATD	JBTS	LCA	MKKS	MKS	MORM	NPHP	OFD	SLS	USH	Other
CLRN1														RP
CSPP1														
DCDC2														
DDX59														
DYNC2H1														
DYNC2LI1														
EVC														
EVC2														
EXOC4														
EXOC8														
GDF6														
GLIS2														
GPR98														
GUCY2D														
HARS														
HYLS1														
IFT27														
IFT74														
IFT52														
IFT80														
IFT122														
IFT140														
IFT172														
IMPDH1														RP
INPP5E														
INVS														
IQCB1														
KCNJ13														
KIAA0586														
KIAA0588														
KIAA0753														
KIF7														
KIF14														

GENE	ALMS	BBS	EVC	JATD	JBTS	LCA	MKKS	MKS	MORM	NPHP	OFD	SLS	USH	Other
LCA5														
MAPKBP1														
MKKS														
MKS1														
MYO7A														
NEK1														
NEK8														
NMNAT1														
NPHP1														
NPHP3														
NPHP4														
OFD1														
PCDH15														
PDE6D														
PDZD7														
PIBF1														
PRPH2														
RD3														
RDH12														
RPE65														
RPGRIP1														RP
RPGRIP1L														COACH
SDCCAG8														
SLSN3														
SPATA7														RP
SRTD12														
SUFU														
TCTEX1D2														
TCTN1														
TCTN2														
TCTN3														
TMEM67														COACH

GENE	ALMS	BBS	EVC	JATD	JBTS	LCA	MKKS	MKS	MORM	NPHP	OFD	SLS	USH	Other
TMEM107														
TMEM216														
TRAF3IP1														
TRIM32														
TTC21B														
TTC8														
TULP1														RP
USH1C														
USH1E														
USH1G														
USH1H														
USH1K														
USH2A														
WDPCP														
WDR19														
WDR35														
WDR60														
WHRN														
XPNPEP3														
ZNF423														

Table A1.4 Genes causing ciliopathies. Data taken from www.omim.org and Mitcheson and Valente(2107)(142)

ALMS- Alstrom syndrome, BBS- Bardet Biedl syndrome, COACH- cerebellar vermis hypo/aplasia, oligophrenia, congenital ataxia, ocular coloboma, and hepatic fibrosis, EVC- Ellis van Creveld syndrome, JATD- Jeune asphyxiating thoracic dystrophy, JBTS- Joubert syndrome, OFD- orofaciodigital syndrome, LCA- Leber congenital amaurosis, MKKS- McKusick-Kaufman syndrome, MKS- Meckel syndrome, MORM- mental retardation-obesity- retinal dystrophy-micropenis, NPH- nephronophthisis, RP- retinitis pigmentosa, SLS- Senior-Loken syndrome, USH- Usher syndrome.

Appendix 2- Gene list filters

ARL6 (BBS3)	*BBS12*	*MKS1 (BBS13)*
BBIP1 (BBS18)	*C8orf37 (BBS21)*	*PTHB1 (BBS9)*
BBS1	*CCDC28B* (modifier)	*SDCCAG8 (BBS16)*
BBS2	*CEP290 (BBS14)*	*TMEM67* (modifier)
BBS4	*IFT27 (BBS19)*	*TRIM32 (BBS11)*
BBS5	*IFT74 (BBS20)*	*TTC8 (BBS8)*
BBS7	*LZTFL1 (BBS17)*	*WDPCP (BBS15)*
BBS10	*MKKS (BBS6)*	

Table A2.1 BBS diagnostic gene list filter, courtesy of North East Thames Regional Genetics Centre

ACVR2B	*C2orf71*	*DNAAF1*	*HYLS1*	*NKX2-5*	*RSPH4A*	*TSC1*
AHI1	*C5orf42*	*DNAAF2*	*IFT43*	*NME8*	*RSPH9*	*TSC2*
AIPL1	*CC2D2A*	*DNAAF3*	*IFT80*	*NODAL*	*SCNN1A*	*TTC21B*
ARL13B	*CCDC28B*	*DNAH11*	*IMPDH1*	*NPHP1*	*SCNN1B*	*TTC8*
ARL6	*CCDC39*	*DNAH5*	*INVS*	*NPHP3*	*SCNN1G*	*TULP1*
ATXN10	*CCDC40*	*DNAI1*	*IQCB1*	*NPHP4*	*SDCCAG8*	*UMOD*
B9D1	*CDH23*	*DNAI2*	*KCNJ13*	*OFD1*	*SPATA7*	*USH1C*
B9D2	*CEP164*	*DNAL1*	*KIF7*	*PCDH15*	*TCTN1*	*USH1G*
BBS1	*CEP290*	*DYNC2H1*	*LCA5*	*PKD2*	*TCTN2*	*USH2A*
BBS10	*CEP41*	*EVC*	*LEFTY2*	*PKHD1*	*TMEM138*	*VHL*
BBS12	*CFTR*	*EVC2*	*LRAT*	*RD3*	*TMEM216*	*WDPCP*
BBS2	*CLRN1*	*FOXH1*	*MKKS*	*RDH12*	*TMEM231*	*WDR19*
BBS4	*CRB1*	*GDF1*	*MKS1*	*RPE65*	*TMEM237*	*WDR35*
BBS5	*CRELD1*	*GLIS2*	*MYO7A*	*RPGR*	*TMEM67*	*XPNPEP3*
BBS7	*CRX*	*GPR98*	*NEK1*	*RPGRIP1*	*TOPORS*	*ZIC3*
BBS9	*DFNB31*	*GUCY2D*	*NEK8*	*RPGRIP1L*	*TRIM32*	*ZNF423*

Table A2.2 Ciliopathy diagnostic gene list filter, courtesy of North East Thames Regional Genetics Centre

AHI1	*CEP89*	*EXOC6*	*KIF24*	*PAFAH1B1*	*SDCCAG8*	*TTBK2*
AK7	*CEP97*	*EXOC6B*	*KIF27*	*PARD3*	*SGK196*	*TTC12*
AK8	*CLDN2*	*FAM161A*	*KIF3A*	*PARD6A*	*SHH*	*TTC21B*
ALMS1	*CLUAP1*	*FBF1*	*KIF3B*	*PCDH15*	*SLC47A2*	*TTC26*
ARF4	*CNGA2*	*FLNA*	*KIF3C*	*PCM1*	*SMO*	*TTC29*
ARL13B	*CNGA4*	*FOPNL*	*KIF7*	*PDE6D*	*SNAP25*	*TTC30A*
ARL3	*CNGB1*	*FOXJ1*	*LCA5*	*PDZD7*	*SNX10*	*TTC30B*
ARL6	*CP110*	*FUZ*	*LRRC6*	*PHF17*	*SPA17*	*TTC8*
ASAP1	*CRB3*	*GAS8*	*LZTFL1*	*PIBF1*	*SPAG16*	*TTK*
ATXN10	*CROCC*	*GLI1*	*MAK*	*PKD1*	*SPAG17*	*TTLL3*
AZI1	*CTNNB1*	*GLI2*	*MAL*	*PKD1L1*	*SPAG6*	*TTLL6*
B9D1	*DCDC2*	*GLI3*	*MAPRE1*	*PKD2*	*SPATA7*	*TTLL9*
B9D2	*DCDC2*	*GLIS2*	*MCHR1*	*PKHD1*	*SPEF2*	*TUBA1A*
BBS1	*DCY3*	*GPR161*	*MDM1*	*PLK1*	*SSNA1*	*TUBA1C*
BBS10	*DFNB31*	*GPR98*	*MKKS*	*POC1A*	*SSTR3*	*TUBA4A*
BBS12	*DISC1*	*GSK3B*	*MKS1*	*PTCH1*	*STIL*	*TUBB2A*
BBS2	*DNAAF1*	*HAP1*	*MLF1*	*PTPDC1*	*STK36*	*TUBB2B*
BBS4	*DNAAF2*	*HEATR2*	*MNS1*	*RAB11A*	*STK38L*	*TUBB3*
BBS5	*DNAAF3*	*HNF1B*	*MYO15A*	*RAB11FIP3*	*STOML3*	*TUBE1*
BBS7	*DNAH1*	*HSPA8*	*MYO7A*	*RAB17*	*STX3*	*TUBGCP2*
BBS9	*DNAH10*	*HSPB11*	*NEK1*	*RAB23*	*SUFU*	*TUBGCP3*
C21orf2	*DNAH11*	*HTR6*	*NEK2*	*RAB3IP*	*SYNE2*	*TUBGCP4*
C2CD3	*DNAH2*	*HTT*	*NEK4*	*RAB8A*	*TBC1D30*	*TUBGCP5*
C2orf71	*DNAH5*	*HYDIN*	*NEK8*	*RABL5*	*TBC1D7*	*TUBGCP6*
C8orf37	*DNAH6*	*HYLS1*	*NGFR*	*RAN*	*TCTN1*	*TULP1*
CBY1	*DNAI1*	*IFT122*	*NIN*	*RANBP1*	*TCTN2*	*TULP3*
CC2D2A	*DNAI2*	*IFT140*	*NINL*	*RFX3*	*TCTN3*	*ULK4*
CCDC103	*DNAL1*	*IFT172*	*NME5*	*RILPL1*	*TEKT2*	*USH1C*
CCDC114	*DNALI1*	*IFT20*	*NME7*	*RILPL2*	*TEKT4*	*USH1G*
CCDC164	*DPCD*	*IFT27*	*NME8*	*ROPN1L*	*TEKT5*	*USH2A*
CCDC28B	*DPYSL2*	*IFT43*	*NOTO*	*RP1*	*TMEM138*	*VDAC3*
CCDC37	*DRD1*	*IFT46*	*NPHP1*	*RP2*	*TMEM216*	*VHL*
CCDC39	*DRD2*	*IFT52*	*NPHP3*	*RPGR*	*TMEM231*	*WDPCP*
CCDC40	*DRD5*	*IFT57*	*NPHP4*	*RPGRIP1*	*TMEM237*	*WDR19*
CCDC41	*DVL1*	*IFT74*	*NUP214*	*RPGRIP1L*	*TMEM67*	*WDR35*
CDH23	*DYNC2H1*	*IFT80*	*NUP35*	*RSPH1*	*TNPO1*	*WDR60*
CENPJ	*DYNLT1*	*IFT81*	*NUP37*	*RSPH3*	*TOPORS*	*WDR78*
CEP104	*DYX1C1*	*IFT88*	*NUP62*	*RSPH4A*	*TPPP2*	*XPNPEP3*
CEP135	*EFHC1*	*INPP5E*	*NUP93*	*RSPH9*	*TRAF3IP1*	*ZNF423*
CEP164	*EVC*	*INTU*	*OCRL*	*RTTN*	*TRAPPC10*	
CEP250	*EVC2*	*INVS*	*ODF2*	*SASS6*	*TRAPPC3*	
CEP290	*EXOC3*	*IQCB1*	*OFD1*	*SCLT1*	*TRAPPC9*	
CEP41	*EXOC4*	*KIF17*	*ORC1*	*SEPT2*	*TRIM32*	
CEP72	*EXOC5*	*KIF19*	*PACRG*	*SEPT7*	*TRIP11*	

Table A2.3 Cilia-associated gene list filter, from www.syscilia.org

ABCA4	*BBIP1*	*CATSPER4*	*CLDN2*	*DVL1*	*FSCB*	*LRRC43*
ABHD6	*BBS1*	*CATSPERB*	*CLTC*	*DYDC1*	*FUZ*	*LRRC46*
ABLIM1	*BBS10*	*CATSPERD*	*CLUAP1*	*DYDC2*	*GABARAP*	*LRRC48*
ABLIM3	*BBS12*	*CATSPERG*	*CNGA1*	*DYNC2H1*	*GAPDHS*	*LRRC49*
ACTL7A	*BBS2*	*CAV1*	*CNGA2*	*DYNC2LI1*	*GAS8*	*LRRC6*
ACTR2	*BBS4*	*CBY1*	*CNGA3*	*DYNLL1*	*GLI1*	*LRRC73*
ADCY3	*BBS5*	*CC2D2A*	*CNGA4*	*DYNLL2*	*GLI2*	*LZTFL1*
ADCY5	*BBS7*	*CCDC103*	*CNGB1*	*DYNLRB1*	*GLI3*	*MAATS1*
ADCY6	*BBS9*	*CCDC104*	*CNGB3*	*DYNLRB2*	*GLIPR1L1*	*MAGI2*
ADGB	*BEST2*	*CCDC11*	*CNOT10*	*DYNLT1*	*GLIS2*	*MAK*
ADRBK1	*C10orf107*	*CCDC113*	*COPS8*	*DYX1C1*	*GNA11*	*MAL*
AGBL2	*C11orf49*	*CCDC114*	*COQ10A*	*DZANK1*	*GNAQ*	*MAP1A*
AGBL4	*C11orf63*	*CCDC13*	*CRB3*	*DZIP1*	*GNAS*	*MAP1B*
AGR3	*C11orf70*	*CCDC135*	*CROCC*	*DZIP1L*	*GNAT1*	*MAP1LC3B*
AHI1	*C11orf74*	*CCDC146*	*CSNK1A1*	*E2F4*	*GNAT2*	*MAP6*
AK1	*C11orf88*	*CCDC147*	*CSNK1D*	*EEF1A1*	*GNAT3*	*MAP9*
AK2	*C12orf10*	*CCDC151*	*CSNK1G1*	*EFCAB1*	*GNB1*	*MAPRE1*
AK7	*C12orf55*	*CCDC164*	*CTD2373H*	*EFCAB12*	*GNB5*	*MAPRE3*
AK8	*C14orf79*	*CCDC17*	*CTNNB1*	*EFCAB2*	*GNGT1*	*MAPT*
AKAP14	*C15orf26*	*CCDC170*	*CYB5D1*	*EFCAB7*	*GPR161*	*MARK4*
AKAP3	*C16orf71*	*CCDC173*	*CYFIP2*	*EFHC1*	*GPR83*	*MCHR1*
AKAP4	*C16orf80*	*CCDC176*	*CYLD*	*EFHC2*	*GPR98*	*MCIN*
AKAP9	*C1orf114*	*CCDC19*	*CYS1*	*EFTUD2*	*GRK1*	*MDH1B*
AKT1	*C1orf173*	*CCDC28B*	*DCDC2*	*EHD1*	*GRXCR1*	*MDM1*
ALDH1A1	*C1orf192*	*CCDC33*	*DCTN1*	*EHD3*	*GSK3B*	*MERTK*
ALMS1	*C1orf194*	*CCDC37*	*DDAH1*	*ELMOD2*	*GSTA1*	*MKKS*
ALS2CR12	*C1orf222*	*CCDC39*	*DFNB31*	*ELMOD3*	*GSTA2*	*MKS1*
AMBRA1	*C20orf26*	*CCDC40*	*DHRS3*	*EML1*	*GSTM3*	*MLF1*
ANKMY2	*C20orf85*	*CCDC41*	*DISC1*	*ENAH*	*GUCA1A*	*MNS1*
ANKS6	*C21orf2*	*CCDC65*	*DLD*	*ENKD1*	*GUCA1B*	*MOK*
ANO2	*C21orf58*	*CCDC74A*	*DMD*	*ENKUR*	*GUCA1C*	*MORN2*
ANXA1	*C21orf59*	*CCDC74B*	*DNAAF1*	*ENO4*	*GUCA2B*	*MORN3*
AP3M2	*C2CD3*	*CCDC78*	*DNAAF2*	*ENPP5*	*GUCY2D*	*MORN5*
APOA1BP	*C2orf71*	*CCDC81*	*DNAAF3*	*EPS15*	*GUCY2F*	*MYB*
APOBEC4	*C2orf73*	*CCL15*	*DNAH1*	*EVC*	*HAP1*	*MYH10*
APOO	*C4orf22*	*CCNO*	*DNAH10*	*EVC2*	*HAVCR1*	*MYO15A*
ARF4	*C5orf30*	*CCP110*	*DNAH11*	*EXOC3*	*HEATR2*	*MYO3B*
ARFGEF2	*C5orf49*	*CCSAP*	*DNAH12*	*EXOC4*	*HHIP*	*MYO5A*
ARL13B	*C6orf118*	*CCT2*	*DNAH14*	*EXOC5*	*HIF1A*	*MYO5B*
ARL2BP	*C6orf165*	*CCT3*	*DNAH17*	*EXOC6*	*HIPK1*	*MYO7A*
ARL3	*C6orf170*	*CDH23*	*DNAH2*	*EXOC6B*	*HK1*	*MYOC*
ARL6	*C7orf57*	*CDHR1*	*DNAH3*	*EZR*	*HMGB2*	*MYRIP*
ARMC3	*C7orf63*	*CDHR3*	*DNAH5*	*FAIM*	*HNF1A*	*NAPEPLD*
ARMC4	*C8orf37*	*CDK20*	*DNAH6*	*FAM154A*	*HNF1B*	*NBEA*
ARR3	*C8orf47*	*CELSR2*	*DNAH7*	*FAM154B*	*HSP90AB1*	*NEDD1*
ASAP1	*C9orf116*	*CELSR3*	*DNAH8*	*FAM161A*	*HSP90B1*	*NEK1*
ATG14	*C9orf117*	*CENPJ*	*DNAH9*	*FAM179A*	*HSPA1L*	*NEK11*
ATG16L1	*C9orf135*	*CEP104*	*DNAI1*	*FAM216B*	*HSPA1L*	*NEK2*
ATG5	*C9orf24*	*CEP135*	*DNAI2*	*FAM229B*	*HSPA1L*	*NEK4*
ATG7	*CABS1*	*CEP164*	*DNAJA1*	*FAM49B*	*HSPA1L*	*NEK8*
ATP2A2	*CABYR*	*CEP19*	*DNAJA4*	*FAM65B*	*HSPA1L*	*NEURL*
ATP2B2	*CALCR*	*CEP250*	*DNAL1*	*FAM81B*	*HSPA4L*	*NGFR*
ATP2B4	*CALML4*	*CEP290*	*DNAL4*	*FANK1*	*HSPA8*	*NIN*
ATP6V0D1	*CAPS*	*CEP41*	*DNALI1*	*FBF1*	*HSPB11*	*HTT*
ATP6V1C1	*CAPSL*	*CEP72*	*DPCD*	*FBXO15*	*HSPBP1*	*HUWE1*
ATP6V1D	*CARS*	*CEP89*	*DPYSL2*	*FLCN*	*LDHA*	*HYDIN*
ATP8A2	*CASC1*	*CEP97*	*DRD1*	*FLNA*	*LDHC*	*HYLS1*
ATXN10	*CASK*	*CETN1*	*DRD1*	*FNBP1L*	*LRBA*	*ICK*
AZI1	*CATSPER1*	*CETN2*	*DRD2*	*FOCAD*	*LRP2BP*	*IFT122*
B9D1	*CATSPER2*	*CETN3*	*DRD5*	*FOPNL*	*LRRC23*	*IFT140*
B9D2	*CATSPER3*	*CLASP1*	*DSTN*	*FOXJ1*	*LRRC34*	*IFT172*

IFT20	*LRRC43*	*NPHP3*	*PIH1D3*	*SLC26A6*	*TEKT5*	*TUBB4A*
IFT27	*LRRC46*	*NPHP4*	*PIK3C3*	*SLC27A2*	*TEX26*	*TUBB4B*
IFT43	*LRRC48*	*NPY2R*	*PIK3R4*	*SLC47A2*	*TEX9*	*TUBE1*
IFT46	*LRRC49*	*NT5C3*	*PIN1*	*SLC9A3R1*	*TMEM107*	*TUBG1*
IFT52	*LRRC6*	*NTPCR*	*PKD1*	*SLC9C1*	*TMEM138*	*TUBGCP2*
IFT57	*LRRC73*	*NUDC*	*PKD1L1*	*SLIRP*	*TMEM17*	*TUBGCP3*
IFT74	*LZTFL1*	*NUP214*	*PKD2*	*SMO*	*TMEM216*	*TUBGCP4*
IFT80	*MAATS1*	*NUP35*	*PKD2L1*	*SNAP25*	*TMEM231*	*TUBGCP5*
IFT81	*MAGI2*	*NUP37*	*PKHD1*	*SNAP29*	*TMEM232*	*TUBGCP6*
IFT88	*MAK*	*NUP62*	*PKHD1L1*	*SNTN*	*TMEM237*	*TULP1*
IGBP1	*MAL*	*NUP93*	*PKIG*	*SNX10*	*TMEM67*	*TULP2*
INHA	*MAP1A*	*NUPL2*	*PKM*	*SORD*	*TNPO1*	*TULP3*
INPP5E	*MAP1B*	*OCRL*	*PLA2G3*	*SPA17*	*TOPORS*	*TULP4*
INTU	*MAP1LC3B*	*ODF1*	*PLCB4*	*SPAG1*	*TP53BP1*	*TXNDC2*
INVS	*MAP6*	*ODF2*	*PLEKHB1*	*SPAG16*	*TPGS1*	*TXNDC8*
IPO5	*MAP9*	*ODF3*	*PLK1*	*SPAG17*	*TPPP2*	*UBE2B*
IQCA1	*MAPRE1*	*ODF4*	*POC1A*	*SPAG4*	*TPPP3*	*UBXN10*
IQCB1	*MAPRE3*	*OFD1*	*POC1B*	*SPAG6*	*TRAF3IP1*	*ULK3*
IQCD	*MAPT*	*ONECUT1*	*POR*	*SPATA17*	*TRAPPC10*	*ULK4*
IQCE	*MARK4*	*ONECUT2*	*PPEF2*	*SPATA18*	*TRAPPC3*	*UMOD*
IQCG	*MCHR1*	*OPN1LW*	*PPID*	*SPATA4*	*TRAPPC9*	*UNC119B*
IQCH	*MCIN*	*OPN1MW*	*PPP1R7*	*SPATA6*	*TRIM32*	*USH1C*
IQCK	*MDH1B*	*OPN1MW2*	*PPP2CB*	*SPATA7*	*TRIM59*	*USH1G*
IQUB	*MDM1*	*OPN1SW*	*PPP5C*	*SPEF1*	*TRIP11*	*USH2A*
IRS1	*MERTK*	*ORC1*	*PQBP1*	*SPEF2*	*TRPV4*	*VANGL2*
KATNAL1	*MKKS*	*OSBPL6*	*PRKACA*	*SPTAN1*	*TSGA10*	*VDAC3*
KATNAL2	*MKS1*	*OSCP1*	*PRKACB*	*SPTBN5*	*TSNAXIP1*	*VHL*
KCNJ16	*MLF1*	*OXCT2*	*PRKACG*	*SRGAP3*	*TSSC1*	*VHLL*
KCNRG	*MNS1*	*PABPC1L*	*PRKAR1A*	*SSNA1*	*TSSK1B*	*VPS35*
KCTD10	*MOK*	*PACRG*	*PRKAR1B*	*SSTR3*	*TTBK2*	*VWA3B*
KIAA0586	*MORN2*	*PAFAH1B1*	*PRKAR2A*	*SSX2IP*	*TTC12*	*WDPCP*
KIAA0753	*MORN3*	*PARD3*	*PRKAR2B*	*STIL*	*TTC18*	*WDR11*
KIAA1009	*MORN5*	*PARD6A*	*PRKCA*	*STK33*	*TTC21A*	*WDR16*
KIAA1377	*MYB*	*PCDH15*	*PROM1*	*STK36*	*TTC21B*	*WDR19*
KIAA1407	*MYH10*	*PCDHB13*	*PROM2*	*STK38L*	*TTC26*	*WDR34*
KIF17	*MYO15A*	*PCDHB15*	*PRPH*	*STOML3*	*TTC29*	*WDR35*
KIF19	*MYO3B*	*PCDP1*	*PRPH2*	*STRC*	*TTC30A*	*WDR47*
KIF24	*MYO5A*	*PCM1*	*PSEN1*	*STX3*	*TTC30B*	*WDR60*
KIF27	*MYO5B*	*PCNT*	*PSEN2*	*SUFU*	*TTC40*	*WDR63*
KIF2A	*MYO7A*	*PDC*	*PSMC2*	*SYNE1*	*TTC8*	*WDR66*
KIF3A	*MYOC*	*PDE1C*	*PSMC5*	*SYNE2*	*TTK*	*WDR69*
KIF3B	*MYRIP*	*PDE4C*	*PTCH1*	*TAS2R4*	*TTLL1*	*WDR78*
KIF3C	*NAPEPLD*	*PDE6A*	*PTCHD3*	*TAS2R43*	*TTLL11*	*WDR96*
KIF5B	*NBEA*	*PDE6B*	*PTGS1*	*TAS2R43*	*TTLL3*	*WRAP53*
KIF7	*NEDD1*	*PDE6D*	*PTPDC1*	*TAS2R46*	*TTLL4*	*XPNPEP3*
KIF9	*NEK1*	*PDE6G*	*PTPN23*	*TAS2R46*	*TTLL5*	*YWHAE*
KIFAP3	*NEK11*	*PDZD7*	*PTPRK*	*TBC1D30*	*TTLL6*	*YWHAQ*
KLC1	*NEK2*	*PFKM*	*RAB10*	*TBC1D7*	*TTLL7*	*ZBBX*
KLC2	*NEK4*	*PFN2*	*RAB11A*	*TBCC*	*TTLL8*	*ZMYND10*
KLC3	*NEK8*	*PGAM4*	*RAB11FIP3*	*TCEA2*	*TTLL9*	*ZMYND12*
KNCN	*NEURL*	*PGK2*	*RAB15*	*TCTEX1D2*	*TUB*	*ZNF423*
LAMA5	*NGFR*	*PHC1*	*RAB17*	*TCTEX1D4*	*TUBA1A*	*ZNF474*
LCA5	*NIN*	*PHF17*	*RAB23*	*TCTN1*	*TUBA1B*	*ZSCAN18*
LDHA	*NINL*	*PHLPP2*	*RAB27A*	*TCTN2*	*TUBA1C*	
LDHC	*NME5*	*PHTF1*	*RAB28*	*TCTN3*	*TUBA3D*	
LRBA	*NME7*	*PIBF1*	*SKP1*	*TEKT1*	*TUBA4A*	
LRP2BP	*NME8*	*PIFO*	*SLC22A4*	*TEKT2*	*TUBB2A*	
LRRC23	*NOTO*	*PIGS*	*SLC25A31*	*TEKT3*	*TUBB2B*	
LRRC34	*NPHP1*	*PIH1D2*	*SLC26A3*	*TEKT4*	*TUBB3*	

Table A2.4 Cilia-associated gene list filter, from www.omictools.com/ciliacarta-tool

Supplementary Information

All supplementary information is available on the supplied CD-ROM. Files are listed in Table S1 below.

Supplementary Information 1- Database		
S1.1	HIGH-5 phenotyping database with anonymised patient information	
S1.2	Empty HIGH-5 phenotyping database	
S1.3	Human Phenotype Ontology terms for cohort	
Supplementary Information 2- Pharmacogenomics		
S2.1	Actionable pharmacogenomics genes with references	
S2.2	WGS haplotype calls	
S2.3	Prescribing advice	
	S2.3.1	Long-form guidance
	S2.3.2	Short-form guidance
S2.4	SNP genotyping validation	

Table S1 Supplementary data available on CD-ROM

Bibliography

1. Tansey EM, Wellcome Trust Centre for the History of Medicine at UCLWS, Reynolds LA, Harper PS, Wellcome Trust Centre for the History of Medicine at UCL, Wellcome T. Clinical genetics in Britain : origins and development : the transcript of a Witness Seminar held by the Wellcome Trust Centre for the History of Medicine at UCL, London, on 23 September 2008. London: Wellcome Trust Centre for the History of Medicine at UCL; 2010. xx, 146 pages : illustrations, portraits ; 24 cm. p.

2. Turner SA, Rao SK, Morgan RH, Vnencak-Jones CL, Wiesner GL. The impact of variant classification on the clinical management of hereditary cancer syndromes. Genet Med. 2018.

3. Nykamp K, Anderson M, Powers M, Garcia J, Herrera B, Ho YY, et al. Sherloc: a comprehensive refinement of the ACMG-AMP variant classification criteria. Genet Med. 2017;19(10):1105-17.

4. Amendola LM, Jarvik GP, Leo MC, McLaughlin HM, Akkari Y, Amaral MD, et al. Performance of ACMG-AMP Variant-Interpretation Guidelines among Nine Laboratories in the Clinical Sequencing Exploratory Research Consortium. American journal of human genetics. 2016;98(6):1067-76.

5. Green RC, Berg JS, Grody WW, Kalia SS, Korf BR, Martin CL, et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. Genet Med. 2013;15(7):565-74.

6. Redekop WK, Mladsi D. The faces of personalized medicine: a framework for understanding its meaning and scope. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2013;16(6 Suppl):S4-9.

7. Eriksson N, Macpherson JM, Tung JY, Hon LS, Naughton B, Saxonov S, et al. Web-based, participant-driven studies yield novel genetic associations for common traits. PLoS Genet. 2010;6(6):e1000993.

8. Higdon R, Earl RK, Stanberry L, Hudac CM, Montague E, Stewart E, et al. The promise of multi-omics and clinical data integration to identify and target personalized healthcare approaches in autism spectrum disorders. OMICS. 2015;19(4):197-208.

9. Chen R, Mias GI, Li-Pook-Than J, Jiang L, Lam HY, Chen R, et al. Personal omics profiling reveals dynamic molecular and medical phenotypes. Cell. 2012;148(6):1293-307.

10. Karaosmanoglu K, Sayar NA, Kurnaz IA, Akbulut BS. Assessment of berberine as a multi-target antimicrobial: a multi-omics study for drug discovery and repositioning. OMICS. 2014;18(1):42-53.

11. Currais A, Goldberg J, Farrokhi C, Chang M, Prior M, Dargusch R, et al. A comprehensive multiomics approach toward understanding the relationship between aging and dementia. Aging (Albany NY). 2015;7(11):937-55.

12. Nishigori M, Yagi H, Mochiduki A, Minamino N. Multiomics approach to identify novel biomarkers for dilated cardiomyopathy: Proteome and transcriptome analyses of 4C30 dilated cardiomyopathy mouse model. Biopolymers. 2016;106(4):491-502.

13. Tebani A, Afonso C, Marret S, Bekri S. Omics-Based Strategies in Precision Medicine: Toward a Paradigm Shift in Inborn Errors of Metabolism Investigations. Int J Mol Sci. 2016;17(9).

14. Union E. Decision No 1295/1999/EC of the European Parliament and of the Council of 29 April 1999 adopting a programme of Community action on rare diseases within the framework for action in the field of public health (1999 to 2003) (OJ L 155, 22.6.1999, p. 1).

Decision repealed by Decision No 1786/2002/EC (OJ L 271, 9.10.2002, p. 1). 1999.

15. Union E. COUNCIL RECOMMENDATION of 8 June 2009 on an action in the field of rare diseases (2009/C 151/02). 2009.

16. Rare Disease UK. Rare Disease in the UK 2016 [Available from: http://www.raredisease.org.uk/what-is-a-rare-disease/.

17. Angelis A, Tordrup D, Kanavos P. Socio-economic burden of rare diseases: A systematic review of cost of illness evidence. Health Policy. 2015;119(7):964-79.

18. Forsythe E, Beales PL. Bardet-Biedl Syndrome. In: Pagon RA, Adam MP, Ardinger HH, Wallace SE, Amemiya A, Bean LJH, et al., editors. GeneReviews(R). Seattle (WA)1993.

19. Zaki MS, Sattar S, Massoudi RA, Gleeson JG. Co-occurrence of distinct ciliopathy diseases in single families suggests genetic modifiers. Am J Med Genet A. 2011;155A(12):3042-9.

20. Williams CL, McIntyre JC, Norris SR, Jenkins PM, Zhang L, Pei Q, et al. Direct evidence for BBSome-associated intraflagellar transport reveals distinct properties of native mammalian cilia. Nat Commun. 2014;5:5813.

21. Khan SA, Muhammad N, Khan MA, Kamal A, Rehman ZU, Khan S. Genetics of human Bardet-Biedl syndrome, an updates. Clin Genet. 2016;90(1):3-15.

22. Seo S, Mullins RF, Dumitrescu AV, Bhattarai S, Gratie D, Wang K, et al. Subretinal gene therapy of mice with Bardet-Biedl syndrome type 1. Investigative ophthalmology & visual science. 2013;54(9):6118-32.

23. Meyer A, Meyer N, Schaeffer M, Gottenberg JE, Geny B, Sibilia J. Incidence and prevalence of inflammatory myopathies: a systematic review. Rheumatology (Oxford). 2015;54(1):50-63.

24. Enders FB, Bader-Meunier B, Baildam E, Constantin T, Dolezalova P, Feldman BM, et al. Consensus-based recommendations for the management of juvenile dermatomyositis. Ann Rheum Dis. 2016.

25. Gowdie PJ, Allen RC, Kornberg AJ, Akikusa JD. Clinical features and disease course of patients with juvenile dermatomyositis. Int J Rheum Dis. 2013;16(5):561-7.

26. Bohan A, Peter JB. Polymyositis and dermatomyositis (first of two parts). N Engl J Med. 1975;292(7):344-7.

27. Bohan A, Peter JB. Polymyositis and dermatomyositis (second of two parts). N Engl J Med. 1975;292(8):403-7.

28. Tollisen A, Sanner H, Flato B, Wahl AK. Quality of life in adults with juvenile-onset dermatomyositis: a case-control study. Arthritis Care Res (Hoboken). 2012;64(7):1020-7.

29. Rothwell S, Cooper RG, Lundberg IE, Miller FW, Gregersen PK, Bowes J, et al. Dense genotyping of immune-related loci in idiopathic inflammatory myopathies confirms HLA alleles as the strongest genetic risk factor and suggests different genetic background for major clinical subgroups. Ann Rheum Dis. 2016;75(8):1558-66.

30. Deakin CT, Yasin SA, Simou S, Arnold KA, Tansley SL, Betteridge ZE, et al. Muscle Biopsy Findings in Combination With Myositis-Specific Autoantibodies Aid Prediction of Outcomes in Juvenile Dermatomyositis. Arthritis Rheumatol. 2016;68(11):2806-16.

31. Shah M, Targoff IN, Rice MM, Miller FW, Rider LG, Childhood Myositis Heterogeneity Collaborative Study G. Brief report: ultraviolet radiation exposure is associated with clinical and autoantibody phenotypes in juvenile myositis. Arthritis Rheum. 2013;65(7):1934-41.

32. Orione MA, Silva CA, Sallum AM, Campos LM, Omori CH, Braga AL, et al. Risk factors for juvenile dermatomyositis: exposure to tobacco and air pollutants during pregnancy. Arthritis Care Res (Hoboken). 2014;66(10):1571-5.

33. Saal HM. Russell-Silver Syndrome. In: Pagon RA, Adam MP, Ardinger HH, Wallace SE, Amemiya A, Bean LJH, et al., editors. GeneReviews(R). Seattle (WA)1993.

34. Giabicani E, Netchine I, Brioude F. New clinical and molecular insights into Silver-Russell syndrome. Curr Opin Pediatr. 2016;28(4):529-35.

35. Wakeling EL, Brioude F, Lokulo-Sodipe O, O'Connell SM, Salem J, Bliek J, et al. Diagnosis and management of Silver-Russell syndrome: first international consensus statement. Nat Rev Endocrinol. 2017;13(2):105-24.

36. Azzi S, Salem J, Thibaud N, Chantot-Bastaraud S, Lieber E, Netchine I, et al. A prospective study validating a clinical scoring system and demonstrating phenotypical-genotypical correlations in Silver-Russell syndrome. J Med Genet. 2015;52(7):446-53.

37. Netchine I, Rossignol S, Dufourg MN, Azzi S, Rousseau A, Perin L, et al. 11p15 imprinting center region 1 loss of methylation is a common and specific cause of typical Russell-Silver syndrome: clinical scoring system and epigenetic-phenotypic correlations. J Clin Endocrinol Metab. 2007;92(8):3148-54.

38. Penaherrera MS, Weindler S, Van Allen MI, Yong SL, Metzger DL, McGillivray B, et al. Methylation profiling in individuals with Russell-Silver syndrome. Am J Med Genet A. 2010;152A(2):347-55.

39. Eggermann T, Begemann M, Binder G, Spengler S. Silver-Russell syndrome: genetic basis and molecular genetic testing. Orphanet J Rare Dis. 2010;5:19.

40. Albanese A, Stanhope R. GH treatment induces sustained catch-up growth in children with intrauterine growth retardation: 7-year results. Horm Res. 1997;48(4):173-7.

41. Sorusch N, Wunderlich K, Bauss K, Nagel-Wolfrum K, Wolfrum U. Usher syndrome protein network functions in the retina and their relation to other retinal ciliopathies. Adv Exp Med Biol. 2014;801:527-33.

42. Saihan Z, Webster AR, Luxon L, Bitner-Glindzicz M. Update on Usher syndrome. Curr Opin Neurol. 2009;22(1):19-27.

43. Lentz J, Keats BJB. Usher Syndrome Type I. In: Pagon RA, Adam MP, Ardinger HH, Wallace SE, Amemiya A, Bean LJH, et al., editors. GeneReviews(R). Seattle (WA)1993.

44. Aparisi MJ, Aller E, Fuster-Garcia C, Garcia-Garcia G, Rodrigo R, Vazquez-Manrique RP, et al. Targeted next generation sequencing for molecular diagnosis of Usher syndrome. Orphanet J Rare Dis. 2014;9:168.

45. Yan D, Liu XZ. Genetics and pathological mechanisms of Usher syndrome. J Hum Genet. 2010;55(6):327-35.

46. Joensuu T, Hamalainen R, Yuan B, Johnson C, Tegelberg S, Gasparini P, et al. Mutations in a novel gene with transmembrane domains underlie Usher syndrome type 3. American journal of human genetics. 2001;69(4):673-84.

47. Kelsen J, Baldassano RN. Inflammatory bowel disease: the difference between children and adults. Inflamm Bowel Dis. 2008;14 Suppl 2:S9-11.

48. Levine A, Griffiths A, Markowitz J, Wilson DC, Turner D, Russell RK, et al. Pediatric modification of the Montreal classification for inflammatory bowel disease: the Paris classification. Inflamm Bowel Dis. 2011;17(6):1314-21.

49. Uhlig HH, Schwerd T, Koletzko S, Shah N, Kammermeier J, Elkadri A, et al. The diagnostic approach to monogenic very early onset inflammatory bowel disease. Gastroenterology. 2014;147(5):990-1007 e3.

50. Benchimol EI, Mack DR, Nguyen GC, Snapper SB, Li W, Mojaverian N, et al. Incidence, outcomes, and health services burden of very early onset inflammatory bowel disease. Gastroenterology. 2014;147(4):803-13 e7; quiz e14-5.

51. Virta LJ, Saarinen MM, Kolho KL. Inflammatory Bowel Disease Incidence is on the Continuous Rise Among All Paediatric Patients Except for the Very Young: A Nationwide Registry-based Study on 28-Year Follow-up. J Crohns Colitis. 2017;11(2):150-6.

52. Prenzel F, Uhlig HH. Frequency of indeterminate colitis in children and adults with IBD - a metaanalysis. J Crohns Colitis. 2009;3(4):277-81.

53. Sawczenko A, Sandhu BK. Presenting features of inflammatory bowel disease in Great Britain and Ireland. Arch Dis Child. 2003;88(11):995-1000.

54. Moeeni V, Day AS. Impact of Inflammatory Bowel Disease upon Growth in Children and Adolescents. ISRN Pediatr. 2011;2011:365712.

55. Van Limbergen J, Russell RK, Drummond HE, Aldhous MC, Round NK, Nimmo ER, et al. Definition of phenotypic characteristics of childhood-onset inflammatory bowel disease. Gastroenterology. 2008;135(4):1114-22.

56. Kammermeier J, Dziubak R, Pescarin M, Drury S, Godwin H, Reeve K, et al. Phenotypic and genotypic characterisation of inflammatory bowel disease presenting before the age of 2 years. J Crohns Colitis. 2016.

57. Moran CJ. Very early onset inflammatory bowel disease. Semin Pediatr Surg. 2017;26(6):356-9.

58. Baldwin KR, Kaplan JL. Medical management of pediatric inflammatory bowel disease. Semin Pediatr Surg. 2017;26(6):360-6.

59. Kotlarz D, Beier R, Murugan D, Diestelhorst J, Jensen O, Boztug K, et al. Loss of interleukin-10 signaling and infantile inflammatory bowel disease: implications for diagnosis and therapy. Gastroenterology. 2012;143(2):347-55.

60. Hoskins B FE. Personal communication- BBS testing strategy in the UK, 2017. In: Kenny J, editor. 2017.

61. Farmer A, Ayme S, de Heredia ML, Maffei P, McCafferty S, Mlynarski W, et al. EURO-WABB: an EU rare diseases registry for Wolfram syndrome, Alstrom syndrome and Bardet-Biedl syndrome. BMC Pediatr. 2013;13:130.

62. Ip SC, Lin SW, Lai KM. An evaluation of the performance of five extraction methods: Chelex(R) 100, QIAamp(R) DNA Blood Mini Kit, QIAamp(R) DNA Investigator Kit, QIAsymphony(R) DNA Investigator(R) Kit and DNA IQ. Sci Justice. 2015;55(3):200-8.

63. Desjardins P, Conklin D. NanoDrop microvolume quantitation of nucleic acids. J Vis Exp. 2010(45).

64. Cunningham F, Amode MR, Barrell D, Beal K, Billis K, Brent S, et al. Ensembl 2015. Nucleic Acids Res. 2015;43(Database issue):D662-9.

65. Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JA. Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Res. 2007;35(Web Server issue):W71-4.

66. Rosenbloom KR, Armstrong J, Barber GP, Casper J, Clawson H, Diekhans M, et al. The UCSC Genome Browser database: 2015 update. Nucleic Acids Res. 2015;43(Database issue):D670-81.

67. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010;26(5):589-95.

68. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-9.

69. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9):1297-303.

70. BROAD Institute Picard Tools 2018 [Available from: http://broadinstitute.github.io/picard/.

71. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013;14(2):178-92.

72. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29(1):24-6.

73. Medicine M-NIoG. Online Mendelian Inheritance in Man, OMIM®. Johns Hopkins University (Baltimore, MD); 2018 [Available from: https://omim.org/.

74. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57-74.

75. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, et al. Accurate whole human genome sequencing using reversible terminator chemistry. Nature. 2008;456(7218):53-9.

76. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. Nature. 2015;526(7571):68-74.

77. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016;536(7616):285-91.

78. (ESP) NGESP. Exome Variant Server Seattle, WA: NHLBI GO Exome Sequencing Project (ESP); 2018 [Available from: http://evs.gs.washington.edu/EVS/.

79. Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, et al. ClinVar: public archive of interpretations of clinically relevant variants. Nucleic Acids Res. 2016;44(D1):D862-8.

80. Stenson PD, Mort M, Ball EV, Shaw K, Phillips A, Cooper DN. The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. Human genetics. 2014;133(1):1-9.

81. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet. 2013;Chapter 7:Unit7 20.

82. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res. 2003;31(13):3812-4.

83. Twist GP, Gaedigk A, Miller NA, Farrow EG, Willig LK, Dinwiddie DL, et al. Constellation: a tool for rapid, automated phenotype assignment of a highly polymorphic pharmacogene, CYP2D6, from whole-genome sequences. Npj Genomic Medicine. 2017;2:16039.

84. Layer RM, Chiang C, Quinlan AR, Hall IM. LUMPY: a probabilistic framework for structural variant discovery. Genome Biol. 2014;15(6):R84.

85. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015;17(5):405-24.

86. Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, et al. STRING 8--a global view on proteins and their functional interactions in 630 organisms. Nucleic Acids Res. 2009;37(Database issue):D412-6.

87. Bank PC, Caudle KE, Swen JJ, Gammal RS, Whirl-Carrillo M, Klein TE, et al. Comparison of the Guidelines of the Clinical Pharmacogenetics Implementation Consortium and the Dutch Pharmacogenetics Working Group. Clin Pharmacol Ther. 2017.

88. Shostak S, Zarhin D, Ottman R. What's at stake? Genetic information from the perspective of people with epilepsy and their family members. Soc Sci Med. 2011;73(5):645-54.

89. Basel D, McCarrier J. Ending a Diagnostic Odyssey: Family Education, Counseling, and Response to Eventual Diagnosis. Pediatr Clin North Am. 2017;64(1):265-72.

90. Roberts JS, LaRusse SA, Katzen H, Whitehouse PJ, Barber M, Post SG, et al. Reasons for seeking genetic susceptibility testing among first-degree relatives of people with Alzheimer disease. Alzheimer Dis Assoc Disord. 2003;17(2):86-93.

91. Boycott KM, Vanstone MR, Bulman DE, MacKenzie AE. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. Nat Rev Genet. 2013;14(10):681-91.

92. Hayward J, Chitty LS. Beyond screening for chromosomal abnormalities: Advances in non-invasive diagnosis of single gene disorders and fetal exome sequencing. Semin Fetal Neonatal Med. 2018.

93. UK RD. Experiences of rare diseases: an insight from patients and families [Report of Survey]. 2010 [Available from: https://www.raredisease.org.uk/media/1594/rduk-family-report.pdf.

94. Carmichael N, Tsipis J, Windmueller G, Mandel L, Estrella E. "Is it going to hurt?": the impact of the diagnostic odyssey on children and their families. J Genet Couns. 2015;24(2):325-35.

95. Lazaridis KN, Schahl KA, Cousin MA, Babovic-Vuksanovic D, Riegert-Johnson DL, Gavrilova RH, et al. Outcome of Whole Exome Sequencing for Diagnostic Odyssey Cases of an Individualized Medicine Clinic: The Mayo Clinic Experience. Mayo Clin Proc. 2016;91(3):297-307.

96. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A. 1977;74(12):5463-7.

97. Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, Connell CR, et al. Fluorescence detection in automated DNA sequence analysis. Nature. 1986;321(6071):674-9.

98. LaFramboise T. Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances. Nucleic Acids Res. 2009;37(13):4181-93.

99. Iacobucci I, Lonetti A, Papayannidis C, Martinelli G. Use of single nucleotide polymorphism array technology to improve the identification of chromosomal lesions in leukemia. Curr Cancer Drug Targets. 2013;13(7):791-810.

100. D'Amours G, Langlois M, Mathonnet G, Fetni R, Nizard S, Srour M, et al. SNP arrays: comparing diagnostic yields for four platforms in children with developmental delay. BMC Med Genomics. 2014;7:70.

101. Lohmann K, Klein C. Next generation sequencing and the future of genetic diagnosis. Neurotherapeutics. 2014;11(4):699-707.

102. NIHGRI. National Institute of Genome Research- The Cost of Sequencing a Human Genome: National Institute of Health; 2016 [updated 06/07/2016. Available from: https://www.genome.gov/sequencingcostsdata/.

103. Buhr S. Illumina wants to sequence your genome for $100 2017 [updated 10/1/2017. Available from: https://techcrunch.com/2017/01/10/illumina-wants-to-sequence-your-whole-genome-for-100/.

104. Judkins T, Leclair B, Bowles K, Gutin N, Trost J, McCulloch J, et al. Development and analytical validation of a 25-gene next generation sequencing panel that includes the BRCA1 and BRCA2 genes to assess hereditary cancer risk. BMC Cancer. 2015;15:215.

105. Moller RS, Larsen LH, Johannesen KM, Talvik I, Talvik T, Vaher U, et al. Gene Panel Testing in Epileptic Encephalopathies and Familial Epilepsies. Mol Syndromol. 2016;7(4):210-9.

106. Beck J, Pittman A, Adamson G, Campbell T, Kenny J, Houlden H, et al. Validation of next-generation sequencing technologies in genetic diagnosis of dementia. Neurobiol Aging. 2014;35(1):261-5.

107. Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, et al. Exome sequencing identifies the cause of a mendelian disorder. Nat Genet. 2010;42(1):30-5.

108. Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, et al. Targeted capture and massively parallel sequencing of 12 human exomes. Nature. 2009;461(7261):272-6.

109. Clarke AJ. Managing the ethical challenges of next-generation sequencing in genomic medicine. Br Med Bull. 2014;111(1):17-30.

110. Clift KE, Halverson CM, Fiksdal AS, Kumbamu A, Sharp RR, McCormick JB. Patients' views on incidental findings from clinical exome sequencing. Appl Transl Genom. 2015;4:38-43.

111. Dorschner MO, Amendola LM, Turner EH, Robertson PD, Shirts BH, Gallego CJ, et al. Actionable, pathogenic incidental findings in 1,000 participants' exomes. American journal of human genetics. 2013;93(4):631-40.

112. Shahmirzadi L, Chao EC, Palmaer E, Parra MC, Tang S, Gonzalez KD. Patient decisions for disclosure of secondary findings among the first 200 individuals undergoing clinical diagnostic exome sequencing. Genet Med. 2014;16(5):395-9.

113. Hicks JK, Shealy A, Schreiber A, Coleridge M, Noss R, Natowicz M, et al. Patient Decisions to Receive Secondary Pharmacogenomic Findings and Development of a Multidisciplinary Practice Model to Integrate Results Into Patient Care. Clin Transl Sci. 2018;11(1):71-6.

114. Daack-Hirsch S, Driessnack M, Hanish A, Johnson VA, Shah LL, Simon CM, et al. 'Information is information': a public perspective on incidental findings in clinical and research genome-based testing. Clin Genet. 2013;84(1):11-8.

115. Belkadi A, Bolze A, Itan Y, Cobat A, Vincent QB, Antipenko A, et al. Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. Proc Natl Acad Sci U S A. 2015;112(17):5473-8.

116. Meynert AM, Ansari M, FitzPatrick DR, Taylor MS. Variant detection sensitivity and biases in whole genome and exome sequencing. BMC Bioinformatics. 2014;15:247.

117. Gilissen C, Hehir-Kwa JY, Thung DT, van de Vorst M, van Bon BW, Willemsen MH, et al. Genome sequencing identifies major causes of severe intellectual disability. Nature. 2014;511(7509):344-7.

118. Lelieveld SH, Spielmann M, Mundlos S, Veltman JA, Gilissen C. Comparison of Exome and Genome Sequencing Technologies for the Complete Capture of Protein-Coding Regions. Human mutation. 2015;36(8):815-22.

119. Campuzano O, Allegue C, Fernandez A, Iglesias A, Brugada R. Determining the pathogenicity of genetic variants associated with cardiac channelopathies. Sci Rep. 2015;5:7953.

120. Duzkale H, Shen J, McLaughlin H, Alfares A, Kelly MA, Pugh TJ, et al. A systematic approach to assessing the clinical significance of genetic variants. Clin Genet. 2013;84(5):453-63.

121. Izarzugaza JM, del Pozo A, Vazquez M, Valencia A. Prioritization of pathogenic mutations in the protein kinase superfamily. BMC Genomics. 2012;13 Suppl 4:S3.

122. Plon SE, Eccles DM, Easton D, Foulkes WD, Genuardi M, Greenblatt MS, et al. Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. Human mutation. 2008;29(11):1282-91.

123. Pepin MG, Murray ML, Bailey S, Leistritz-Kessler D, Schwarze U, Byers PH. The challenge of comprehensive and consistent sequence variant interpretation between clinical laboratories. Genet Med. 2016;18(1):20-4.

124. Sobreira N, Schiettecatte F, Valle D, Hamosh A. GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. Human mutation. 2015;36(10):928-30.

125. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010;7(4):248-9.

126. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc. 2009;4(7):1073-81.

127. Thusberg J, Olatubosun A, Vihinen M. Performance of mutation pathogenicity prediction methods on missense variants. Human mutation. 2011;32(4):358-68.

128. Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic Acids Res. 2009;37(9):e67.

129. Faa V, Coiana A, Incani F, Costantino L, Cao A, Rosatelli MC. A synonymous mutation in the CFTR gene causes aberrant splicing in an italian patient affected by a mild form of cystic fibrosis. J Mol Diagn. 2010;12(3):380-3.

130. Hunt RC, Simhadri VL, Iandoli M, Sauna ZE, Kimchi-Sarfaty C. Exposing synonymous mutations. Trends Genet. 2014;30(7):308-21.

131. Sauna ZE, Kimchi-Sarfaty C. Understanding the contribution of synonymous mutations to human disease. Nat Rev Genet. 2011;12(10):683-91.

132. Stefl S, Nishi H, Petukh M, Panchenko AR, Alexov E. Molecular mechanisms of disease-causing missense mutations. J Mol Biol. 2013;425(21):3919-36.

133. Guerreiro R, Bras J, Wojtas A, Rademakers R, Hardy J, Graff-Radford N. Nonsense mutation in PRNP associated with clinical Alzheimer's disease. Neurobiol Aging. 2014;35(11):2656 e13- e16.

134. Nagasaki K, Nishimura G, Kikuchi T, Nyuzuki H, Sasaki S, Ogawa Y, et al. Nonsense mutations in FZD2 cause autosomal-dominant omodysplasia: Robinow syndrome-like phenotypes. Am J Med Genet A. 2018.

135. Bai J, Qu Y, Cao Y, Yang L, Ge L, Jin Y, et al. The SMN1 common variant c.22 dupA in Chinese patients causes spinal muscular atrophy by nonsense-mediated mRNA decay in humans. Gene. 2018;644:49-55.

136. Shahbazi S, Baniahmad F, Zakiani-Roudsari M, Raigani M, Mahdian R. Nonsense mediated decay of VWF mRNA subsequent to c.7674-7675insC mutation in type3 VWD patients. Blood Cells Mol Dis. 2012;49(1):48-52.

137. Forster L, Ardakani RM, Qadah T, Finlayson J, Ghassemifar R. The Effect of Nonsense Mediated Decay on Transcriptional Activity Within the Novel beta-Thalassemia Mutation HBB: c.129delT. Hemoglobin. 2015;39(5):334-9.

138. Apellaniz-Ruiz M, de Kock L, Sabbaghian N, Guaraldi F, Ghizzoni L, Beccuti G, et al. Familial multinodular goiter and Sertoli-Leydig cell tumors associated with a large intragenic in-frame DICER1 deletion. Eur J Endocrinol. 2018;178(2):K11-K9.

139. Mall M, Kreda SM, Mengos A, Jensen TJ, Hirtz S, Seydewitz HH, et al. The DeltaF508 mutation results in loss of CFTR function and mature protein in native human colon. Gastroenterology. 2004;126(1):32-41.

140. den Dunnen JT, Antonarakis SE. Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. Human mutation. 2000;15(1):7-12.

141. den Dunnen JT, Dalgleish R, Maglott DR, Hart RK, Greenblatt MS, McGowan-Jordan J, et al. HGVS Recommendations for the Description of Sequence Variants: 2016 Update. Human mutation. 2016;37(6):564-9.

142. Mitchison HM, Valente EM. Motile and non-motile cilia in human pathology: from function to phenotypes. J Pathol. 2017;241(2):294-309.

143. Praetorius HA, Spring KR. A physiological view of the primary cilium. Annu Rev Physiol. 2005;67:515-29.

144. Venkatesh D. Primary cilia. J Oral Maxillofac Pathol. 2017;21(1):8-10.

145. Reiter JF, Blacque OE, Leroux MR. The base of the cilium: roles for transition fibres and the transition zone in ciliary formation, maintenance and compartmentalization. EMBO Rep. 2012;13(7):608-18.

146. Sasai N, Briscoe J. Primary cilia and graded Sonic Hedgehog signaling. Wiley Interdiscip Rev Dev Biol. 2012;1(5):753-72.

147. Abou Alaiwi WA, Lo ST, Nauli SM. Primary cilia: highly sophisticated biological sensors. Sensors (Basel). 2009;9(9):7003-20.

148. Deane JA, Cole DG, Seeley ES, Diener DR, Rosenbaum JL. Localization of intraflagellar transport protein IFT52 identifies basal body transitional fibers as the docking site for IFT particles. Curr Biol. 2001;11(20):1586-90.

149. Nachury MV, Loktev AV, Zhang Q, Westlake CJ, Peranen J, Merdes A, et al. A core complex of BBS proteins cooperates with the GTPase Rab8 to promote ciliary membrane biogenesis. Cell. 2007;129(6):1201-13.

150. Xu Q, Zhang Y, Wei Q, Huang Y, Li Y, Ling K, et al. BBS4 and BBS5 show functional redundancy in the BBSome to regulate the degradative sorting of ciliary sensory receptors. Sci Rep. 2015;5:11855.

151. Loktev AV, Zhang Q, Beck JS, Searby CC, Scheetz TE, Bazan JF, et al. A BBSome subunit links ciliogenesis, microtubule stability, and acetylation. Dev Cell. 2008;15(6):854-65.

152. Wei Q, Zhang Y, Li Y, Zhang Q, Ling K, Hu J. The BBSome controls IFT assembly and turnaround in cilia. Nat Cell Biol. 2012;14(9):950-7.

153. van Dam TK, J.; van der Lee, R.; de Vrieze, E.; Wunderlich, K.; Rix, S.; Dougherty, G.; Lambacher, N.; Li, C.; Jensen, V.; Leroux, M.; Hjeij, R.; Horn, N.; Texier, Y.; Wissinger, Y.; van Reeuwijk, J.; Wheway, G.; Knapp,B.; Scheel, J.; Franco, B.; Mans, D.; van Wijk, E.; Képès, F.; Slaats, G.; Toedt, G.; Kremer, H.; Omran, H.; Szymanska, K.; Koutroumpas, K.; Ueffing, M.; Nguyen, T.; Letteboer, S.; Oug, M.; van Beersum, S.; Schmidts, M.; Beales, P.; Lu, Q.; Giles, R.; Szklarczyk, R.; Russell, R.; Gibson, T.; Johnson, C.; Blacque, O.; Wolfrum,U.; Boldt, K.; Roepman, R.; Hernandez-Hernandez, V.; Huynen, M. CiliaCarta: An Integrated And Validated Compendium Of Ciliary Genes. 2017.

154. Waters AM, Beales PL. Ciliopathies: an expanding disease spectrum. Pediatr Nephrol. 2011;26(7):1039-56.

155. Huangfu D, Liu A, Rakeman AS, Murcia NS, Niswander L, Anderson KV. Hedgehog signalling in the mouse requires intraflagellar transport proteins. Nature. 2003;426(6962):83-7.

156. Choudhry Z, Rikani AA, Choudhry AM, Tariq S, Zakaria F, Asghar MW, et al. Sonic hedgehog signalling pathway: a complex network. Ann Neurosci. 2014;21(1):28-31.

157. Goetz SC, Anderson KV. The primary cilium: a signalling centre during vertebrate development. Nat Rev Genet. 2010;11(5):331-44.

158. Ross AJ, May-Simera H, Eichers ER, Kai M, Hill J, Jagger DJ, et al. Disruption of Bardet-Biedl syndrome ciliary proteins perturbs planar cell polarity in vertebrates. Nat Genet. 2005;37(10):1135-40.

159. Wallingford JB, Mitchell B. Strange as it may seem: the many links between Wnt signaling, planar cell polarity, and cilia. Genes Dev. 2011;25(3):201-13.

160. Komiya Y, Habas R. Wnt signal transduction pathways. Organogenesis. 2008;4(2):68-75.

161. Lancaster MA, Schroth J, Gleeson JG. Subcellular spatial regulation of canonical Wnt signalling at the primary cilium. Nat Cell Biol. 2011;13(6):700-7.

162. Gerdes JM, Liu Y, Zaghloul NA, Leitch CC, Lawson SS, Kato M, et al. Disruption of the basal body compromises proteasomal function and perturbs intracellular Wnt response. Nat Genet. 2007;39(11):1350-60.

163. Simons M, Gloy J, Ganner A, Bullerkotte A, Bashkurov M, Kronig C, et al. Inversin, the gene product mutated in nephronophthisis type II, functions as a molecular switch between Wnt signaling pathways. Nat Genet. 2005;37(5):537-43.

164. Eggenschwiler JT, Anderson KV. Cilia and developmental signaling. Annu Rev Cell Dev Biol. 2007;23:345-73.

165. Schou KB, Pedersen LB, Christensen ST. Ins and outs of GPCR signaling in primary cilia. EMBO Rep. 2015;16(9):1099-113.

166. Green JA, Mykytyn K. Neuronal ciliary signaling in homeostasis and disease. Cell Mol Life Sci. 2010;67(19):3287-97.

167. Praetorius HA, Spring KR. Bending the MDCK cell primary cilium increases intracellular calcium. J Membr Biol. 2001;184(1):71-9.

168. Delling M, Indzhykulian AA, Liu X, Li Y, Xie T, Corey DP, et al. Primary cilia are not calcium-responsive mechanosensors. Nature. 2016;531(7596):656-60.

169. Nag S, Resnick A. Biophysics and biofluid dynamics of primary cilia: evidence for and against the flow-sensing function. Am J Physiol Renal Physiol. 2017;313(3):F706-F20.

170. Valente EM, Dallapiccola B, Bertini E. Joubert syndrome and related disorders. Handb Clin Neurol. 2013;113:1879-88.

171. Hampshire DJ, Ayub M, Springell K, Roberts E, Jafri H, Rashid Y, et al. MORM syndrome (mental retardation, truncal obesity, retinal dystrophy and micropenis), a new autosomal recessive disorder, links to 9q34. Eur J Hum Genet. 2006;14(5):543-8.

172. Khanna H. Photoreceptor Sensory Cilium: Traversing the Ciliary Gate. Cells. 2015;4(4):674-86.

173. Tran PV, Sharma M, Li X, Calvet JP. Developmental signaling: does it bridge the gap between cilia dysfunction and renal cystogenesis? Birth Defects Res C Embryo Today. 2014;102(2):159-73.

174. Heon E, Kim G, Qin S, Garrison JE, Tavares E, Vincent A, et al. Mutations in C8ORF37 cause Bardet Biedl syndrome (BBS21). Hum Mol Genet. 2016;25(11):2283-94.

175. Schmidts M. Clinical genetics and pathobiology of ciliary chondrodysplasias. J Pediatr Genet. 2014;3(2):46-94.

176. Leitch CC, Zaghloul NA, Davis EE, Stoetzel C, Diaz-Font A, Rix S, et al. Hypomorphic mutations in syndromic encephalocele genes are associated with Bardet-Biedl syndrome. Nat Genet. 2008;40(4):443-8.

177. Helou J, Otto EA, Attanasio M, Allen SJ, Parisi MA, Glass I, et al. Mutation analysis of NPHP6/CEP290 in patients with Joubert syndrome and Senior-Loken syndrome. J Med Genet. 2007;44(10):657-63.

178. Baala L, Audollent S, Martinovic J, Ozilou C, Babron MC, Sivanandamoorthy S, et al. Pleiotropic effects of CEP290 (NPHP6) mutations extend to Meckel syndrome. American journal of human genetics. 2007;81(1):170-9.

179. den Hollander AI, Koenekoop RK, Yzer S, Lopez I, Arends ML, Voesenek KE, et al. Mutations in the CEP290 (NPHP6) gene are a frequent cause of Leber congenital amaurosis. American journal of human genetics. 2006;79(3):556-61.

180. Sayer JA, Otto EA, O'Toole JF, Nurnberg G, Kennedy MA, Becker C, et al. The centrosomal protein nephrocystin-6 is mutated in Joubert syndrome and activates transcription factor ATF4. Nat Genet. 2006;38(6):674-81.

181. Brancati F, Barrano G, Silhavy JL, Marsh SE, Travaglini L, Bielas SL, et al. CEP290 mutations are frequently identified in the oculo-renal form of Joubert syndrome-related disorders. American journal of human genetics. 2007;81(1):104-13.

182. Davis EE, Zhang Q, Liu Q, Diplas BH, Davey LM, Hartley J, et al. TTC21B contributes both causal and modifying alleles across the ciliopathy spectrum. Nat Genet. 2011;43(3):189-96.

183. Badano JL, Leitch CC, Ansley SJ, May-Simera H, Lawson S, Lewis RA, et al. Dissection of epistasis in oligogenic Bardet-Biedl syndrome. Nature. 2006;439(7074):326-30.

184. Laurence JZ, Moon RC. Four cases of "retinitis pigmentosa" occurring in the same family, and accompanied by general imperfections of development. 1866. Obes Res. 1995;3(4):400-3.

185. Bardet G. On congenital obesity syndrome with polydactyly and retinitis pigmentosa (a contribution to the study of clinical forms of hypophyseal obesity). 1920. Obes Res. 1995;3(4):387-99.

186. Biedl A. A pair of siblings with adiposo-genital dystrophy. 1922. Obes Res. 1995;3(4):404.

187. Forsythe E, Beales PL. Bardet-Biedl syndrome. Eur J Hum Genet. 2013;21(1):8-13.

188. Swiderski RE, Nishimura DY, Mullins RF, Olvera MA, Ross JL, Huang J, et al. Gene expression analysis of photoreceptor cell loss in bbs4-knockout mice reveals an early stress gene response and photoreceptor cell damage. Investigative ophthalmology & visual science. 2007;48(7):3329-40.

189. Mockel A, Perdomo Y, Stutzmann F, Letsch J, Marion V, Dollfus H. Retinal dystrophy in Bardet-Biedl syndrome and related syndromic ciliopathies. Prog Retin Eye Res. 2011;30(4):258-74.

190. Forsythe E, Sparks K, Hoskins BE, Bagkeris E, McGowan BM, Carroll PV, et al. Genetic predictors of cardiovascular morbidity in Bardet-Biedl syndrome. Clin Genet. 2015;87(4):343-9.

191. Seo S, Guo DF, Bugge K, Morgan DA, Rahmouni K, Sheffield VC. Requirement of Bardet-Biedl syndrome proteins for leptin receptor signaling. Hum Mol Genet. 2009;18(7):1323-31.

192. Guo DF, Rahmouni K. Molecular basis of the obesity associated with Bardet-Biedl syndrome. Trends Endocrinol Metab. 2011;22(7):286-93.

193. Loktev AV, Jackson PK. Neuropeptide Y family receptors traffic via the Bardet-Biedl syndrome pathway to signal in neuronal primary cilia. Cell Rep. 2013;5(5):1316-29.

194. Grarup N, Moltke I, Andersen MK, Dalby M, Vitting-Seerup K, Kern T, et al. Loss-of-function variants in ADCY3 increase risk of obesity and type 2 diabetes. Nat Genet. 2018;50(2):172-4.

195. Bishop GA, Berbari NF, Lewis J, Mykytyn K. Type III adenylyl cyclase localizes to primary cilia throughout the adult mouse brain. J Comp Neurol. 2007;505(5):562-71.

196. Forsythe E, Kenny J, Bacchelli C, Beales PL. Managing Bardet–Biedl Syndrome—Now and in the Future. Frontiers in Pediatrics. 2018;6(23).

197. Tayeh MK, Yen HJ, Beck JS, Searby CC, Westfall TA, Griesbach H, et al. Genetic interaction between Bardet-Biedl syndrome genes and implications for limb patterning. Hum Mol Genet. 2008;17(13):1956-67.

198. Zhang Q, Seo S, Bugge K, Stone EM, Sheffield VC. BBS proteins interact genetically with the IFT pathway to influence SHH-related phenotypes. Hum Mol Genet. 2012;21(9):1945-53.

199. Robert B, Lallemand Y. Anteroposterior patterning in the limb and digit specification: contribution of mouse genetics. Dev Dyn. 2006;235(9):2337-52.

200. Forsythe E, Sparks K, Best S, Borrows S, Hoskins B, Sabir A, et al. Risk Factors for Severe Renal Disease in Bardet-Biedl Syndrome. Journal of the American Society of Nephrology : JASN. 2017;28(3):963-70.

201. Haws RM, Joshi A, Shah SA, Alkandari O, Turman MA. Renal transplantation in Bardet-Biedl Syndrome. Pediatr Nephrol. 2016;31(11):2153-61.

202. Habbig S, Liebau MC. Ciliopathies - from rare inherited cystic kidney diseases to basic cellular function. Mol Cell Pediatr. 2015;2(1):8.

203. Stoler JM, Herrin JT, Holmes LB. Genital abnormalities in females with Bardet-Biedl syndrome. Am J Med Genet. 1995;55(3):276-8.

204. Mykytyn K, Mullins RF, Andrews M, Chiang AP, Swiderski RE, Yang B, et al. Bardet-Biedl syndrome type 4 (BBS4)-null mice implicate Bbs4 in flagella formation but not global cilia assembly. Proc Natl Acad Sci U S A. 2004;101(23):8664-9.

205. Lin C, Yin Y, Veith GM, Fisher AV, Long F, Ma L. Temporal and spatial dissection of Shh signaling in genital tubercle development. Development. 2009;136(23):3959-67.

206. Beales PL, Elcioglu N, Woolf AS, Parker D, Flinter FA. New criteria for improved diagnosis of Bardet-Biedl syndrome: results of a population survey. J Med Genet. 1999;36(6):437-46.

207. Breunig JJ, Sarkisian MR, Arellano JI, Morozov YM, Ayoub AE, Sojitra S, et al. Primary cilia regulate hippocampal neurogenesis by mediating sonic hedgehog signaling. Proc Natl Acad Sci U S A. 2008;105(35):13127-32.

208. Amador-Arjona A, Elliott J, Miller A, Ginbey A, Pazour GJ, Enikolopov G, et al. Primary cilia regulate proliferation of amplifying progenitors in adult hippocampus: implications for learning and memory. J Neurosci. 2011;31(27):9933-44.

209. Lindstrand A, Frangakis S, Carvalho CM, Richardson EB, McFadden KA, Willer JR, et al. Copy-Number Variation Contributes to the Mutational Load of Bardet-Biedl Syndrome. American journal of human genetics. 2016;99(2):318-36.

210. Aldahmesh MA, Li Y, Alhashem A, Anazi S, Alkuraya H, Hashem M, et al. IFT27, encoding a small GTPase component of IFT particles, is mutated in a consanguineous family with Bardet-Biedl syndrome. Hum Mol Genet. 2014;23(12):3307-15.

211. Jin H, White SR, Shida T, Schulz S, Aguiar M, Gygi SP, et al. The conserved Bardet-Biedl syndrome proteins assemble a coat that traffics membrane proteins to cilia. Cell. 2010;141(7):1208-19.

212. Mourao A, Nager AR, Nachury MV, Lorentzen E. Structural basis for membrane targeting of the BBSome by ARL6. Nat Struct Mol Biol. 2014;21(12):1035-41.

213. Seo S, Baye LM, Schulz NP, Beck JS, Zhang Q, Slusarski DC, et al. BBS6, BBS10, and BBS12 form a complex with CCT/TRiC family chaperonins and mediate BBSome assembly. Proc Natl Acad Sci U S A. 2010;107(4):1488-93.

214. Chiang AP, Beck JS, Yen HJ, Tayeh MK, Scheetz TE, Swiderski RE, et al. Homozygosity mapping with SNP arrays identifies TRIM32, an E3 ubiquitin ligase, as a Bardet-Biedl syndrome gene (BBS11). Proc Natl Acad Sci U S A. 2006;103(16):6287-92.

215. Dawe HR, Smith UM, Cullinane AR, Gerrelli D, Cox P, Badano JL, et al. The Meckel-Gruber Syndrome proteins MKS1 and meckelin interact and are required for primary cilium formation. Hum Mol Genet. 2007;16(2):173-86.

216. Barbelanne M, Hossain D, Chan DP, Peranen J, Tsang WY. Nephrocystin proteins NPHP5 and Cep290 regulate BBSome integrity, ciliary trafficking and cargo delivery. Hum Mol Genet. 2015;24(8):2185-200.

217. Coppieters F, Lefever S, Leroy BP, De Baere E. CEP290, a gene with many faces: mutation overview and presentation of CEP290base. Human mutation. 2010;31(10):1097-108.

218. Kim SK, Shindo A, Park TJ, Oh EC, Ghosh S, Gray RS, et al. Planar cell polarity acts through septins to control collective cell movement and ciliogenesis. Science (New York, NY). 2010;329(5997):1337-40.

219. Chaki M, Airik R, Ghosh AK, Giles RH, Chen R, Slaats GG, et al. Exome capture reveals ZNF423 and CEP164 mutations, linking renal ciliopathies to DNA damage response signaling. Cell. 2012;150(3):533-48.

220. Seo S, Zhang Q, Bugge K, Breslow DK, Searby CC, Nachury MV, et al. A novel protein LZTFL1 regulates ciliary trafficking of the BBSome and Smoothened. PLoS Genet. 2011;7(11):e1002358.

221. Huet D, Blisnick T, Perrot S, Bastin P. The GTPase IFT27 is involved in both anterograde and retrograde intraflagellar transport. Elife. 2014;3:e02419.

222. Liew GM, Ye F, Nager AR, Murphy JP, Lee JS, Aguiar M, et al. The intraflagellar transport protein IFT27 promotes BBSome exit from cilia through the GTPase ARL6/BBS3. Dev Cell. 2014;31(3):265-78.

223. Eguether T, San Agustin JT, Keady BT, Jonassen JA, Liang Y, Francis R, et al. IFT27 links the BBSome to IFT for maintenance of the ciliary signaling compartment. Dev Cell. 2014;31(3):279-90.

224. Bhogaraju S, Cajanek L, Fort C, Blisnick T, Weber K, Taschner M, et al. Molecular basis of tubulin transport within the cilium by IFT74 and IFT81. Science (New York, NY). 2013;341(6149):1009-12.

225. Sharif AS, Yu D, Loertscher S, Austin R, Nguyen K, Mathur PD, et al. C8ORF37 is required for photoreceptor outer segment disc morphogenesis by maintaining outer segment membrane protein homeostasis. J Neurosci. 2018.

226. Otto EA, Hurd TW, Airik R, Chaki M, Zhou W, Stoetzel C, et al. Candidate exome capture identifies mutation of SDCCAG8 as the cause of a retinal-renal ciliopathy. Nat Genet. 2010;42(10):840-50.

227. Mykytyn K, Nishimura DY, Searby CC, Beck G, Bugge K, Haines HL, et al. Evaluation of complex inheritance involving the most common Bardet-Biedl syndrome locus (BBS1). American journal of human genetics. 2003;72(2):429-37.

228. Stoetzel C, Laurier V, Davis EE, Muller J, Rix S, Badano JL, et al. BBS10 encodes a vertebrate-specific chaperonin-like protein and is a major BBS locus. Nat Genet. 2006;38(5):521-4.

229. Katsanis N, Ansley SJ, Badano JL, Eichers ER, Lewis RA, Hoskins BE, et al. Triallelic inheritance in Bardet-Biedl syndrome, a Mendelian recessive disorder. Science (New York, NY). 2001;293(5538):2256-9.

230. Beales PL, Badano JL, Ross AJ, Ansley SJ, Hoskins BE, Kirsten B, et al. Genetic interaction of BBS1 mutations with alleles at other BBS loci can result in non-Mendelian Bardet-Biedl syndrome. American journal of human genetics. 2003;72(5):1187-99.

231. Kousi M, Katsanis N. Genetic modifiers and oligogenic inheritance. Cold Spring Harb Perspect Med. 2015;5(6).

232. Boyle MP. Strategies for identifying modifier genes in cystic fibrosis. Proc Am Thorac Soc. 2007;4(1):52-7.

233. Tory K, Lacoste T, Burglen L, Moriniere V, Boddaert N, Macher MA, et al. High NPHP1 and NPHP6 mutation rate in patients with Joubert syndrome and nephronophthisis: potential epistatic effect of NPHP6 and AHI1 mutations in patients with NPHP1 mutations. Journal of the American Society of Nephrology : JASN. 2007;18(5):1566-75.

234. Schaefer E, Lauer J, Durand M, Pelletier V, Obringer C, Claussmann A, et al. Mesoaxial polydactyly is a major feature in Bardet-Biedl syndrome patients with LZTFL1 (BBS17) mutations. Clin Genet. 2014;85(5):476-81.

235. Feuillan PP, Ng D, Han JC, Sapp JC, Wetsch K, Spaulding E, et al. Patients with Bardet-Biedl syndrome have hyperleptinemia suggestive of leptin resistance. J Clin Endocrinol Metab. 2011;96(3):E528-35.

236. Daniels AB, Sandberg MA, Chen J, Weigel-DiFranco C, Fielding Hejtmancic J, Berson EL. Genotype-phenotype correlations in Bardet-Biedl syndrome. Arch Ophthalmol. 2012;130(7):901-7.

237. UKGTN UGTN. Gene sequencing for Bardet-Biedl syndrome 2018 [Available from: https://ukgtn.nhs.uk/find-a-test/search-by-disorder-gene/details/6372/.

238. van Dam TJ, Wheway G, Slaats GG, Group SS, Huynen MA, Giles RH. The SYSCILIA gold standard (SCGSv1) of known ciliary components and its applications within a systems biology consortium. Cilia. 2013;2(1):7.

239. Van Dam T, Wheway,G, Glaats, G, The SysCilia Consortium, Huynen, G, Giles, R. Syscilia gene list home page 2013 [Available from: http://www.syscilia.org/goldstandard.shtml.

240. Boldt K, van Reeuwijk J, Lu Q, Koutroumpas K, Nguyen TM, Texier Y, et al. An organelle-specific protein landscape identifies novel diseases and molecular mechanisms. Nat Commun. 2016;7:11491.

241. Coppieters F. *CEP290* Mutation Database: Centrum Medische Genetica Gent; [Available from: https://cep290base.cmgg.be/overview.php.

242. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014;46(3):310-5.

243. Jacobson SG, Cideciyan AV, Sumaroka A, Roman AJ, Charng J, Lu M, et al. Outcome Measures for Clinical Trials of Leber Congenital Amaurosis Caused by the Intronic Mutation in the CEP290 Gene. Investigative ophthalmology & visual science. 2017;58(5):2609-22.

244. Williams CL, Li C, Kida K, Inglis PN, Mohan S, Semenec L, et al. MKS and NPHP modules cooperate to establish basal body/transition zone membrane associations and ciliary gate function during ciliogenesis. J Cell Biol. 2011;192(6):1023-41.

245. Otto EA, Ramaswami G, Janssen S, Chaki M, Allen SJ, Zhou W, et al. Mutation analysis of 18 nephronophthisis associated ciliopathy disease genes using a DNA pooling and next generation sequencing strategy. J Med Genet. 2011;48(2):105-16.

246. Chaki M, Hoefele J, Allen SJ, Ramaswami G, Janssen S, Bergmann C, et al. Genotype-phenotype correlation in 440 patients with NPHP-related ciliopathies. Kidney Int. 2011;80(11):1239-45.

247. Yamada H, Shinmura K, Tsuneyoshi T, Sugimura H. Effect of splice-site polymorphisms of the TMPRSS4, NPHP4 and ORCTL4 genes on their mRNA expression. J Genet. 2005;84(2):131-6.

248. Konta T, Takasaki S, Ichikawa K, Emi M, Toriyama S, Satoh H, et al. The novel and independent association between single-point SNP of NPHP4 gene and renal function in non-diabetic Japanese population: the Takahata study. J Hum Genet. 2010;55(12):791-5.

249. Rattner A, Smallwood PM, Williams J, Cooke C, Savchenko A, Lyubarsky A, et al. A photoreceptor-specific cadherin is essential for the structural integrity of the outer segment and for photoreceptor survival. Neuron. 2001;32(5):775-86.

250. Ostergaard E, Batbayli M, Duno M, Vilhelmsen K, Rosenberg T. Mutations in PCDH21 cause autosomal recessive cone-rod dystrophy. J Med Genet. 2010;47(10):665-9.

251. Bredrup C, Saunier S, Oud MM, Fiskerstrand T, Hoischen A, Brackman D, et al. Ciliopathies with skeletal anomalies and renal insufficiency due to mutations in the IFT-A gene WDR19. American journal of human genetics. 2011;89(5):634-43.

252. Halbritter J, Porath JD, Diaz KA, Braun DA, Kohl S, Chaki M, et al. Identification of 99 novel mutations in a worldwide cohort of 1,056 patients with a nephronophthisis-related ciliopathy. Human genetics. 2013;132(8):865-84.

253. Coussa RG, Otto EA, Gee HY, Arthurs P, Ren H, Lopez I, et al. WDR19: an ancient, retrograde, intraflagellar ciliary protein is mutated in autosomal recessive retinitis pigmentosa and in Senior-Loken syndrome. Clin Genet. 2013;84(2):150-9.

254. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res. 2018;46(D1):D1062-D7.

255. Li G, Vega R, Nelms K, Gekakis N, Goodnow C, McNamara P, et al. A role for Alstrom syndrome protein, alms1, in kidney ciliogenesis and cellular quiescence. PLoS Genet. 2007;3(1):e8.

256. Marshall JD, Muller J, Collin GB, Milan G, Kingsmore SF, Dinwiddie D, et al. Alstrom Syndrome: Mutation Spectrum of ALMS1. Human mutation. 2015;36(7):660-8.

257. Sharp AM, Messiaen LM, Page G, Antignac C, Gubler MC, Onuchic LF, et al. Comprehensive genomic analysis of PKHD1 mutations in ARPKD cohorts. J Med Genet. 2005;42(4):336-49.

258. Abouelhoda M, Faquih T, El-Kalioby M, Alkuraya FS. Revisiting the morbid genome of Mendelian disorders. Genome Biol. 2016;17(1):235.

259. Bergmann C, Senderek J, Schneider F, Dornia C, Kupper F, Eggermann T, et al. PKHD1 mutations in families requesting prenatal diagnosis for autosomal recessive polycystic kidney disease (ARPKD). Human mutation. 2004;23(5):487-95.

260. Kubo A, Sasaki H, Yuba-Kubo A, Tsukita S, Shiina N. Centriolar satellites: molecular characterization, ATP-dependent movement toward centrioles and possible involvement in ciliogenesis. J Cell Biol. 1999;147(5):969-80.

261. Kim J, Krishnaswami SR, Gleeson JG. CEP290 interacts with the centriolar satellite component PCM-1 and is required for Rab8 localization to the primary cilium. Hum Mol Genet. 2008;17(23):3796-805.

262. Keryer G, Pineda JR, Liot G, Kim J, Dietrich P, Benstaali C, et al. Ciliogenesis is regulated by a huntingtin-HAP1-PCM1 pathway and is altered in Huntington disease. J Clin Invest. 2011;121(11):4372-82.

263. Stoetzel C, Laurier V, Faivre L, Megarbane A, Perrin-Schmitt F, Verloes A, et al. BBS8 is rarely mutated in a cohort of 128 Bardet-Biedl syndrome families. J Hum Genet. 2006;51(1):81-4.

264. Ansley SJ, Badano JL, Blacque OE, Hill J, Hoskins BE, Leitch CC, et al. Basal body dysfunction is a likely cause of pleiotropic Bardet-Biedl syndrome. Nature. 2003;425(6958):628-33.

265. Vaisberg EA, Grissom PM, McIntosh JR. Mammalian cells express three distinct dynein heavy chains that are localized to different cytoplasmic organelles. J Cell Biol. 1996;133(4):831-42.

266. Radmanesh F, Caglayan AO, Silhavy JL, Yilmaz C, Cantagrel V, Omar T, et al. Mutations in LAMB1 cause cobblestone brain malformation without muscular or ocular abnormalities. American journal of human genetics. 2013;92(3):468-74.

267. Tonduti D, Dorboz I, Renaldo F, Masliah-Planchon J, Elmaleh-Berges M, Dalens H, et al. Cystic leukoencephalopathy with cortical dysplasia related to LAMB1 mutations. Neurology. 2015;84(21):2195-7.

268. Zenker M, Pierson M, Jonveaux P, Reis A. Demonstration of two novel LAMB2 mutations in the original Pierson syndrome family reported 42 years ago. Am J Med Genet A. 2005;138(1):73-4.

269. Hochgreb-Hagele T, Yin C, Koo DE, Bronner ME, Stainier DY. Laminin beta1a controls distinct steps during the establishment of digestive organ laterality. Development. 2013;140(13):2734-45.

270. Katsanis SH, Jabs EW. Treacher Collins Syndrome. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJH, Stephens K, et al., editors. GeneReviews((R)). Seattle (WA)1993.

271. Marsh KL, Dixon J, Dixon MJ. Mutations in the Treacher Collins syndrome gene lead to mislocalization of the nucleolar protein treacle. Hum Mol Genet. 1998;7(11):1795-800.

272. Splendore A, Jabs EW, Passos-Bueno MR. Screening of TCOF1 in patients from different populations: confirmation of mutational hot spots and identification of a novel missense mutation that suggests an important functional domain in the protein treacle. J Med Genet. 2002;39(7):493-5.

273. Bowman M, Oldridge M, Archer C, O'Rourke A, McParland J, Brekelmans R, et al. Gross deletions in TCOF1 are a cause of Treacher-Collins-Franceschetti syndrome. Eur J Hum Genet. 2012;20(7):769-77.

274. Baum L, Chan WM, Li WY, Lam DS, Wang PB, Pang CP. ABCA4 sequence variants in Chinese patients with age-related macular degeneration or Stargardt's disease. Ophthalmologica. 2003;217(2):111-4.

275. Singh HP, Jalali S, Hejtmancik JF, Kannabiran C. Homozygous null mutations in the ABCA4 gene in two families with autosomal recessive retinal dystrophy. Am J Ophthalmol. 2006;141(5):906-13.

276. Nasonkin I, Illing M, Koehler MR, Schmid M, Molday RS, Weber BH. Mapping of the rod photoreceptor ABC transporter (ABCR) to 1p21-p22.1 and identification of novel mutations in Stargardt's disease. Human genetics. 1998;102(1):21-6.

277. Eisenberger T, Neuhaus C, Khan AO, Decker C, Preising MN, Friedburg C, et al. Increasing the yield in targeted next-generation sequencing by implicating CNV analysis, non-coding exons and the overall variant load: the example of retinal dystrophies. PloS one. 2013;8(11):e78496.

278. Jiang F, Pan Z, Xu K, Tian L, Xie Y, Zhang X, et al. Screening of ABCA4 Gene in a Chinese Cohort With Stargardt Disease or Cone-Rod Dystrophy With a Report on 85 Novel Mutations. Investigative ophthalmology & visual science. 2016;57(1):145-52.

279. Sun H, Smallwood PM, Nathans J. Biochemical defects in ABCR protein variants associated with human retinopathies. Nat Genet. 2000;26(2):242-6.

280. Zernant J, Lee W, Collison FT, Fishman GA, Sergeev YV, Schuerch K, et al. Frequent hypomorphic alleles account for a significant fraction of ABCA4 disease and distinguish it from age-related macular degeneration. J Med Genet. 2017;54(6):404-12.

281. Jacoby M, Cox JJ, Gayral S, Hampshire DJ, Ayub M, Blockmans M, et al. INPP5E mutations cause primary cilium signaling defects, ciliary instability and ciliopathies in human and mouse. Nat Genet. 2009;41(9):1027-31.

282. Bielas SL, Silhavy JL, Brancati F, Kisseleva MV, Al-Gazali L, Sztriha L, et al. Mutations in INPP5E, encoding inositol polyphosphate-5-phosphatase E, link phosphatidyl inositol signaling to the ciliopathies. Nat Genet. 2009;41(9):1032-6.

283. Humbert MC, Weihbrecht K, Searby CC, Li Y, Pope RM, Sheffield VC, et al. ARL13B, PDE6D, and CEP164 form a functional network for INPP5E ciliary targeting. Proc Natl Acad Sci U S A. 2012;109(48):19691-6.

284. Travaglini L, Brancati F, Silhavy J, Iannicelli M, Nickerson E, Elkhartoufi N, et al. Phenotypic spectrum and prevalence of INPP5E mutations in Joubert syndrome and related disorders. Eur J Hum Genet. 2013;21(10):1074-8.

285. Graser S, Stierhof YD, Lavoie SB, Gassner OS, Lamla S, Le Clech M, et al. Cep164, a novel centriole appendage protein required for primary cilium formation. J Cell Biol. 2007;179(2):321-30.

286. Sivasubramaniam S, Sun X, Pan YR, Wang S, Lee EY. Cep164 is a mediator protein required for the maintenance of genomic stability through modulation of MDC1, RPA, and CHK1. Genes Dev. 2008;22(5):587-600.

287. Onat OE, Gulsuner S, Bilguvar K, Nazli Basak A, Topaloglu H, Tan M, et al. Missense mutation in the ATPase, aminophospholipid transporter protein ATP8A2 is associated with cerebellar atrophy and quadrupedal locomotion. Eur J Hum Genet. 2013;21(3):281-5.

288. Cacciagli P, Haddad MR, Mignon-Ravix C, El-Waly B, Moncla A, Missirian C, et al. Disruption of the ATP8A2 gene in a patient with a t(10;13) de novo balanced translocation and a severe neurological phenotype. Eur J Hum Genet. 2010;18(12):1360-3.

289. Martin-Hernandez E, Rodriguez-Garcia ME, Camacho A, Matilla-Duenas A, Garcia-Silva MT, Quijada-Fraile P, et al. New ATP8A2 gene mutations associated with a novel syndrome: encephalopathy, intellectual disability, severe hypotonia, chorea and optic atrophy. Neurogenetics. 2016;17(4):259-63.

290. Chalat M, Moleschi K, Molday RS. C-terminus of the P4-ATPase ATP8A2 functions in protein folding and regulation of phospholipid flippase activity. Mol Biol Cell. 2017;28(3):452-62.

291. Lai CK, Gupta N, Wen X, Rangell L, Chih B, Peterson AS, et al. Functional characterization of putative cilia genes by high-content analysis. Mol Biol Cell. 2011;22(7):1104-19.

292. Cantagrel V, Silhavy JL, Bielas SL, Swistun D, Marsh SE, Bertrand JY, et al. Mutations in the cilia gene ARL13B lead to the classical form of Joubert syndrome. American journal of human genetics. 2008;83(2):170-9.

293. Bramswig NC, Ludecke HJ, Pettersson M, Albrecht B, Bernier RA, Cremer K, et al. Identification of new TRIP12 variants and detailed clinical evaluation of individuals with non-syndromic intellectual disability with or without autism. Human genetics. 2017;136(2):179-92.

294. O'Roak BJ, Stessman HA, Boyle EA, Witherspoon KT, Martin B, Lee C, et al. Recurrent de novo mutations implicate novel genes underlying simplex autism risk. Nat Commun. 2014;5:5595.

295. Zhang J, Gambin T, Yuan B, Szafranski P, Rosenfeld JA, Balwi MA, et al. Haploinsufficiency of the E3 ubiquitin-protein ligase gene TRIP12 causes intellectual disability with or without autism spectrum disorders, speech delay, and dysmorphic features. Human genetics. 2017;136(4):377-86.

296. Shearer RF, Saunders DN. Regulation of primary cilia formation by the ubiquitin-proteasome system. Biochem Soc Trans. 2016;44(5):1265-71.

297. Long H, Wang Q, Huang K. Ciliary/Flagellar Protein Ubiquitination. Cells. 2015;4(3):474-82.

298. Kasahara K, Kawakami Y, Kiyono T, Yonemura S, Kawamura Y, Era S, et al. Ubiquitin-proteasome system controls ciliogenesis at the initial step of axoneme extension. Nat Commun. 2014;5:5081.

299. Kim J, Lee JE, Heynen-Genel S, Suyama E, Ono K, Lee K, et al. Functional genomic screen for modulators of ciliogenesis and cilium length. Nature. 2010;464(7291):1048-51.

300. Gudjonsson T, Altmeyer M, Savic V, Toledo L, Dinant C, Grofte M, et al. TRIP12 and UBR5 suppress spreading of chromatin ubiquitylation at damaged chromosomes. Cell. 2012;150(4):697-709.

301. Schrier Vergano S, Santen G, Wieczorek D, Wollnik B, Matsumoto N, Deardorff MA. Coffin-Siris Syndrome. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJH, Stephens K, et al., editors. GeneReviews((R)). Seattle (WA)1993.

302. Hoyer J, Ekici AB, Endele S, Popp B, Zweier C, Wiesener A, et al. Haploinsufficiency of ARID1B, a member of the SWI/SNF-a chromatin-remodeling complex, is a frequent cause of intellectual disability. American journal of human genetics. 2012;90(3):565-72.

303. Yu Y, Yao R, Wang L, Fan Y, Huang X, Hirschhorn J, et al. De novo mutations in ARID1B associated with both syndromic and non-syndromic short stature. BMC Genomics. 2015;16:701.

304. Costa T, Ramsby G, Cassia F, Peters KR, Soares J, Correa J, et al. Grebe syndrome: clinical and radiographic findings in affected individuals and heterozygous carriers. Am J Med Genet. 1998;75(5):523-9.

305. Quelce-Salgado A. A rare genetic syndrome. Lancet (London, England). 1968;1(7557):1430.

306. Thomas JT, Kilpatrick MW, Lin K, Erlacher L, Lembessis P, Costa T, et al. Disruption of human limb morphogenesis by a dominant negative mutation in CDMP1. Nat Genet. 1997;17(1):58-64.

307. Travieso-Suarez L, Pereda A, Pozo-Roman J, Perez de Nanclares G, Argente J. [Brachydactyly type C due to a nonsense mutation in the GDF5 gene]. An Pediatr (Barc). 2018;88(2):107-9.

308. Lee HW, Choi J, Shin H, Kim K, Yang J, Na M, et al. Preso, a novel PSD-95-interacting FERM and PDZ domain protein that regulates dendritic spine morphogenesis. J Neurosci. 2008;28(53):14546-56.

309. Hu H, Haas SA, Chelly J, Van Esch H, Raynaud M, de Brouwer AP, et al. X-exome sequencing of 405 unresolved families identifies seven novel intellectual disability genes. Mol Psychiatry. 2016;21(1):133-48.

310. Savina-Christou S, Beales, P.L,. BBS4 null mice have reduced numbers of dendritic spines. 2016.

311. Hernandez-Hernandez V, Pravincumar P, Diaz-Font A, May-Simera H, Jenkins D, Knight M, et al. Bardet-Biedl syndrome proteins control the cilia length through regulation of actin polymerization. Hum Mol Genet. 2013;22(19):3858-68.

312. Markov AG, Aschenbach JR, Amasheh S. Claudin clusters as determinants of epithelial barrier function. IUBMB Life. 2015;67(1):29-35.

313. Heiskala M, Peterson PA, Yang Y. The roles of claudin superfamily proteins in paracellular transport. Traffic. 2001;2(2):93-8.

314. Larre I, Castillo A, Flores-Maldonado C, Contreras RG, Galvan I, Munoz-Estrada J, et al. Ouabain modulates ciliogenesis in epithelial cells. Proc Natl Acad Sci U S A. 2011;108(51):20591-6.

315. Nishiyama K, Sakaguchi H, Hu JG, Bok D, Hollyfield JG. Claudin localization in cilia of the retinal pigment epithelium. Anat Rec. 2002;267(3):196-203.

316. Konrad M, Schaller A, Seelow D, Pandey AV, Waldegger S, Lesslauer A, et al. Mutations in the tight-junction gene claudin 19 (CLDN19) are associated with renal magnesium wasting, renal failure, and severe ocular involvement. American journal of human genetics. 2006;79(5):949-57.

317. Ekinci Z, Karabas L, Konrad M. Hypomagnesemia-hypercalciuria-nephrocalcinosis and ocular findings: a new claudin-19 mutation. Turk J Pediatr. 2012;54(2):168-70.

318. Pater JA, Benteau T, Griffin A, Penney C, Stanton SG, Predham S, et al. A common variant in CLDN14 causes precipitous, prelingual sensorineural hearing loss in multiple families due to founder effect. Human genetics. 2016.

319. Bashir ZE, Latief N, Belyantseva IA, Iqbal F, Riazuddin SA, Khan SN, et al. Phenotypic variability of CLDN14 mutations causing DFNB29 hearing loss in the Pakistani population. J Hum Genet. 2013;58(2):102-8.

320. Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, et al. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. Curr Protoc Bioinformatics. 2016;54:1 30 1-1 3.

321. Marion V, Stutzmann F, Gerard M, De Melo C, Schaefer E, Claussmann A, et al. Exome sequencing identifies mutations in LZTFL1, a BBSome and smoothened trafficking regulator, in a family with Bardet--Biedl syndrome with situs inversus and insertional polydactyly. J Med Genet. 2012;49(5):317-21.

322. Li Z, Schonberg R, Guidugli L, Johnson AK, Arnovitz S, Yang S, et al. A novel mutation in the promoter of RARS2 causes pontocerebellar hypoplasia in two siblings. J Hum Genet. 2015;60(7):363-9.

323. Gotoh L, Inoue K, Helman G, Mora S, Maski K, Soul JS, et al. GJC2 promoter mutations causing Pelizaeus-Merzbacher-like disease. Mol Genet Metab. 2014;111(3):393-8.

324. Gardella R, Barlati S, Zoppi N, Tadini G, Colombi M. A -96C-->T mutation in the promoter of the collagen type VII gene (COL7A1) abolishing transcription in a patient affected by recessive dystrophic epidermolysis bullosa. Human mutation. 2000;16(3):275.

325. Papon JF, Perrault I, Coste A, Louis B, Gerard X, Hanein S, et al. Abnormal respiratory cilia in non-syndromic Leber congenital amaurosis with CEP290 mutations. J Med Genet. 2010;47(12):829-34.

326. Tsang WY, Bossard C, Khanna H, Peranen J, Swaroop A, Malhotra V, et al. CP110 suppresses primary cilia formation through its interaction with CEP290, a protein deficient in human ciliary disease. Dev Cell. 2008;15(2):187-97.

327. Gorden NT, Arts HH, Parisi MA, Coene KL, Letteboer SJ, van Beersum SE, et al. CC2D2A is mutated in Joubert syndrome and interacts with the ciliopathy-associated basal body protein CEP290. American journal of human genetics. 2008;83(5):559-71.

328. Schafer T, Putz M, Lienkamp S, Ganner A, Bergbreiter A, Ramachandran H, et al. Genetic and physical interaction between the NPHP5 and NPHP6 gene products. Hum Mol Genet. 2008;17(23):3655-62.

329. Drivas TG, Holzbaur EL, Bennett J. Disruption of CEP290 microtubule/membrane-binding domains causes retinal degeneration. J Clin Invest. 2013;123(10):4525-39.

330. Zhang Y, Seo S, Bhattarai S, Bugge K, Searby CC, Zhang Q, et al. BBS mutations modify phenotypic expression of CEP290-related ciliopathies. Hum Mol Genet. 2014;23(1):40-51.

331. Beales PL. Personal communication: Patients with *CEP290* mutations seen in the National BBS clinic. 2018.

332. Drivas TG, Wojno AP, Tucker BA, Stone EM, Bennett J. Basal exon skipping and genetic pleiotropy: A predictive model of disease pathogenesis. Science translational medicine. 2015;7(291):291ra97.

333. Doherty D, Parisi MA, Finn LS, Gunay-Aygun M, Al-Mateen M, Bates D, et al. Mutations in 3 genes (MKS3, CC2D2A and RPGRIP1L) cause COACH syndrome (Joubert syndrome with congenital hepatic fibrosis). J Med Genet. 2010;47(1):8-21.

334. Otto EA, Loeys B, Khanna H, Hellemans J, Sudbrak R, Fan S, et al. Nephrocystin-5, a ciliary IQ domain protein, is mutated in Senior-Loken syndrome and interacts with RPGR and calmodulin. Nat Genet. 2005;37(3):282-8.

335. Baala L, Romano S, Khaddour R, Saunier S, Smith UM, Audollent S, et al. The Meckel-Gruber syndrome gene, MKS3, is mutated in Joubert syndrome. American journal of human genetics. 2007;80(1):186-94.

336. Burnight ER, Wiley LA, Drack AV, Braun TA, Anfinson KR, Kaalberg EE, et al. CEP290 gene transfer rescues Leber congenital amaurosis cellular phenotype. Gene therapy. 2014;21(7):662-72.

337. Shimada H, Lu Q, Insinna-Kettenhofen C, Nagashima K, English MA, Semler EM, et al. In Vitro Modeling Using Ciliopathy-Patient-Derived Cells Reveals Distinct Cilia Dysfunctions Caused by CEP290 Mutations. Cell Rep. 2017;20(2):384-96.

338. Forsythe E. Personal communication: Appearance of cilia in cultured fibroblasts from patients with Bardet-Biedl syndrome. 2018.

339. Nakamura K, Fujii W, Tsuboi M, Tanihata J, Teramoto N, Takeuchi S, et al. Generation of muscular dystrophy model rats with a CRISPR/Cas system. Sci Rep. 2014;4:5635.

340. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex genome engineering using CRISPR/Cas systems. Science (New York, NY). 2013;339(6121):819-23.

341. Mollet G, Silbermann F, Delous M, Salomon R, Antignac C, Saunier S. Characterization of the nephrocystin/nephrocystin-4 complex and subcellular localization of nephrocystin-4 to primary cilia and centrosomes. Hum Mol Genet. 2005;14(5):645-56.

342. Mollet G, Salomon R, Gribouval O, Silbermann F, Bacq D, Landthaler G, et al. The gene mutated in juvenile nephronophthisis type 4 encodes a novel protein that interacts with nephrocystin. Nat Genet. 2002;32(2):300-5.

343. Henderson RH, Li Z, Abd El Aziz MM, Mackay DS, Eljinini MA, Zeidan M, et al. Biallelic mutation of protocadherin-21 (PCDH21) causes retinal degeneration in humans. Mol Vis. 2010;16:46-52.

344. Deltas C. Digenic inheritance and genetic modifiers. Clin Genet. 2018;93(3):429-38.

345. Schrauwen I, Chakchouk I, Acharya A, Liaqat K, Irfanullah, University of Washington Center for Mendelian G, et al. Novel digenic inheritance of PCDH15 and USH1G underlies profound non-syndromic hearing impairment. BMC Med Genet. 2018;19(1):122.

346. Wang L, Xiao Y, Tian T, Jin L, Lei Y, Finnell RH, et al. Digenic variants of planar cell polarity genes in human neural tube defect patients. Mol Genet Metab. 2018;124(1):94-100.

347. Fallerini C, Baldassarri M, Trevisson E, Morbidoni V, La Manna A, Lazzarin R, et al. Alport syndrome: impact of digenic inheritance in patients management. Clin Genet. 2017;92(1):34-44.

348. Lee JE, Silhavy JL, Zaki MS, Schroth J, Bielas SL, Marsh SE, et al. CEP41 is mutated in Joubert syndrome and is required for tubulin glutamylation at the cilium. Nat Genet. 2012;44(2):193-9.

349. Liu YP, Bosch DG, Siemiatkowska AM, Rendtorff ND, Boonstra FN, Moller C, et al. Putative digenic inheritance of heterozygous RP1L1 and C2orf71 null mutations in syndromic retinal dystrophy. Ophthalmic Genet. 2017;38(2):127-32.

350. Kikuno R, Nagase T, Ishikawa K, Hirosawa M, Miyajima N, Tanaka A, et al. Prediction of the coding sequences of unidentified human genes. XIV. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro. DNA Res. 1999;6(3):197-205.

351. Schalk A, Greff G, Drouot N, Obringer C, Dollfus H, Laugel V, et al. Deep intronic variation in splicing regulatory element of the ERCC8 gene associated with severe but long-term survival Cockayne syndrome. Eur J Hum Genet. 2018;26(4):527-36.

352. Bax NM, Sangermano R, Roosing S, Thiadens AA, Hoefsloot LH, van den Born LI, et al. Heterozygous deep-intronic variants and deletions in ABCA4 in persons with retinal dystrophies and one exonic ABCA4 variant. Human mutation. 2015;36(1):43-7.

353. Mendes de Almeida R, Tavares J, Martins S, Carvalho T, Enguita FJ, Brito D, et al. Whole gene sequencing identifies deep-intronic variants with potential functional impact in patients with hypertrophic cardiomyopathy. PloS one. 2017;12(8):e0182946.

354. Mercer TR, Clark MB, Andersen SB, Brunck ME, Haerty W, Crawford J, et al. Genome-wide discovery of human splicing branchpoints. Genome Res. 2015;25(2):290-303.

355. Pineda JMB, Bradley RK. Most human introns are recognized via multiple and tissue-specific branchpoints. Genes Dev. 2018.

356. Spielmann M, Klopocki E. CNVs of noncoding cis-regulatory elements in human disease. Curr Opin Genet Dev. 2013;23(3):249-56.

357. Smedley D, Schubach M, Jacobsen JO, Kohler S, Zemojtel T, Spielmann M, et al. A Whole-Genome Analysis Framework for Effective Identification of Pathogenic Regulatory Variants in Mendelian Disease. American journal of human genetics. 2016;99(3):595-606.

358. Ross MG, Russ C, Costello M, Hollinger A, Lennon NJ, Hegarty R, et al. Characterizing and measuring bias in sequence data. Genome Biol. 2013;14(5):R51.

359. Meienberg J, Bruggmann R, Oexle K, Matyas G. Clinical sequencing: is WGS the better WES? Human genetics. 2016;135(3):359-62.

360. Wang X, Li X, Cheng Y, Sun X, Sun X, Self S, et al. Copy number alterations detected by whole-exome and whole-genome sequencing of esophageal adenocarcinoma. Hum Genomics. 2015;9:22.

361. Hehir-Kwa JY, Pfundt R, Veltman JA. Exome sequencing and whole genome sequencing for the detection of copy number variation. Expert Rev Mol Diagn. 2015;15(8):1023-32.

362. Alkan C, Coe BP, Eichler EE. Genome structural variation discovery and genotyping. Nat Rev Genet. 2011;12(5):363-76.

363. Merker JD, Wenger AM, Sneddon T, Grove M, Zappala Z, Fresard L, et al. Long-read genome sequencing identifies causal structural variation in a Mendelian disease. Genet Med. 2018;20(1):159-63.

364. Zhao X, Weber AM, Mills RE. A recurrence-based approach for validating structural variation using long-read sequencing technology. Gigascience. 2017;6(8):1-9.

365. Beck TF, Mullikin JC, Program NCS, Biesecker LG. Systematic Evaluation of Sanger Validation of Next-Generation Sequencing Variants. Clin Chem. 2016;62(4):647-54.

366. Strom SP, Lee H, Das K, Vilain E, Nelson SF, Grody WW, et al. Assessing the necessity of confirmatory testing for exome-sequencing results in a clinical molecular diagnostic laboratory. Genet Med. 2014;16(7):510-5.

367. Popejoy AB, Fullerton SM. Genomics is failing on diversity. Nature. 2016;538(7624):161-4.

368. Kobayashi Y, Yang S, Nykamp K, Garcia J, Lincoln SE, Topper SE. Pathogenic variant burden in the ExAC database: an empirical approach to evaluating population data for clinical variant interpretation. Genome Med. 2017;9(1):13.

369. Lam HY, Clark MJ, Chen R, Chen R, Natsoulis G, O'Huallachain M, et al. Performance comparison of whole-genome sequencing platforms. Nat Biotechnol. 2011;30(1):78-82.

370. Nakano K, Shiroma A, Shimoji M, Tamotsu H, Ashimine N, Ohki S, et al. Advantages of genome sequencing by long-read sequencer using SMRT technology in medical area. Hum Cell. 2017;30(3):149-61.

371. Sevim V, Bashir A, Chin CS, Miga KH. Alpha-CENTAURI: assessing novel centromeric repeat sequence variation with long read sequencing. Bioinformatics. 2016;32(13):1921-4.

372. Suzuki A, Suzuki M, Mizushima-Sugano J, Frith MC, Makalowski W, Kohno T, et al. Sequencing and phasing cancer mutations in lung cancers using a long-read portable sequencer. DNA Res. 2017;24(6):585-96.

373. Laver TW, Caswell RC, Moore KA, Poschmann J, Johnson MB, Owens MM, et al. Pitfalls of haplotype phasing from amplicon-based long-read sequencing. Sci Rep. 2016;6:21746.

374. Decision in Williams vs Quest Diagnostics et al, (2018).

375. Ayuso C, Millan JM, Mancheno M, Dal-Re R. Informed consent for whole-genome sequencing studies in the clinical setting. Proposed recommendations on essential content and process. Eur J Hum Genet. 2013;21(10):1054-9.

376. Mayer AN, Dimmock DP, Arca MJ, Bick DP, Verbsky JW, Worthey EA, et al. A timely arrival for genomic medicine. Genet Med. 2011;13(3):195-6.

377. Bell CJ, Dinwiddie DL, Miller NA, Hateley SL, Ganusova EE, Mudge J, et al. Carrier testing for severe childhood recessive diseases by next-generation sequencing. Science translational medicine. 2011;3(65):65ra4.

378. Leong IU, Stuckey A, Lai D, Skinner JR, Love DR. Assessment of the predictive accuracy of five in silico prediction tools, alone or in combination, and two metaservers to classify long QT syndrome gene mutations. BMC Med Genet. 2015;16:34.

379. Hoskinson DC, Dubuc AM, Mason-Suares H. The current state of clinical interpretation of sequence variants. Curr Opin Genet Dev. 2017;42:33-9.

380. Vos J, Otten W, van Asperen C, Jansen A, Menko F, Tibben A. The counsellees' view of an unclassified variant in BRCA1/2: recall, interpretation, and impact on life. Psychooncology. 2008;17(8):822-30.

381. Ackerman MJ. Genetic purgatory and the cardiac channelopathies: Exposing the variants of uncertain/unknown significance issue. Heart Rhythm. 2015;12(11):2325-31.

382. England G. The 100,000 Genomes Project: Genomics England; 2018 [25/4/2018]. Available from: https://www.genomicsengland.co.uk/taking-part/results/.

383. Dewey FE, Murray MF, Overton JD, Habegger L, Leader JB, Fetterolf SN, et al. Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study. Science (New York, NY). 2016;354(6319).

384. Bennette CS, Gallego CJ, Burke W, Jarvik GP, Veenstra DL. The cost-effectiveness of returning incidental findings from next-generation genomic sequencing. Genet Med. 2015;17(7):587-95.

385. Johannsen W. Elemente der exakten Ereblichkeitslehre (Elements of an exact theory of heredity): Jena,G. Fischer; 1909.

386. Nachtomy O, Shavit A, Yakhini Z. Gene expression and the concept of the phenotype. Stud Hist Philos Biol Biomed Sci. 2007;38(1):238-54.

387. Stitziel NO, Fouchier SW, Sjouke B, Peloso GM, Moscoso AM, Auer PL, et al. Exome sequencing and directed clinical phenotyping diagnose cholesterol ester storage disease presenting as autosomal recessive hypercholesterolemia. Arterioscler Thromb Vasc Biol. 2013;33(12):2909-14.

388. de Goede C, Yue WW, Yan G, Ariyaratnam S, Chandler KE, Downes L, et al. Role of reverse phenotyping in interpretation of next generation sequencing data and a review of INPP5E related disorders. Eur J Paediatr Neurol. 2016;20(2):286-95.

389. Deveault C, Billingsley G, Duncan JL, Bin J, Theal R, Vincent A, et al. BBS genotype-phenotype assessment of a multiethnic patient cohort calls for a revision of the disease definition. Human mutation. 2011;32(6):610-9.

390. Ackerman JP, Bartos DC, Kapplinger JD, Tester DJ, Delisle BP, Ackerman MJ. The Promise and Peril of Precision Medicine: Phenotyping Still Matters Most. Mayo Clin Proc. 2016.

391. Heller ER, Khan SG, Kuschal C, Tamura D, DiGiovanna JJ, Kraemer KH. Mutations in the TTDN1 gene are associated with a distinct trichothiodystrophy phenotype. The Journal of investigative dermatology. 2015;135(3):734-41.

392. Huang XF, Huang F, Wu KC, Wu J, Chen J, Pang CP, et al. Genotype-phenotype correlation and mutation spectrum in a large cohort of patients with inherited retinal dystrophy revealed by next-generation sequencing. Genet Med. 2015;17(4):271-8.

393. Newton KF, Mallinson EK, Bowen J, Lalloo F, Clancy T, Hill J, et al. Genotype-phenotype correlation in colorectal polyposis. Clin Genet. 2012;81(6):521-31.

394. Darnell AJ, Austin H, Bluemke DA, Cannon RO, 3rd, Fischbeck K, Gahl W, et al. A Clinical Service to Support the Return of Secondary Genomic Findings in Human Research. American journal of human genetics. 2016;98(3):435-41.

395. Lyon GJ, Wang K. Identifying disease mutations in genomic medicine settings: current challenges and how to accelerate progress. Genome Med. 2012;4(7):58.

396. Lythgoe MP, Rhodes CJ, Ghataorhe P, Attard M, Wharton J, Wilkins MR. Why drugs fail in clinical trials in pulmonary arterial hypertension, and strategies to succeed in the future. Pharmacology & therapeutics. 2016;164:195-203.

397. Cree IA, Booton R, Cane P, Gosney J, Ibrahim M, Kerr K, et al. PD-L1 testing for lung cancer in the UK: recognizing the challenges for implementation. Histopathology. 2016;69(2):177-86.

398. Hamid JS, Hu P, Roslin NM, Ling V, Greenwood CM, Beyene J. Data integration in genetics and genomics: methods and challenges. Hum Genomics Proteomics. 2009;2009.

399. Lin D, Zhang J, Li J, He H, Deng HW, Wang YP. Integrative analysis of multiple diverse omics datasets by sparse group multitask regression. Front Cell Dev Biol. 2014;2:62.

400. Begley CG, Ellis LM. Drug development: Raise standards for preclinical cancer research. Nature. 2012;483(7391):531-3.

401. Guey LT, Kravic J, Melander O, Burtt NP, Laramie JM, Lyssenko V, et al. Power in the phenotypic extremes: a simulation study of power in discovery and replication of rare variants. Genet Epidemiol. 2011;35(4):236-46.

402. Ahmadi Adl A, Lee HS, Qian X. Detecting pairwise interactive effects of continuous random variables for bimarker identification with small sample size. IEEE/ACM Trans Comput Biol Bioinform. 2016.

403. Delude CM. Deep phenotyping: The details of disease. Nature. 2015;527(7576):S14-5.

404. Baynam G, Walters M, Claes P, Kung S, LeSouef P, Dawkins H, et al. Phenotyping: targeting genotype's rich cousin for diagnosis. J Paediatr Child Health. 2015;51(4):381-6.

405. Robinson PN. Deep phenotyping for precision medicine. Human mutation. 2012;33(5):777-80.

406. Fassihi H, Sethi M, Fawcett H, Wing J, Chandler N, Mohammed S, et al. Deep phenotyping of 89 xeroderma pigmentosum patients reveals unexpected heterogeneity dependent on the precise molecular defect. Proc Natl Acad Sci U S A. 2016;113(9):E1236-45.

407. Westbury SK, Turro E, Greene D, Lentaigne C, Kelly AM, Bariana TK, et al. Human phenotype ontology annotation and cluster analysis to unravel genetic defects in 707 cases with unexplained bleeding and platelet disorders. Genome Med. 2015;7(1):36.

408. Wang Y, Liu P, Xu Y, Zhang W, Tong L, Guo Z, et al. Preoperative neutrophil-to-lymphocyte ratio predicts response to first-line platinum-based chemotherapy and prognosis in serous ovarian cancer. Cancer Chemother Pharmacol. 2015;75(2):255-62.

409. Lyssenko V, Bianchi C, Del Prato S. Personalized Therapy by Phenotype and Genotype. Diabetes Care. 2016;39 Suppl 2:S127-36.

410. Sun W, Rumshisky A, Uzuner O. Evaluating temporal relations in clinical text: 2012 i2b2 Challenge. J Am Med Inform Assoc. 2013;20(5):806-13.

411. O'Malley KJ, Cook KF, Price MD, Wildes KR, Hurdle JF, Ashton CM. Measuring diagnoses: ICD code accuracy. Health Serv Res. 2005;40(5 Pt 2):1620-39.

412. Shen W, Wong B, Chin JY, Lee M, Coulter C, Braund R. Comparison of documentation of patient reported adverse drug reactions on both paper-based medication charts and electronic medication charts at a New Zealand hospital. N Z Med J. 2016;129(1444):90-6.

413. Klitzman R. Exclusion of genetic information from the medical record: ethical and medical dilemmas. JAMA. 2010;304(10):1120-1.

414. Frey LJ, Lenert L, Lopez-Campos G. EHR Big Data Deep Phenotyping. Contribution of the IMIA Genomic Medicine Working Group. Yearb Med Inform. 2014;9:206-11.

415. Roden DM, Denny JC. Integrating electronic health record genotype and phenotype datasets to transform patient care. Clin Pharmacol Ther. 2016;99(3):298-305.

416. Richesson RL, Sun J, Pathak J, Kho AN, Denny JC. Clinical phenotyping in selected national networks: demonstrating the need for high-throughput, portable, and computational methods. Artif Intell Med. 2016;71:57-61.

417. Wei WQ, Teixeira PL, Mo H, Cronin RM, Warner JL, Denny JC. Combining billing codes, clinical notes, and medications from electronic health records provides superior phenotyping performance. J Am Med Inform Assoc. 2016;23(e1):e20-7.

418. Bodenreider OB, A. Towards Desiderata for an Ontology of Diseases for the Annotation of Biological Datasets ICBO: International Conference on Biomedical Ontology; Buffalo, NY, USA2009.

419. Robinson PN, Kohler S, Bauer S, Seelow D, Horn D, Mundlos S. The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease. American journal of human genetics. 2008;83(5):610-5.

420. Kohler S, Doelken SC, Mungall CJ, Bauer S, Firth HV, Bailleul-Forestier I, et al. The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data. Nucleic Acids Res. 2014;42(Database issue):D966-74.

421. Kohler SR, P. The Human Phenotype Ontology website 2016 [08/11/2016]. Available from: http://human-phenotype-ontology.github.io/.

422. Deans AR, Lewis SE, Huala E, Anzaldo SS, Ashburner M, Balhoff JP, et al. Finding our way through phenotypes. PLoS Biol. 2015;13(1):e1002033.

423. Deng Y, Gao L, Wang B, Guo X. HPOSim: an R package for phenotypic similarity measure and enrichment analysis based on the human phenotype ontology. PloS one. 2015;10(2):e0115692.

424. Girdea M, Dumitriu S, Fiume M, Bowdin S, Boycott KM, Chenier S, et al. PhenoTips: patient phenotyping software for clinical and research use. Human mutation. 2013;34(8):1057-65.

425. Kohler S, Vasilevsky NA, Engelstad M, Foster E, McMurry J, Ayme S, et al. The Human Phenotype Ontology in 2017. Nucleic Acids Res. 2017;45(D1):D865-D76.

426. Robinson PN, Mungall CJ, Haendel M. Capturing phenotypes for precision medicine. Cold Spring Harb Mol Case Stud. 2015;1(1):a000372.

427. Buske OJ, Girdea M, Dumitriu S, Gallinger B, Hartley T, Trang H, et al. PhenomeCentral: a portal for phenotypic and genotypic matchmaking of patients with rare genetic diseases. Human mutation. 2015;36(10):931-40.

428. Sobreira NLM, Arachchi H, Buske OJ, Chong JX, Hutton B, Foreman J, et al. Matchmaker Exchange. Curr Protoc Hum Genet. 2017;95:9 31 1-9 15.

429. Organisation IHTSD. SNOMED [Available from: https://www.snomed.org/snomed-ct.

430. Heja G, Surjan G, Varga P. Ontological analysis of SNOMED CT. BMC Med Inform Decis Mak. 2008;8 Suppl 1:S8.

431. Winnenburg R, ; Bodenreider, O. Coverage of Phenotypes in Standard Terminologies: National Library of Medicine; 2014 [Available from: https://lhncbc.nlm.nih.gov/system/files/pub8937.pdf.

432. Dhombres F, Bodenreider O. Interoperability between phenotypes in research and healthcare terminologies--Investigating partial mappings between HPO and SNOMED CT. J Biomed Semantics. 2016;7:3.

433. Organisation WH. International Statistical Classification of Diseases and Related Health Problems, 10th Revision 2010 [cited 2018 10/03/2018]. Available from: http://apps.who.int/classifications/icd10/browse/2010/en.

434. Taboada M, Rodriguez H, Martinez D, Pardo M, Sobrido MJ. Automated semantic annotation of rare disease cases: a case study. Database (Oxford). 2014;2014.

435. Arighi CN, Carterette B, Cohen KB, Krallinger M, Wilbur WJ, Fey P, et al. An overview of the BioCreative 2012 Workshop Track III: interactive text mining task. Database (Oxford). 2013;2013:bas056.

436. Stark Z, Dashnow H, Lunke S, Tan TY, Yeung A, Sadedin S, et al. A clinically driven variant prioritization framework outperforms purely computational approaches for the diagnostic analysis of singleton WES data. Eur J Hum Genet. 2017;25(11):1268-72.

437. Mungall CJ, McMurry JA, Kohler S, Balhoff JP, Borromeo C, Brush M, et al. The Monarch Initiative: an integrative data and analytic platform connecting phenotypes to genotypes across species. Nucleic Acids Res. 2017;45(D1):D712-D22.

438. Maiella S, Olry A, Hanauer M, Lanneau V, Lourghi H, Donadille B, et al. Harmonising phenomics information for a better interoperability in the rare disease field. Eur J Med Genet. 2018.

439. Government HM. Data Protection Act: Her Majesty's Stationary Office; 1998 [Available from: https://www.legislation.gov.uk/ukpga/1998/29/contents.

440. Government HM. Data Protection Act: Her Majesty's Stationary Office; 2018 [Available from: http://www.legislation.gov.uk/ukpga/2018/12/contents/enacted.

441. UCL Data Safe Haven 2016 [8/11/2016]. Available from: https://www.ucl.ac.uk/isd/itforslms/services/handling-sens-data/tech-soln.

442. Gonzaga-Jauregui C, Harel T, Gambin T, Kousi M, Griffin LB, Francescatto L, et al. Exome Sequence Analysis Suggests that Genetic Burden Contributes to Phenotypic Variability and Complex Neuropathy. Cell Rep. 2015;12(7):1169-83.

443. Klassen T, Davis C, Goldman A, Burgess D, Chen T, Wheeler D, et al. Exome sequencing of ion channel genes reveals complex profiles confounding personal risk assessment in epilepsy. Cell. 2011;145(7):1036-48.

444. Grzasko N, Hus M, Pluta A, Jurczyszyn A, Walter-Croneck A, Morawska M, et al. Additional genetic abnormalities significantly worsen poor prognosis associated with 1q21 amplification in multiple myeloma patients. Hematol Oncol. 2013;31(1):41-8.

445. Al-Mulla F, Keith WN, Pickford IR, Going JJ, Birnie GD. Comparative genomic hybridization analysis of primary colorectal carcinomas and their synchronous metastases. Genes Chromosomes Cancer. 1999;24(4):306-14.

446. Davis EE, Katsanis N. The ciliopathies: a transitional model into systems biology of human genetic disease. Curr Opin Genet Dev. 2012;22(3):290-303.

447. Khanna H, Davis EE, Murga-Zamalloa CA, Estrada-Cuzcano A, Lopez I, den Hollander AI, et al. A common allele in RPGRIP1L is a modifier of retinal degeneration in ciliopathies. Nat Genet. 2009;41(6):739-45.

448. Asimit JL, Day-Williams AG, Morris AP, Zeggini E. ARIEL and AMELIA: testing for an accumulation of rare variants using next-generation sequencing data. Hum Hered. 2012;73(2):84-94.

449. Li B, Leal SM. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. American journal of human genetics. 2008;83(3):311-21.

450. Emond MJ, Louie T, Emerson J, Zhao W, Mathias RA, Knowles MR, et al. Exome sequencing of extreme phenotypes identifies DCTN4 as a modifier of chronic Pseudomonas aeruginosa infection in cystic fibrosis. Nat Genet. 2012;44(8):886-9.

451. Sakurai K, Miyaso H, Eguchi A, Matsuno Y, Yamamoto M, Todaka E, et al. Chiba study of Mother and Children's Health (C-MACH): cohort study with omics analyses. BMJ Open. 2016;6(1):e010531.

452. Li YC, Yen JC, Chiu WT, Jian WS, Syed-Abdul S, Hsu MH. Building a national electronic medical record exchange system - experiences in Taiwan. Comput Methods Programs Biomed. 2015;121(1):14-20.

453. Ventura ML, Battan AM, Zorloni C, Abbiati L, Colombo M, Farina S, et al. The electronic medical record: pros and cons. J Matern Fetal Neonatal Med. 2011;24 Suppl 1:163-6.

454. Gummadi S, Housri N, Zimmers TA, Koniaris LG. Electronic medical record: a balancing act of patient safety, privacy and health care delivery. Am J Med Sci. 2014;348(3):238-43.

455. Perry JJ, Sutherland J, Symington C, Dorland K, Mansour M, Stiell IG. Assessment of the impact on time to complete medical record using an electronic medical record versus a paper record on emergency department patients: a study. Emerg Med J. 2014;31(12):980-5.

456. Parker S. The pooling of manpower and resources through the establishment of European reference networks and rare disease patient registries is a necessary area of collaboration for rare renal disorders. Nephrol Dial Transplant. 2014;29 Suppl 4:iv9-14.

457. Choquet R, Maaroufi M, de Carrara A, Messiaen C, Luigi E, Landais P. A methodology for a minimum data set for rare diseases to support national centers of excellence for healthcare and research. J Am Med Inform Assoc. 2015;22(1):76-85.

458. Kennell TI, Jr., Willig JH, Cimino JJ. Clinical Informatics Researcher's Desiderata for the Data Content of the Next Generation Electronic Health Record. Appl Clin Inform. 2017;8(4):1159-72.

459. Vogel F. Vogel, F. 1959. Moderne Probleme der Humangenetik. Ergebnisse der Inneren Medizin und Kinderheilkunde. 1959;12:52–125.

460. Marshall A. Genset-Abbott deal heralds pharmacogenomics era. Nat Biotechnol. 1997;15(9):829-30.

461. Zinkham WH, Lenhard RE, Jr., Childs B. A deficiency of glucose-6-phosphate dehydrogenase activity in erythrocytes from patients with favism. Bull Johns Hopkins Hosp. 1958;102(4):169-75.

462. Nebert DW, Zhang G, Vesell ES. From human genetics and genomics to pharmacogenetics and pharmacogenomics: past lessons, future directions. Drug Metab Rev. 2008;40(2):187-224.

463. Cordes WB, M.A. Experiences with plasmochin in malaria. Annual report. United Fruit Co; 1926 1926.

464. Alving AS, Carson PE, Flanagan CL, Ickes CE. Enzymatic deficiency in primaquine-sensitive erythrocytes. Science (New York, NY). 1956;124(3220):484-5.

465. Lehmann H, Ryan E. The familial incidence of low pseudocholinesterase level. Lancet (London, England). 1956;271(6934):124.

466. Motulsky AG. Drug reactions enzymes, and biochemical genetics. J Am Med Assoc. 1957;165(7):835-7.

467. Lehmann H, Liddell J. Genetical Variants of Human Serum Pseudocholinesterase. Prog Med Genet. 1964;23:75-105.

468. Eichelbaum M, Bertilsson L, Sawe J, Zekorn C. Polymorphic oxidation of sparteine and debrisoquine: related pharmacogenetic entities. Clin Pharmacol Ther. 1982;31(2):184-6.

469. Marez D, Legrand M, Sabbagh N, Lo Guidice JM, Spire C, Lafitte JJ, et al. Polymorphism of the cytochrome P450 CYP2D6 gene in a European population: characterization of 48 mutations and 53 alleles, their frequencies and evolution. Pharmacogenetics. 1997;7(3):193-202.

470. Jones DS. How personalized medicine became genetic, and racial: Werner Kalow and the formations of pharmacogenetics. J Hist Med Allied Sci. 2013;68(1):1-48.

471. Whirl-Carrillo M, McDonagh EM, Hebert JM, Gong L, Sangkuhl K, Thorn CF, et al. Pharmacogenomics knowledge for personalized medicine. Clin Pharmacol Ther. 2012;92(4):414-7.

472. Wang L, Pelleymounter L, Weinshilboum R, Johnson JA, Hebert JM, Altman RB, et al. Very important pharmacogene summary: thiopurine S-methyltransferase. Pharmacogenet Genomics. 2010;20(6):401-5.

473. Thorn CF, Klein TE, Altman RB. PharmGKB summary: very important pharmacogene information for angiotensin-converting enzyme. Pharmacogenet Genomics. 2010;20(2):143-6.

474. Thorn CF, Lamba JK, Lamba V, Klein TE, Altman RB. PharmGKB summary: very important pharmacogene information for CYP2B6. Pharmacogenet Genomics. 2010;20(8):520-3.

475. Thorn CF, Ji Y, Weinshilboum RM, Altman RB, Klein TE. PharmGKB summary: very important pharmacogene information for GSTT1. Pharmacogenet Genomics. 2012;22(8):646-51.

476. Thorn CF, Grosser T, Klein TE, Altman RB. PharmGKB summary: very important pharmacogene information for PTGS2. Pharmacogenet Genomics. 2011;21(9):607-13.

477. Thorn CF, Aklillu E, Klein TE, Altman RB. PharmGKB summary: very important pharmacogene information for CYP1A2. Pharmacogenet Genomics. 2012;22(1):73-7.

478. Scott SA, Sangkuhl K, Shuldiner AR, Hulot JS, Thorn CF, Altman RB, et al. PharmGKB summary: very important pharmacogene information for cytochrome P450, family 2, subfamily C, polypeptide 19. Pharmacogenet Genomics. 2012;22(2):159-65.

479. Poon AH, Gong L, Brasch-Andersen C, Litonjua AA, Raby BA, Hamid Q, et al. Very important pharmacogene summary for VDR. Pharmacogenet Genomics. 2012;22(10):758-63.

480. Oshiro C, Mangravite L, Klein T, Altman R. PharmGKB very important pharmacogene: SLCO1B1. Pharmacogenet Genomics. 2010;20(3):211-6.

481. Medina MW, Sangkuhl K, Klein TE, Altman RB. PharmGKB: very important pharmacogene--HMGCR. Pharmacogenet Genomics. 2011;21(2):98-101.

482. McDonagh EM, Thorn CF, Bautista JM, Youngster I, Altman RB, Klein TE. PharmGKB summary: very important pharmacogene information for G6PD. Pharmacogenet Genomics. 2012;22(3):219-28.

483. McDonagh EM, Clancy JP, Altman RB, Klein TE. PharmGKB summary: very important pharmacogene information for CFTR. Pharmacogenet Genomics. 2015;25(3):149-56.

484. McDonagh EM, Boukouvala S, Aklillu E, Hein DW, Altman RB, Klein TE. PharmGKB summary: very important pharmacogene information for N-acetyltransferase 2. Pharmacogenet Genomics. 2014;24(8):409-25.

485. Litonjua AA, Gong L, Duan QL, Shin J, Moore MJ, Weiss ST, et al. Very important pharmacogene summary ADRB2. Pharmacogenet Genomics. 2010;20(1):64-9.

486. Lamba J, Hebert JM, Schuetz EG, Klein TE, Altman RB. PharmGKB summary: very important pharmacogene information for CYP3A5. Pharmacogenet Genomics. 2012;22(7):555-8.

487. Hodoglugil U, Carrillo MW, Hebert JM, Karachaliou N, Rosell RC, Altman RB, et al. PharmGKB summary: very important pharmacogene information for the epidermal growth factor receptor. Pharmacogenet Genomics. 2013;23(11):636-42.

488. Hodges LM, Markova SM, Chinn LW, Gow JM, Kroetz DL, Klein TE, et al. Very important pharmacogene summary: ABCB1 (MDR1, P-glycoprotein). Pharmacogenet Genomics. 2011;21(3):152-61.

489. Hildebrandt M, Adjei A, Weinshilboum R, Johnson JA, Berlin DS, Klein TE, et al. Very important pharmacogene summary: sulfotransferase 1A1. Pharmacogenet Genomics. 2009;19(6):404-6.

490. Goswami S, Gong L, Giacomini K, Altman RB, Klein TE. PharmGKB summary: very important pharmacogene information for SLC22A1. Pharmacogenet Genomics. 2014;24(6):324-8.

491. Barbarino JM, Kroetz DL, Klein TE, Altman RB. PharmGKB summary: very important pharmacogene information for human leukocyte antigen B. Pharmacogenet Genomics. 2015;25(4):205-21.

492. Aquilante CL, Niemi M, Gong L, Altman RB, Klein TE. PharmGKB summary: very important pharmacogene information for cytochrome P450, family 2, subfamily C, polypeptide 8. Pharmacogenet Genomics. 2013;23(12):721-8.

493. Alvarellos ML, Sangkuhl K, Daneshjou R, Whirl-Carrillo M, Altman RB, Klein TE. PharmGKB summary: very important pharmacogene information for CYP4F2. Pharmacogenet Genomics. 2015;25(1):41-7.

494. Alvarellos ML, Krauss RM, Wilke RA, Altman RB, Klein TE. PharmGKB summary: very important pharmacogene information for RYR1. Pharmacogenet Genomics. 2016;26(3):138-44.

495. (FDA) FaDA. Table of Pharmacogenomic Biomarkers in Drug Labeling USA2016 [18/11/2016]. Available from: http://www.fda.gov/Drugs/ScienceResearch/ResearchAreas/Pharmacogenetics/ucm083378.htm .

496. Ehmann F, Caneva L, Prasad K, Paulmichl M, Maliepaard M, Llerena A, et al. Pharmacogenomic information in drug labels: European Medicines Agency perspective. Pharmacogenomics J. 2015;15(3):201-10.

497. Grosse SD, Khoury MJ. What is the clinical utility of genetic testing? Genet Med. 2006;8(7):448-50.

498. Swen JJ, Wilting I, de Goede AL, Grandia L, Mulder H, Touw DJ, et al. Pharmacogenetics: from bench to byte. Clin Pharmacol Ther. 2008;83(5):781-7.

499. Swen JJ, Nijenhuis M, de Boer A, Grandia L, Maitland-van der Zee AH, Mulder H, et al. Pharmacogenetics: from bench to byte--an update of guidelines. Clin Pharmacol Ther. 2011;89(5):662-73.

500. Fohner AE, McDonagh EM, Clancy JP, Whirl Carrillo M, Altman RB, Klein TE. PharmGKB summary: ivacaftor pathway, pharmacokinetics/pharmacodynamics. Pharmacogenet Genomics. 2017;27(1):39-42.

501. Whiting P, Al M, Burgers L, Westwood M, Ryder S, Hoogendoorn M, et al. Ivacaftor for the treatment of patients with cystic fibrosis and the G551D mutation: a systematic review and cost-effectiveness analysis. Health Technol Assess. 2014;18(18):1-106.

502. De Boeck K, Munck A, Walker S, Faro A, Hiatt P, Gilmartin G, et al. Efficacy and safety of ivacaftor in patients with cystic fibrosis and a non-G551D gating mutation. J Cyst Fibros. 2014;13(6):674-80.

503. Clancy JP, Johnson SG, Yee SW, McDonagh EM, Caudle KE, Klein TE, et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) guidelines for ivacaftor therapy in the context of CFTR genotype. Clin Pharmacol Ther. 2014;95(6):592-7.

504. Cystic Fibrosis Genotype-Phenotype C. Correlation between genotype and phenotype in patients with cystic fibrosis. N Engl J Med. 1993;329(18):1308-13.

505. Jarrar YB, Lee SJ. Molecular functionality of CYP2C9 polymorphisms and their influence on drug therapy. Drug Metabol Drug Interact. 2014;29(4):211-20.

506. Lee CR, Goldstein JA, Pieper JA. Cytochrome P450 2C9 polymorphisms: a comprehensive review of the in-vitro and human data. Pharmacogenetics. 2002;12(3):251-63.

507. Gaedigk A, Ingelman-Sundberg M, Miller NA, Leeder JS, Whirl-Carrillo M, Klein TE, et al. The Pharmacogene Variation (PharmVar) Consortium: Incorporation of the Human Cytochrome P450 (CYP) Allele Nomenclature Database. Clin Pharmacol Ther. 2018;103(3):399-401.

508. Caudle KE, Rettie AE, Whirl-Carrillo M, Smith LH, Mintzer S, Lee MT, et al. Clinical pharmacogenetics implementation consortium guidelines for CYP2C9 and HLA-B genotypes and phenytoin dosing. Clin Pharmacol Ther. 2014;96(5):542-8.

509. Johnson JA, Gong L, Whirl-Carrillo M, Gage BF, Scott SA, Stein CM, et al. Clinical Pharmacogenetics Implementation Consortium Guidelines for CYP2C9 and VKORC1 genotypes and warfarin dosing. Clin Pharmacol Ther. 2011;90(4):625-9.

510. Johnson JA, Caudle KE, Gong L, Whirl-Carrillo M, Stein CM, Scott SA, et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for Pharmacogenetics-Guided Warfarin Dosing: 2017 Update. Clin Pharmacol Ther. 2017;102(3):397-404.

511. Kirchheiner J, Brockmoller J. Clinical consequences of cytochrome P450 2C9 polymorphisms. Clin Pharmacol Ther. 2005;77(1):1-16.

512. Aithal GP, Day CP, Kesteven PJ, Daly AK. Association of polymorphisms in the cytochrome P450 CYP2C9 with warfarin dose requirement and risk of bleeding complications. Lancet (London, England). 1999;353(9154):717-9.

513. Sistonen J, Fuselli S, Palo JU, Chauhan N, Padh H, Sajantila A. Pharmacogenetic variation at CYP2C9, CYP2C19, and CYP2D6 at global and microgeographic scales. Pharmacogenet Genomics. 2009;19(2):170-9.

514. Solus JF, Arietta BJ, Harris JR, Sexton DP, Steward JQ, McMunn C, et al. Genetic variation in eleven phase I drug metabolism genes in an ethnically diverse population. Pharmacogenomics. 2004;5(7):895-931.

515. Cespedes-Garro C, Fricke-Galindo I, Naranjo ME, Rodrigues-Soares F, Farinas H, de Andres F, et al. Worldwide interethnic variability and geographical distribution of CYP2C9 genotypes and phenotypes. Expert Opin Drug Metab Toxicol. 2015;11(12):1893-905.

516. Hicks JK, Bishop JR, Sangkuhl K, Muller DJ, Ji Y, Leckband SG, et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for CYP2D6 and CYP2C19 Genotypes and Dosing of Selective Serotonin Reuptake Inhibitors. Clin Pharmacol Ther. 2015;98(2):127-34.

517. Hicks JK, Swen JJ, Thorn CF, Sangkuhl K, Kharasch ED, Ellingrod VL, et al. Clinical Pharmacogenetics Implementation Consortium guideline for CYP2D6 and CYP2C19 genotypes and dosing of tricyclic antidepressants. Clin Pharmacol Ther. 2013;93(5):402-8.

518. Scott SA, Sangkuhl K, Stein CM, Hulot JS, Mega JL, Roden DM, et al. Clinical Pharmacogenetics Implementation Consortium guidelines for CYP2C19 genotype and clopidogrel therapy: 2013 update. Clin Pharmacol Ther. 2013;94(3):317-23.

519. Scott SA, Sangkuhl K, Gardner EE, Stein CM, Hulot JS, Johnson JA, et al. Clinical Pharmacogenetics Implementation Consortium guidelines for cytochrome P450-2C19 (CYP2C19) genotype and clopidogrel therapy. Clin Pharmacol Ther. 2011;90(2):328-32.

520. Moriyama B, Obeng AO, Barbarino J, Penzak SR, Henning SA, Scott SA, et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) Guidelines for CYP2C19 and Voriconazole Therapy. Clin Pharmacol Ther. 2016.

521. Kuo CH, Lu CY, Shih HY, Liu CJ, Wu MC, Hu HM, et al. CYP2C19 polymorphism influences Helicobacter pylori eradication. World J Gastroenterol. 2014;20(43):16029-36.

522. Goldstein JA, Ishizaki T, Chiba K, de Morais SM, Bell D, Krahn PM, et al. Frequencies of the defective CYP2C19 alleles responsible for the mephenytoin poor metabolizer phenotype in various Oriental, Caucasian, Saudi Arabian and American black populations. Pharmacogenetics. 1997;7(1):59-64.

523. Ingelman-Sundberg M, Sim SC, Gomez A, Rodriguez-Antona C. Influence of cytochrome P450 polymorphisms on drug therapies: pharmacogenetic, pharmacoepigenetic and clinical aspects. Pharmacology & therapeutics. 2007;116(3):496-526.

524. Goetz MP, Sangkuhl K, Guchelaar HJ, Schwab M, Province M, Whirl-Carrillo M, et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for CYP2D6 and Tamoxifen Therapy. Clin Pharmacol Ther. 2018;103(5):770-7.

525. Bell GC, Caudle KE, Whirl-Carrillo M, Gordon RJ, Hikino K, Prows CA, et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) guideline for CYP2D6 genotype and use of ondansetron and tropisetron. Clin Pharmacol Ther. 2016.

526. Crews KR, Gaedigk A, Dunnenberger HM, Klein TE, Shen DD, Callaghan JT, et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) guidelines for codeine therapy in the context of cytochrome P450 2D6 (CYP2D6) genotype. Clin Pharmacol Ther. 2012;91(2):321-6.

527. Crews KR, Gaedigk A, Dunnenberger HM, Leeder JS, Klein TE, Caudle KE, et al. Clinical Pharmacogenetics Implementation Consortium guidelines for cytochrome P450 2D6 genotype and codeine therapy: 2014 update. Clin Pharmacol Ther. 2014;95(4):376-82.

528. Gaedigk A. Complexities of CYP2D6 gene analysis and interpretation. Int Rev Psychiatry. 2013;25(5):534-53.

529. Rojas L, Neumann I, Herrero MJ, Boso V, Reig J, Poveda JL, et al. Effect of CYP3A5*3 on kidney transplant recipients treated with tacrolimus: a systematic review and meta-analysis of observational studies. Pharmacogenomics J. 2015;15(1):38-48.

530. Birdwell KA, Decker B, Barbarino JM, Peterson JF, Stein CM, Sadee W, et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) Guidelines for CYP3A5 Genotype and Tacrolimus Dosing. Clin Pharmacol Ther. 2015;98(1):19-24.

531. Kuehl P, Zhang J, Lin Y, Lamba J, Assem M, Schuetz J, et al. Sequence diversity in CYP3A promoters and characterization of the genetic basis of polymorphic CYP3A5 expression. Nat Genet. 2001;27(4):383-91.

532. Thompson EE, Kuttab-Boulos H, Witonsky D, Yang L, Roe BA, Di Rienzo A. CYP3A variation and the evolution of salt-sensitivity variants. American journal of human genetics. 2004;75(6):1059-69.

533. Kurose K, Sugiyama E, Saito Y. Population differences in major functional polymorphisms of pharmacokinetics/pharmacodynamics-related genes in Eastern Asians and Europeans: implications in the clinical trials for novel drug development. Drug Metab Pharmacokinet. 2012;27(1):9-54.

534. Van Kuilenburg AB, Vreken P, Abeling NG, Bakker HD, Meinsma R, Van Lenthe H, et al. Genotype and phenotype in patients with dihydropyrimidine dehydrogenase deficiency. Human genetics. 1999;104(1):1-9.

535. Van Kuilenburg AB, Meinsma R, Zoetekouw L, Van Gennip AH. Increased risk of grade IV neutropenia after administration of 5-fluorouracil due to a dihydropyrimidine dehydrogenase deficiency: high prevalence of the IVS14+1g>a mutation. Int J Cancer. 2002;101(3):253-8.

536. McLeod HL, Collie-Duguid ES, Vreken P, Johnson MR, Wei X, Sapone A, et al. Nomenclature for human DPYD alleles. Pharmacogenetics. 1998;8(6):455-9.

537. Johnson MR, Wang K, Diasio RB. Profound dihydropyrimidine dehydrogenase deficiency resulting from a novel compound heterozygote genotype. Clin Cancer Res. 2002;8(3):768-74.

538. Mattison LK, Johnson MR, Diasio RB. A comparative analysis of translated dihydropyrimidine dehydrogenase cDNA; conservation of functional domains and relevance to genetic polymorphisms. Pharmacogenetics. 2002;12(2):133-44.

539. Amstutz U, Henricks LM, Offer SM, Barbarino J, Schellens JHM, Swen JJ, et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for Dihydropyrimidine Dehydrogenase Genotype and Fluoropyrimidine Dosing: 2017 Update. Clin Pharmacol Ther. 2018;103(2):210-6.

540. Thorelli E. Mechanisms that regulate the anticoagulant function of coagulation factor V. Scand J Clin Lab Invest Suppl. 1999;229:19-26.

541. Bertina RM, Koeleman BP, Koster T, Rosendaal FR, Dirven RJ, de Ronde H, et al. Mutation in blood coagulation factor V associated with resistance to activated protein C. Nature. 1994;369(6475):64-7.

542. Beauchamp NJ, Daly ME, Hampton KK, Cooper PC, Preston FE, Peake IR. High prevalence of a mutation in the factor V gene within the U.K. population: relationship to activated protein C resistance and familial thrombosis. British journal of haematology. 1994;88(1):219-22.

543. Ridker PM, Miletich JP, Stampfer MJ, Goldhaber SZ, Lindpaintner K, Hennekens CH. Factor V Leiden and risks of recurrent idiopathic venous thromboembolism. Circulation. 1995;92(10):2800-2.

544. de Bruijn SF, Stam J, Koopman MM, Vandenbroucke JP. Case-control study of risk of cerebral sinus thrombosis in oral contraceptive users and in [correction of who are] carriers of hereditary prothrombotic conditions. The Cerebral Venous Sinus Thrombosis Study Group. BMJ. 1998;316(7131):589-92.

545. Takizawa T, Huang IY, Ikuta T, Yoshida A. Human glucose-6-phosphate dehydrogenase: primary structure and cDNA cloning. Proc Natl Acad Sci U S A. 1986;83(12):4157-61.

546. Rinaldi A, Filippi G, Siniscalco M. Variability of red cell phenotypes between and within individuals in an unbiased sample of 77 heterozygotes for G6PD deficiency in Sardinia. American journal of human genetics. 1976;28(5):496-505.

547. Youngster I, Arcavi L, Schechmaster R, Akayzen Y, Popliski H, Shimonov J, et al. Medications and glucose-6-phosphate dehydrogenase deficiency: an evidence-based review. Drug Saf. 2010;33(9):713-26.

548. Relling MV, McDonagh EM, Chang T, Caudle KE, McLeod HL, Haidar CE, et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) guidelines for rasburicase therapy in the context of G6PD deficiency genotype. Clin Pharmacol Ther. 2014;96(2):169-74.

549. Minucci A, Moradkhani K, Hwang MJ, Zuppi C, Giardina B, Capoluongo E. Glucose-6-phosphate dehydrogenase (G6PD) mutations database: review of the "old" and update of the new mutations. Blood Cells Mol Dis. 2012;48(3):154-65.

550. Luzzatto L, Seneca E. G6PD deficiency: a classic example of pharmacogenetics with on-going clinical implications. British journal of haematology. 2014;164(4):469-80.

551. Nkhoma ET, Poole C, Vannappagari V, Hall SA, Beutler E. The global prevalence of glucose-6-phosphate dehydrogenase deficiency: a systematic review and meta-analysis. Blood Cells Mol Dis. 2009;42(3):267-78.

552. Guindo A, Fairhurst RM, Doumbo OK, Wellems TE, Diallo DA. X-linked G6PD deficiency protects hemizygous males but not heterozygous females against severe malaria. PLoS Med. 2007;4(3):e66.

553. Mockenhaupt FP, Mandelkow J, Till H, Ehrhardt S, Eggelte TA, Bienzle U. Reduced prevalence of Plasmodium falciparum infection and of concomitant anaemia in pregnant women with heterozygous G6PD deficiency. Trop Med Int Health. 2003;8(2):118-24.

554. Hedrick SM. Dawn of the hunt for nonclassical MHC function. Cell. 1992;70(2):177-80.

555. Phillips EJ, Sukasem C, Whirl-Carrillo M, Muller DJ, Dunnenberger HM, Chantratita W, et al. Clinical Pharmacogenetics Implementation Consortium Guideline for HLA Genotype and Use of Carbamazepine and Oxcarbazepine: 2017 Update. Clin Pharmacol Ther. 2018;103(4):574-81.

556. Kaniwa N, Saito Y. The risk of cutaneous adverse reactions among patients with the HLA-A* 31:01 allele who are given carbamazepine, oxcarbazepine or eslicarbazepine: a perspective review. Ther Adv Drug Saf. 2013;4(6):246-53.

557. Amstutz U, Shear NH, Rieder MJ, Hwang S, Fung V, Nakamura H, et al. Recommendations for HLA-B*15:02 and HLA-A*31:01 genetic testing to reduce the risk of carbamazepine-induced hypersensitivity reactions. Epilepsia. 2014;55(4):496-506.

558. Hershfield MS, Callaghan JT, Tassaneeyakul W, Mushiroda T, Thorn CF, Klein TE, et al. Clinical Pharmacogenetics Implementation Consortium guidelines for human leukocyte antigen-B genotype and allopurinol dosing. Clin Pharmacol Ther. 2013;93(2):153-8.

559. Martin MA, Klein TE, Dong BJ, Pirmohamed M, Haas DW, Kroetz DL, et al. Clinical pharmacogenetics implementation consortium guidelines for HLA-B genotype and abacavir dosing. Clin Pharmacol Ther. 2012;91(4):734-8.

560. Leckband SG, Kelsoe JR, Dunnenberger HM, George AL, Jr., Tran E, Berger R, et al. Clinical Pharmacogenetics Implementation Consortium guidelines for HLA-B genotype and carbamazepine dosing. Clin Pharmacol Ther. 2013;94(3):324-8.

561. Ko TM, Tsai CY, Chen SY, Chen KS, Yu KH, Chu CS, et al. Use of HLA-B*58:01 genotyping to prevent allopurinol induced severe cutaneous adverse reactions in Taiwan: national prospective cohort study. BMJ. 2015;351:h4848.

562. Gonzalez-Galarza FF, Takeshita LY, Santos EJ, Kempson F, Maia MH, da Silva AL, et al. Allele frequency net 2015 update: new features for HLA epitopes, KIR and disease and HLA adverse drug reaction associations. Nucleic Acids Res. 2015;43(Database issue):D784-8.

563. Mallal S, Phillips E, Carosi G, Molina JM, Workman C, Tomazic J, et al. HLA-B*5701 screening for hypersensitivity to abacavir. N Engl J Med. 2008;358(6):568-79.

564. Sheppard P, Kindsvogel W, Xu W, Henderson K, Schlutsmeyer S, Whitmore TE, et al. IL-28, IL-29 and their class II cytokine receptor IL-28R. Nat Immunol. 2003;4(1):63-8.

565. Muir AJ, Gong L, Johnson SG, Lee MT, Williams MS, Klein TE, et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) guidelines for IFNL3 (IL28B) genotype and PEG interferon-alpha-based regimens. Clin Pharmacol Ther. 2014;95(2):141-6.

566. Conjeevaram HS, Fried MW, Jeffers LJ, Terrault NA, Wiley-Lucas TE, Afdhal N, et al. Peginterferon and ribavirin treatment in African American and Caucasian American patients with hepatitis C genotype 1. Gastroenterology. 2006;131(2):470-7.

567. Tanaka Y, Nishida N, Sugiyama M, Kurosaki M, Matsuura K, Sakamoto N, et al. Genome-wide association of IL28B with response to pegylated interferon-alpha and ribavirin therapy for chronic hepatitis C. Nat Genet. 2009;41(10):1105-9.

568. Aminkeng F, Ross CJ, Rassekh SR, Hwang S, Rieder MJ, Bhavsar AP, et al. Recommendations for genetic testing to reduce the incidence of anthracycline-induced cardiotoxicity. Br J Clin Pharmacol. 2016;82(3):683-95.

569. van de Steeg E, Stranecky V, Hartmannova H, Noskova L, Hrebicek M, Wagenaar E, et al. Complete OATP1B1 and OATP1B3 deficiency causes human Rotor syndrome by interrupting conjugated bilirubin reuptake into the liver. J Clin Invest. 2012;122(2):519-28.

570. Wilke RA, Ramsey LB, Johnson SG, Maxwell WD, McLeod HL, Voora D, et al. The clinical pharmacogenomics implementation consortium: CPIC guideline for SLCO1B1 and simvastatin-induced myopathy. Clin Pharmacol Ther. 2012;92(1):112-7.

571. Relling MV, Gardner EE, Sandborn WJ, Schmiegelow K, Pui CH, Yee SW, et al. Clinical Pharmacogenetics Implementation Consortium guidelines for thiopurine methyltransferase genotype and thiopurine dosing. Clin Pharmacol Ther. 2011;89(3):387-91.

572. Krynetski E, Evans WE. Drug methylation in cancer therapy: lessons from the TPMT polymorphism. Oncogene. 2003;22(47):7403-13.

573. Allan PW, Bennett LL, Jr. 6-Methylthioguanylic acid, a metabolite of 6-thioguanine. Biochem Pharmacol. 1971;20(4):847-52.

574. Weinshilboum RM, Sladek SL. Mercaptopurine pharmacogenetics: monogenic inheritance of erythrocyte thiopurine methyltransferase activity. American journal of human genetics. 1980;32(5):651-62.

575. Lennard L, Lilleyman JS, Van Loon J, Weinshilboum RM. Genetic variation in response to 6-mercaptopurine for childhood acute lymphoblastic leukaemia. Lancet (London, England). 1990;336(8709):225-9.

576. Lee JW, Pussegoda K, Rassekh SR, Monzon JG, Liu G, Hwang S, et al. Clinical Practice Recommendations for the Management and Prevention of Cisplatin-Induced Hearing Loss Using Pharmacogenetic Markers. Ther Drug Monit. 2016;38(4):423-31.

577. Schaeffeler E, Fischer C, Brockmeier D, Wernet D, Moerike K, Eichelbaum M, et al. Comprehensive analysis of thiopurine S-methyltransferase phenotype-genotype correlation in a large population of German-Caucasians and identification of novel TPMT variants. Pharmacogenetics. 2004;14(7):407-17.

578. Newman WG, Payne K, Tricker K, Roberts SA, Fargher E, Pushpakom S, et al. A pragmatic randomized controlled trial of thiopurine methyltransferase genotyping prior to azathioprine treatment: the TARGET study. Pharmacogenomics. 2011;12(6):815-26.

579. Coenen MJ, de Jong DJ, van Marrewijk CJ, Derijks LJ, Vermeulen SH, Wong DR, et al. Identification of Patients With Variants in TPMT and Dose Reduction Reduces Hematologic Events During Thiopurine Treatment of Inflammatory Bowel Disease. Gastroenterology. 2015;149(4):907-17 e7.

580. Basu NK, Ciotti M, Hwang MS, Kole L, Mitra PS, Cho JW, et al. Differential and special properties of the major human UGT1-encoded gastrointestinal UDP-glucuronosyltransferases enhance potential to control chemical uptake. The Journal of biological chemistry. 2004;279(2):1429-41.

581. Gammal RS, Court MH, Haidar CE, Iwuchukwu OF, Gaur AH, Alvarellos M, et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for UGT1A1 and Atazanavir Prescribing. Clin Pharmacol Ther. 2016;99(4):363-9.

582. Etienne-Grimaldi MC, Boyer JC, Thomas F, Quaranta S, Picard N, Loriot MA, et al. UGT1A1 genotype and irinotecan therapy: general review and implementation in routine practice. Fundam Clin Pharmacol. 2015;29(3):219-37.

583. Hall D, Ybazeta G, Destro-Bisol G, Petzl-Erler ML, Di Rienzo A. Variability at the uridine diphosphate glucuronosyltransferase 1A1 promoter in human populations and primates. Pharmacogenetics. 1999;9(5):591-9.

584. Ross KA, Bigham AW, Edwards M, Gozdzik A, Suarez-Kurtz G, Parra EJ. Worldwide allele frequency distribution of four polymorphisms associated with warfarin dose requirements. J Hum Genet. 2010;55(9):582-9.

585. Schalekamp T, Brasse BP, Roijers JF, Chahid Y, van Geest-Daalderop JH, de Vries-Goldschmeding H, et al. VKORC1 and CYP2C9 genotypes and acenocoumarol anticoagulation status: interaction between both genotypes affects overanticoagulation. Clin Pharmacol Ther. 2006;80(1):13-22.

586. Shaw K, Amstutz U, Kim RB, Lesko LJ, Turgeon J, Michaud V, et al. Clinical Practice Recommendations on Genetic Testing of CYP2C9 and VKORC1 Variants in Warfarin Therapy. Ther Drug Monit. 2015;37(4):428-36.

587. Perera MA, Cavallari LH, Limdi NA, Gamazon ER, Konkashbaev A, Daneshjou R, et al. Genetic variants associated with warfarin dose in African-American individuals: a genome-wide association study. Lancet (London, England). 2013;382(9894):790-6.

588. Danese E, Montagnana M, Johnson JA, Rettie AE, Zambon CF, Lubitz SA, et al. Impact of the CYP4F2 p.V433M polymorphism on coumarin dose requirement: systematic review and meta-analysis. Clin Pharmacol Ther. 2012;92(6):746-56.

589. Elens L, Haufroid V. Genotype-based tacrolimus dosing guidelines: with or without CYP3A4*22? Pharmacogenomics. 2017;18(16):1473-80.

590. Organisation WH. International drug monitoring: the role of national centres. Tech Rep Ser 4: World Health Organisation; 1972

591. Zhang J, Tian L, Huang J, Huang S, Chai T, Shen J. Cytochrome P450 2C9 gene polymorphism and warfarin maintenance dosage in pediatric patients: A systematic review and meta-analysis. Cardiovasc Ther. 2016.

592. Boyle KL, Rosenbaum CD. Oxycodone overdose in the pediatric population: case files of the University of Massachusetts Medical Toxicology Fellowship. J Med Toxicol. 2014;10(3):280-5.

593. Grover S, Kukreti R. HLA alleles and hypersensitivity to carbamazepine: an updated systematic review with meta-analysis. Pharmacogenet Genomics. 2014;24(2):94-112.

594. Chung WH, Wang CW, Dao RL. Severe cutaneous adverse drug reactions. J Dermatol. 2016;43(7):758-66.

595. Jensen BC, McLeod HL. Pharmacogenomics as a risk mitigation strategy for chemotherapeutic cardiotoxicity. Pharmacogenomics. 2013;14(2):205-13.

596. Stausberg J. International prevalence of adverse drug events in hospitals: an analysis of routine data from England, Germany, and the USA. BMC Health Serv Res. 2014;14:125.

597. Walsh D, Lavan A, Cushen AM, Williams D. Adverse drug reactions as a cause of admission to a Dublin-based university teaching hospital. Ir J Med Sci. 2015;184(2):441-7.

598. Pedros C, Formiga F, Corbella X, Arnau JM. Adverse drug reactions leading to urgent hospital admission in an elderly population: prevalence and main features. Eur J Clin Pharmacol. 2016;72(2):219-26.

599. Pirmohamed M, James S, Meakin S, Green C, Scott AK, Walley TJ, et al. Adverse drug reactions as cause of admission to hospital: prospective analysis of 18 820 patients. BMJ. 2004;329(7456):15-9.

600. Gallagher RM, Mason JR, Bird KA, Kirkham JJ, Peak M, Williamson PR, et al. Adverse drug reactions causing admission to a paediatric hospital. PloS one. 2012;7(12):e50127.

601. Damodaran SE, Pradhan SC, Umamaheswaran G, Kadambari D, Reddy KS, Adithan C. Genetic polymorphisms of CYP2D6 increase the risk for recurrence of breast cancer in

patients receiving tamoxifen as an adjuvant therapy. Cancer Chemother Pharmacol. 2012;70(1):75-81.

602. Romero-Gomez M, Gonzalez-Escribano MF, Torres B, Barroso N, Montes-Cano MA, Sanchez-Munoz D, et al. HLA class I B44 is associated with sustained response to interferon + ribavirin therapy in patients with chronic hepatitis C. Am J Gastroenterol. 2003;98(7):1621-6.

603. Woodward J. Bi-allelic SNP genotyping using the TaqMan(R) assay. Methods Mol Biol. 2014;1145:67-74.

604. Drogemoller BI, Wright GE, Niehaus DJ, Emsley R, Warnich L. Next-generation sequencing of pharmacogenes: a critical analysis focusing on schizophrenia treatment. Pharmacogenet Genomics. 2013;23(12):666-74.

605. Churko JM, Mantalas GL, Snyder MP, Wu JC. Overview of high throughput sequencing technologies to elucidate molecular pathways in cardiovascular diseases. Circ Res. 2013;112(12):1613-23.

606. Erguner B, Ustek D, Sagiroglu MS. Performance comparison of Next Generation sequencing platforms. Conf Proc IEEE Eng Med Biol Soc. 2015;2015:6453-6.

607. Fox EJ, Reid-Bayliss KS, Emond MJ, Loeb LA. Accuracy of Next Generation Sequencing Platforms. Next Gener Seq Appl. 2014;1.

608. Kuleshov V, Xie D, Chen R, Pushkarev D, Ma Z, Blauwkamp T, et al. Whole-genome haplotyping using long reads and statistical methods. Nat Biotechnol. 2014;32(3):261-6.

609. Ley B, Bancone G, von Seidlein L, Thriemer K, Richards JS, Domingo GJ, et al. Methods for the field evaluation of quantitative G6PD diagnostics: a review. Malar J. 2017;16(1):361.

610. Lennard L. Implementation of TPMT testing. Br J Clin Pharmacol. 2014;77(4):704-14.

611. Booth RA, Ansari MT, Loit E, Tricco AC, Weeks L, Doucette S, et al. Assessment of thiopurine S-methyltransferase activity in patients prescribed thiopurines: a systematic review. Ann Intern Med. 2011;154(12):814-23, W-295-8.

612. Hindorf U, Appell ML. Genotyping should be considered the primary choice for pre-treatment evaluation of thiopurine methyltransferase function. J Crohns Colitis. 2012;6(6):655-9.

613. Tangamornsuksan W, Lohitnavy O, Kongkaew C, Chaiyakunapruk N, Reisfeld B, Scholfield NC, et al. Association of HLA-B*5701 genotypes and abacavir-induced hypersensitivity reaction: a systematic review and meta-analysis. J Pharm Pharm Sci. 2015;18(1):68-76.

614. Tan JC, Murrell DF, Hersch MI. Genetic screening for human leukocyte antigen alleles prior to carbamazepine treatment. J Clin Neurosci. 2015;22(12):1992-3.

615. Bennett LL, Turcotte K. Eliglustat tartrate for the treatment of adults with type 1 Gaucher disease. Drug Des Devel Ther. 2015;9:4639-47.

616. Ford LT, Berg JD. Thiopurine S-methyltransferase (TPMT) assessment prior to starting thiopurine drug treatment; a pharmacogenomic test whose time has come. J Clin Pathol. 2010;63(4):288-95.

617. Pirmohamed M, Burnside G, Eriksson N, Jorgensen AL, Toh CH, Nicholson T, et al. A randomized trial of genotype-guided dosing of warfarin. N Engl J Med. 2013;369(24):2294-303.

618. Stubbins MJ, Harries LW, Smith G, Tarbit MH, Wolf CR. Genetic analysis of the human cytochrome P450 CYP2C9 locus. Pharmacogenetics. 1996;6(5):429-39.

619. Zhou Y, Ingelman-Sundberg M, Lauschke VM. Worldwide Distribution of Cytochrome P450 Alleles: A Meta-analysis of Population-scale Sequencing Projects. Clin Pharmacol Ther. 2017;102(4):688-700.

620. King BP, Leathart JB, Mutch E, Williams FM, Daly AK. CYP3A5 phenotype-genotype correlations in a British population. Br J Clin Pharmacol. 2003;55(6):625-9.

621. Caudle KE, Thorn CF, Klein TE, Swen JJ, McLeod HL, Diasio RB, et al. Clinical Pharmacogenetics Implementation Consortium guidelines for dihydropyrimidine dehydrogenase genotype and fluoropyrimidine dosing. Clin Pharmacol Ther. 2013;94(6):640-5.

622. Rees DC, Cox M, Clegg JB. World distribution of factor V Leiden. Lancet (London, England). 1995;346(8983):1133-4.

623. Erdohazi MH, W.J. Glucose-6-Phosphate-Dehydrogenase deficiency in Britain. Lancet (London, England). 1962;280(7268):1274.

624. Howes RE, Piel FB, Patil AP, Nyangiri OA, Gething PW, Dewi M, et al. G6PD deficiency prevalence and estimates of affected populations in malaria endemic countries: a geostatistical model-based map. PLoS Med. 2012;9(11):e1001339.

625. Pillai NE, Okada Y, Saw WY, Ong RT, Wang X, Tantoso E, et al. Predicting HLA alleles from high-resolution SNP data in three Southeast Asian populations. Hum Mol Genet. 2014;23(16):4443-51.

626. Liu X, Sun J, Yu H, Chen H, Wang J, Zou H, et al. Tag SNPs for HLA-B alleles that are associated with drug response and disease risk in the Chinese Han population. Pharmacogenomics J. 2015;15(5):467-72.

627. Aminkeng F, Bhavsar AP, Visscher H, Rassekh SR, Li Y, Lee JW, et al. A coding variant in RARG confers susceptibility to anthracycline-induced cardiotoxicity in childhood cancer. Nat Genet. 2015;47(9):1079-84.

628. Collie-Duguid ES, Pritchard SC, Powrie RH, Sludden J, Collier DA, Li T, et al. The frequency and distribution of thiopurine methyltransferase alleles in Caucasian and Asian populations. Pharmacogenetics. 1999;9(1):37-42.

629. Samwald M, Blagec K, Hofer S, Freimuth RR. Analyzing the potential for incorrect haplotype calls with different pharmacogenomic assays in different populations: a simulation based on 1000 Genomes data. Pharmacogenomics. 2015;16(15):1713-21.

630. Bachtiar M, Lee CGL. Genetics of Population Differences in Drug Response. Current Genetic Medicine Reports. 2013;1(3):162-70.

631. Glusman G, Cox HC, Roach JC. Whole-genome haplotyping approaches and genomic medicine. Genome Med. 2014;6(9):73.

632. Ammar R, Paton TA, Torti D, Shlien A, Bader GD. Long read nanopore sequencing for detection of HLA and CYP2D6 variants and haplotypes. F1000Res. 2015;4:17.

633. Twist GP, Gaedigk A, Miller NA, Farrow EG, Willig LK, Dinwiddie DL, et al. Constellation: a tool for rapid, automated phenotype assignment of a highly polymorphic pharmacogene, CYP2D6, from whole-genome sequences. Npj Genomic Medicine. 2016;1:15007.

634. Government HM. Ethnicity in the U.K. In: Statistics OfN, editor. 2011.

635. Zhou HH, Wood AJ. Stereoselective disposition of carvedilol is determined by CYP2D6. Clin Pharmacol Ther. 1995;57(5):518-24.

636. Reisberg S, Krebs K, Lepamets M, Kals M, Magi R, Metsalu K, et al. Translating genotype data of 44,000 biobank participants into clinical pharmacogenetic recommendations: challenges and solutions. Genet Med. 2018.

637. He Y, Hoskins JM, Clark S, Campbell NH, Wagner K, Motsinger-Reif AA, et al. Accuracy of SNPs to predict risk of HLA alleles associated with drug-induced hypersensitivity events across racial groups. Pharmacogenomics. 2015;16(8):817-24.

638. Temple NJ, Fraser J. Food labels: a critical assessment. Nutrition. 2014;30(3):257-60.

639. van der Wouden CH, Cambon-Thomsen A, Cecchin E, Cheung KC, Davila-Fajardo CL, Deneer VH, et al. Implementing Pharmacogenomics in Europe: Design and Implementation Strategy of the Ubiquitous Pharmacogenomics Consortium. Clin Pharmacol Ther. 2017;101(3):341-58.

640. Obach RS, Cox LM, Tremaine LM. Sertraline is metabolized by multiple cytochrome P450 enzymes, monoamine oxidases, and glucuronyl transferases in human: an in vitro study. Drug Metab Dispos. 2005;33(2):262-70.

641. Alfaro CL, Lam YW, Simpson J, Ereshefsky L. CYP2D6 inhibition by fluoxetine, paroxetine, sertraline, and venlafaxine in a crossover study: intraindividual variability and plasma concentration correlations. J Clin Pharmacol. 2000;40(1):58-66.

642. Fenech GAG, G. Pharmacogenetics and personalized medicine: does gender have a role? Journal of the Malta College of Pharmacy Practice 2014;20:7-10.

643. Reichwagen A, Ziepert M, Kreuz M, Godtel-Armbrust U, Rixecker T, Poeschel V, et al. Association of NADPH oxidase polymorphisms with anthracycline-induced cardiotoxicity in the RICOVER-60 trial of patients with aggressive CD20(+) B-cell lymphoma. Pharmacogenomics. 2015;16(4):361-72.

644. Vulsteke C, Pfeil AM, Maggen C, Schwenkglenks M, Pettengell R, Szucs TD, et al. Clinical and genetic risk factors for epirubicin-induced cardiac toxicity in early breast cancer patients. Breast Cancer Res Treat. 2015;152(1):67-76.

645. Koren G, Cairns J, Chitayat D, Gaedigk A, Leeder SJ. Pharmacogenetics of morphine poisoning in a breastfed neonate of a codeine-prescribed mother. Lancet (London, England). 2006;368(9536):704.

646. Kearns GL, Abdel-Rahman SM, Alander SW, Blowey DL, Leeder JS, Kauffman RE. Developmental pharmacology--drug disposition, action, and therapy in infants and children. N Engl J Med. 2003;349(12):1157-67.

647. Hawcutt DB, Thompson B, Smyth RL, Pirmohamed M. Paediatric pharmacogenomics: an overview. Arch Dis Child. 2013;98(3):232-7.

648. Zhao W, Leroux S, Biran V, Jacqz-Aigrain E. Developmental pharmacogenetics of CYP2C19 in neonates and young infants: omeprazole as a probe drug. Br J Clin Pharmacol. 2018;84(5):997-1005.

649. Maagdenberg H, Vijverberg SJ, Bierings MB, Carleton BC, Arets HG, de Boer A, et al. Pharmacogenomics in Pediatric Patients: Towards Personalized Medicine. Paediatr Drugs. 2016;18(4):251-60.

650. Blaisdell J, Jorge-Nebert LF, Coulter S, Ferguson SS, Lee SJ, Chanas B, et al. Discovery of new potentially defective alleles of human CYP2C9. Pharmacogenetics. 2004;14(8):527-37.

651. Allabi AC, Gala JL, Horsmans Y. CYP2C9, CYP2C19, ABCB1 (MDR1) genetic polymorphisms and phenytoin metabolism in a Black Beninese population. Pharmacogenet Genomics. 2005;15(11):779-86.

652. Liu Y, Jeong H, Takahashi H, Drozda K, Patel SR, Shapiro NL, et al. Decreased warfarin clearance associated with the CYP2C9 R150H (*8) polymorphism. Clin Pharmacol Ther. 2012;91(4):660-5.

653. McLaughlin HM, Sakaguchi R, Giblin W, Program NCS, Wilson TE, Biesecker L, et al. A recurrent loss-of-function alanyl-tRNA synthetase (AARS) mutation in patients with Charcot-Marie-Tooth disease type 2N (CMT2N). Human mutation. 2012;33(1):244-53.

654. Simons C, Griffin LB, Helman G, Golas G, Pizzino A, Bloom M, et al. Loss-of-function alanyl-tRNA synthetase mutations cause an autosomal-recessive early-onset epileptic encephalopathy with persistent myelination defect. American journal of human genetics. 2015;96(4):675-81.

655. Yang XR, Xiong Y, Duan H, Gong RR. Identification of genes associated with methotrexate resistance in methotrexate-resistant osteosarcoma cell lines. J Orthop Surg Res. 2015;10:136.

656. Juliano RL, Ling V. A surface glycoprotein modulating drug permeability in Chinese hamster ovary cell mutants. Biochimica et biophysica acta. 1976;455(1):152-62.

657. Bodor M, Kelly EJ, Ho RJ. Characterization of the human MDR1 gene. AAPS J. 2005;7(1):E1-5.

658. Strautnieks SS, Bull LN, Knisely AS, Kocoshis SA, Dahl N, Arnell H, et al. A gene encoding a liver-specific ABC transporter is mutated in progressive familial intrahepatic cholestasis. Nat Genet. 1998;20(3):233-8.

659. Ulzurrun E, Stephens C, Crespo E, Ruiz-Cabello F, Ruiz-Nunez J, Saenz-Lopez P, et al. Role of chemical structures and the 1331T>C bile salt export pump polymorphism in idiosyncratic drug-induced liver injury. Liver Int. 2013;33(9):1378-85.

660. Visscher H, Rassekh SR, Sandor GS, Caron HN, van Dalen EC, Kremer LC, et al. Genetic variants in SLC22A17 and SLC22A7 are associated with anthracycline-induced cardiotoxicity in children. Pharmacogenomics. 2015;16(10):1065-76.

661. Fohner AE, Brackman DJ, Giacomini KM, Altman RB, Klein TE. PharmGKB summary: very important pharmacogene information for ABCG2. Pharmacogenet Genomics. 2017;27(11):420-7.

662. Zhao J, Li W, Zhu D, Yu Q, Zhang Z, Sun M, et al. Association of single nucleotide polymorphisms in MTHFR and ABCG2 with the different efficacy of first-line chemotherapy in metastatic colorectal cancer. Med Oncol. 2014;31(1):802.

663. DeGorter MK, Tirona RG, Schwarz UI, Choi YH, Dresser GK, Suskin N, et al. Clinical and pharmacogenetic predictors of circulating atorvastatin and rosuvastatin concentrations in routine clinical care. Circ Cardiovasc Genet. 2013;6(4):400-8.

664. De Mattia E, Toffoli G, Polesel J, D'Andrea M, Corona G, Zagonel V, et al. Pharmacogenetics of ABC and SLC transporters in metastatic colorectal cancer patients receiving first-line FOLFIRI treatment. Pharmacogenet Genomics. 2013;23(10):549-57.

665. Stamp LK, Chapman PT, O'Donnell JL, Zhang M, James J, Frampton C, et al. Polymorphisms within the folate pathway predict folate concentrations but are not associated with disease activity in rheumatoid arthritis patients on methotrexate. Pharmacogenet Genomics. 2010;20(6):367-76.

666. Wen CC, Yee SW, Liang X, Hoffmann TJ, Kvale MN, Banda Y, et al. Genome-wide association study identifies ABCG2 (BCRP) as an allopurinol transporter and a determinant of drug response. Clin Pharmacol Ther. 2015;97(5):518-25.

667. Kobilka BK, Matsui H, Kobilka TS, Yang-Feng TL, Francke U, Caron MG, et al. Cloning, sequencing, and expression of the gene coding for the human platelet alpha 2-adrenergic receptor. Science (New York, NY). 1987;238(4827):650-6.

668. Comings DE, Gade-Andavolu R, Gonzalez N, Blake H, Wu S, MacMurray JP. Additive effect of three noradrenergic genes (ADRA2a, ADRA2C, DBH) on attention-deficit hyperactivity disorder and learning disabilities in Tourette syndrome subjects. Clin Genet. 1999;55(3):160-72.

669. Rosengren AH, Jokubka R, Tojjar D, Granhall C, Hansson O, Li DQ, et al. Overexpression of alpha2A-adrenergic receptors contributes to type 2 diabetes. Science (New York, NY). 2010;327(5962):217-20.

670. McCracken JT, Badashova KK, Posey DJ, Aman MG, Scahill L, Tierney E, et al. Positive effects of methylphenidate on hyperactivity are moderated by monoaminergic gene variants in children with autism spectrum disorders. Pharmacogenomics J. 2014;14(3):295-302.

671. Liljedahl U, Kahan T, Malmqvist K, Melhus H, Syvanen AC, Lind L, et al. Single nucleotide polymorphisms predict the change in left ventricular mass in response to antihypertensive treatment. J Hypertens. 2004;22(12):2321-8.

672. Yin L, Zhang YY, Zhang X, Yu T, He G, Sun XL. TPH, SLC6A2, SLC6A3, DRD2 and DRD4 Polymorphisms and Neuroendocrine Factors Predict SSRIs Treatment Outcome in the Chinese Population with Major Depression. Pharmacopsychiatry. 2015;48(3):95-103.

673. Ranade K, Jorgenson E, Sheu WH, Pei D, Hsiung CA, Chiang FT, et al. A polymorphism in the beta1 adrenergic receptor is associated with resting heart rate. American journal of human genetics. 2002;70(4):935-42.

674. Metra M, Zani C, Covolo L, Nodari S, Pezzali N, Gelatti U, et al. Role of beta1- and alpha2c-adrenergic receptor polymorphisms and their combination in heart failure: a case-control study. Eur J Heart Fail. 2006;8(2):131-5.

675. Wenzel K, Felix SB, Bauer D, Heere P, Flachmeier C, Podlowski S, et al. Novel variants in 3 kb of 5'UTR of the beta 1-adrenergic receptor gene (-93C>T, -210C>T, and -2146T>C): -2146C homozygotes present in patients with idiopathic dilated cardiomyopathy and coronary heart disease. Human mutation. 2000;16(6):534.

676. Liu J, Liu ZQ, Yu BN, Xu FH, Mo W, Zhou G, et al. beta1-Adrenergic receptor polymorphisms influence the response to metoprolol monotherapy in patients with essential hypertension. Clin Pharmacol Ther. 2006;80(1):23-32.

677. Kim KM, Murray MD, Tu W, Robarge J, Ding Y, Brater DC, et al. Pharmacogenetics and healthcare outcomes in patients with chronic heart failure. Eur J Clin Pharmacol. 2012;68(11):1483-91.

678. Hawkins GA, Weiss ST, Bleecker ER. Clinical consequences of ADRbeta2 polymorphisms. Pharmacogenomics. 2008;9(3):349-58.

679. Reihsaus E, Innis M, MacIntyre N, Liggett SB. Mutations in the gene encoding for the beta 2-adrenergic receptor in normal and asthmatic subjects. Am J Respir Cell Mol Biol. 1993;8(3):334-9.

680. Iaccarino G, Trimarco V, Lanni F, Cipolletta E, Izzo R, Arcucci O, et al. beta-Blockade and increased dyslipidemia in patients bearing Glu27 variant of beta2 adrenergic receptor gene. Pharmacogenomics J. 2005;5(5):292-7.

681. Palmer CN, Lipworth BJ, Lee S, Ismail T, Macgregor DF, Mukhopadhyay S. Arginine-16 beta2 adrenoceptor genotype predisposes to exacerbations in young asthmatics taking regular salmeterol. Thorax. 2006;61(11):940-4.

682. Lipworth BJ, Basu K, Donald HP, Tavendale R, Macgregor DF, Ogston SA, et al. Tailored second-line therapy in asthmatic children with the Arg(16) genotype. Clin Sci (Lond). 2013;124(8):521-8.

683. Bonnardeaux A, Davies E, Jeunemaitre X, Fery I, Charru A, Clauser E, et al. Angiotensin II type 1 receptor gene polymorphisms in human essential hypertension. Hypertension. 1994;24(1):63-9.

684. Gribouval O, Gonzales M, Neuhaus T, Aziza J, Bieth E, Laurent N, et al. Mutations in genes in the renin-angiotensin system are associated with autosomal recessive renal tubular dysgenesis. Nat Genet. 2005;37(9):964-8.

685. Murphy TJ, Alexander RW, Griendling KK, Runge MS, Bernstein KE. Isolation of a cDNA encoding the vascular type-1 angiotensin II receptor. Nature. 1991;351(6323):233-6.

686. Bozkurt O, de Boer A, Grobbee DE, de Leeuw PW, Kroon AA, Schiffers P, et al. Variation in Renin-Angiotensin system and salt-sensitivity genes and the risk of diabetes mellitus associated with the use of thiazide diuretics. Am J Hypertens. 2009;22(5):545-51.

687. Miller JA, Thai K, Scholey JW. Angiotensin II type 1 receptor gene polymorphism predicts response to losartan and angiotensin II. Kidney Int. 1999;56(6):2173-80.

688. Brugts JJ, Boersma E, Simoons ML. Tailored therapy of ACE inhibitors in stable coronary artery disease: pharmacogenetic profiling of treatment benefit. Pharmacogenomics. 2010;11(8):1115-26.

689. Neville MJ, Johnstone EC, Walton RT. Identification and characterization of ANKK1: a novel kinase gene closely linked to DRD2 on chromosome band 11q23.1. Human mutation. 2004;23(6):540-5.

690. Li H, Wang X, Zhou Y, Ni G, Su Q, Chen Z, et al. Association of LEPR and ANKK1 Gene Polymorphisms with Weight Gain in Epilepsy Patients Receiving Valproic Acid. Int J Neuropsychopharmacol. 2015;18(7):pyv021.

691. Roke Y, van Harten PN, Franke B, Galesloot TE, Boot AM, Buitelaar JK. The effect of the Taq1A variant in the dopamine D(2) receptor gene and common CYP2D6 alleles on prolactin levels in risperidone-treated boys. Pharmacogenet Genomics. 2013;23(9):487-93.

692. Swan GE, Valdes AM, Ring HZ, Khroyan TV, Jack LM, Ton CC, et al. Dopamine receptor DRD2 genotype and smoking cessation outcome following treatment with bupropion SR. Pharmacogenomics J. 2005;5(1):21-9.

693. Savitsky K, Bar-Shira A, Gilad S, Rotman G, Ziv Y, Vanagaite L, et al. A single ataxia telangiectasia gene with a product similar to PI-3 kinase. Science (New York, NY). 1995;268(5218):1749-53.

694. Out M, Becker ML, van Schaik RH, Lehert P, Stehouwer CD, Kooy A. A gene variant near ATM affects the response to metformin and metformin plasma levels: a post hoc analysis of an RCT. Pharmacogenomics. 2018;19(8):715-26.

695. Souza DG, Lomez ES, Pinho V, Pesquero JB, Bader M, Pesquero JL, et al. Role of bradykinin B2 and B1 receptors in the local, remote, and systemic inflammatory responses that follow intestinal ischemia and reperfusion injury. J Immunol. 2004;172(4):2542-8.

696. Splawski I, Timothy KW, Sharpe LM, Decher N, Kumar P, Bloise R, et al. Ca(V)1.2 calcium channel dysfunction causes a multisystem disorder including arrhythmia and autism. Cell. 2004;119(1):19-31.

697. Antzelevitch C, Pollevick GD, Cordeiro JM, Casis O, Sanguinetti MC, Aizawa Y, et al. Loss-of-function mutations in the cardiac calcium channel underlie a new clinical entity characterized by ST-segment elevation, short QT intervals, and sudden cardiac death. Circulation. 2007;115(4):442-9.

698. Beitelshees AL, Navare H, Wang D, Gong Y, Wessel J, Moss JI, et al. CACNA1C gene polymorphisms, cardiovascular disease outcomes, and treatment response. Circ Cardiovasc Genet. 2009;2(4):362-70.

699. Bremer T, Man A, Kask K, Diamond C. CACNA1C polymorphisms are associated with the efficacy of calcium channel blockers in the treatment of hypertension. Pharmacogenomics. 2006;7(3):271-9.

700. Casamassima F, Huang J, Fava M, Sachs GS, Smoller JW, Cassano GB, et al. Phenotypic effects of a bipolar liability gene among individuals with major depressive disorder. Am J Med Genet B Neuropsychiatr Genet. 2010;153B(1):303-9.

701. Ohmori O, Shinkai T, Kojima H, Terao T, Suzuki T, Mita T, et al. Association study of a functional catechol-O-methyltransferase gene polymorphism in Japanese schizophrenics. Neurosci Lett. 1998;243(1-3):109-12.

702. Inada T, Nakamura A, Iijima Y. Relationship between catechol-O-methyltransferase polymorphism and treatment-resistant schizophrenia. Am J Med Genet B Neuropsychiatr Genet. 2003;120B(1):35-9.

703. Vandenbergh DJ, Rodriguez LA, Miller IT, Uhl GR, Lachman HM. High-activity catechol-O-methyltransferase allele is more prevalent in polysubstance abusers. Am J Med Genet. 1997;74(4):439-42.

704. Shimada T, Yamazaki H, Mimura M, Inui Y, Guengerich FP. Interindividual variations in human liver cytochrome P-450 enzymes involved in the oxidation of drugs, carcinogens and toxic chemicals: studies with liver microsomes of 30 Japanese and 30 Caucasians. J Pharmacol Exp Ther. 1994;270(1):414-23.

705. Perera V, Gross AS, Polasek TM, Qin Y, Rao G, Forrest A, et al. Considering CYP1A2 phenotype and genotype for optimizing the dose of olanzapine in the management of schizophrenia. Expert Opin Drug Metab Toxicol. 2013;9(9):1115-37.

706. Perera V, Gross AS, McLachlan AJ. Measurement of CYP1A2 activity: a focus on caffeine as a probe. Curr Drug Metab. 2012;13(5):667-78.

707. Sinxadi PZ, Leger PD, McIlleron HM, Smith PJ, Dave JA, Levitt NS, et al. Pharmacogenetics of plasma efavirenz exposure in HIV-infected adults and children in South Africa. Br J Clin Pharmacol. 2015;80(1):146-56.

708. Gao Y, Zhang LR, Fu Q. CYP3A4*1G polymorphism is associated with lipid-lowering efficacy of atorvastatin but not of simvastatin. Eur J Clin Pharmacol. 2008;64(9):877-82.

709. McKinney EF, Walton RT, Yudkin P, Fuller A, Haldar NA, Mant D, et al. Association between polymorphisms in dopamine metabolic enzymes and tobacco consumption in smokers. Pharmacogenetics. 2000;10(6):483-91.

710. Randesi M, van den Brink W, Levran O, Yuferov V, Blanken P, van Ree JM, et al. Dopamine gene variants in opioid addiction: comparison of dependent patients, nondependent users and healthy controls. Pharmacogenomics. 2018;19(2):95-104.

711. Arias AJ, Gelernter J, Gueorguieva R, Ralevski E, Petrakis IL. Pharmacogenetics of naltrexone and disulfiram in alcohol dependent, dually diagnosed veterans. Am J Addict. 2014;23(3):288-93.

712. Lencz T, Robinson DG, Xu K, Ekholm J, Sevy S, Gunduz-Bruce H, et al. DRD2 promoter region variation as a predictor of sustained response to antipsychotic medication in first-episode schizophrenia patients. Am J Psychiatry. 2006;163(3):529-31.

713. Mi H, Thomas PD, Ring HZ, Jiang R, Sangkuhl K, Klein TE, et al. PharmGKB summary: dopamine receptor D2. Pharmacogenet Genomics. 2011;21(6):350-6.

714. Nyholm D. Pharmacokinetic optimisation in the treatment of Parkinson's disease : an update. Clin Pharmacokinet. 2006;45(2):109-36.

715. Muller DJ, Zai CC, Sicard M, Remington E, Souza RP, Tiwari AK, et al. Systematic analysis of dopamine receptor genes (DRD1-DRD5) in antipsychotic-induced weight gain. Pharmacogenomics J. 2012;12(2):156-64.

716. Sullivan D, Pinsonneault JK, Papp AC, Zhu H, Lemeshow S, Mash DC, et al. Dopamine transporter DAT and receptor DRD2 variants affect risk of lethal cocaine abuse: a gene-gene-environment interaction. Transl Psychiatry. 2013;3:e222.

717. Crettol S, Besson J, Croquette-Krokar M, Hammig R, Gothuey I, Monnat M, et al. Association of dopamine and opioid receptor genetic polymorphisms with response to methadone maintenance treatment. Prog Neuropsychopharmacol Biol Psychiatry. 2008;32(7):1722-7.

718. Lancellotti S, De Cristofaro R. Congenital prothrombin deficiency. Seminars in thrombosis and hemostasis. 2009;35(4):367-81.

719. Lefkowitz JB, Haver T, Clarke S, Jacobson L, Weller A, Nuss R, et al. The prothrombin Denver patient has two different prothrombin point mutations resulting in Glu-300-->Lys and Glu-309-->Lys substitutions. British journal of haematology. 2000;108(1):182-7.

720. Miyawaki Y, Suzuki A, Fujita J, Maki A, Okuyama E, Murata M, et al. Thrombosis from a prothrombin mutation conveying antithrombin resistance. N Engl J Med. 2012;366(25):2390-6.

721. Legnani C, Palareti G, Guazzaloca G, Cosmi B, Lunghi B, Bernardi F, et al. Venous thromboembolism in young women; role of thrombophilic mutations and oral contraceptive use. Eur Heart J. 2002;23(12):984-90.

722. D'Ambrosio RL, D'Andrea G, Cappucci F, Chetta M, Di Perna P, Brancaccio V, et al. Polymorphisms in factor II and factor VII genes modulate oral anticoagulation with warfarin. Haematologica. 2004;89(12):1510-6.

723. Baughman G, Wiederrecht GJ, Campbell NF, Martin MM, Bourgeois S. FKBP51, a novel T-cell-specific immunophilin capable of calcineurin inhibition. Mol Cell Biol. 1995;15(8):4395-402.

724. Zou YF, Wang F, Feng XL, Li WF, Tao JH, Pan FM, et al. Meta-analysis of FKBP5 gene polymorphisms association with treatment response in patients with mood disorders. Neurosci Lett. 2010;484(1):56-61.

725. Maltese P, Palma L, Sfara C, de Rocco P, Latiano A, Palmieri O, et al. Glucocorticoid resistance in Crohn's disease and ulcerative colitis: an association study investigating GR and FKBP5 gene polymorphisms. Pharmacogenomics J. 2012;12(5):432-8.

726. Szpirer C, Molne M, Antonacci R, Jenkins NA, Finelli P, Szpirer J, et al. The genes encoding the glutamate receptor subunits KA1 and KA2 (GRIK4 and GRIK5) are located on separate chromosomes in human, mouse, and rat. Proc Natl Acad Sci U S A. 1994;91(25):11849-53.

727. Pu M, Zhang Z, Xu Z, Shi Y, Geng L, Yuan Y, et al. Influence of genetic polymorphisms in the glutamatergic and GABAergic systems and their interactions with environmental stressors on antidepressant response. Pharmacogenomics. 2013;14(3):277-88.

728. Matta JA, Ashby MC, Sanz-Clemente A, Roche KW, Isaac JT. mGluR5 and NMDA receptors drive the experience- and activity-dependent NMDA receptor NR2B to NR2A subunit switch. Neuron. 2011;70(2):339-51.

729. Lemke JR, Hendrickx R, Geider K, Laube B, Schwake M, Harvey RJ, et al. GRIN2B mutations in West syndrome and intellectual disability with focal epilepsy. Ann Neurol. 2014;75(1):147-54.

730. Endele S, Rosenberger G, Geider K, Popp B, Tamer C, Stefanova I, et al. Mutations in GRIN2A and GRIN2B encoding regulatory subunits of NMDA receptors cause variable neurodevelopmental phenotypes. Nat Genet. 2010;42(11):1021-6.

731. Lopez-Rodriguez R, Cabaleiro T, Ochoa D, Roman M, Borobia AM, Carcas AJ, et al. Pharmacodynamic genetic variants related to antipsychotic adverse reactions in healthy volunteers. Pharmacogenomics. 2013;14(10):1203-14.

732. Taylor DL, Tiwari AK, Lieberman JA, Potkin SG, Meltzer HY, Knight J, et al. Genetic association analysis of N-methyl-D-aspartate receptor subunit gene GRIN2B and clinical response to clozapine. Hum Psychopharmacol. 2016;31(2):121-34.

733. Michal AM, So CH, Beeharry N, Shankar H, Mashayekhi R, Yen TJ, et al. G Protein-coupled receptor kinase 5 is localized to centrosomes and regulates cell cycle progression. The Journal of biological chemistry. 2012;287(9):6928-40.

734. Liggett SB, Cresci S, Kelly RJ, Syed FM, Matkovich SJ, Hahn HS, et al. A GRK5 polymorphism that inhibits beta-adrenergic receptor signaling is protective in heart failure. Nat Med. 2008;14(5):510-7.

735. Ji Y, Biernacka JM, Hebbring S, Chai Y, Jenkins GD, Batzler A, et al. Pharmacogenomics of selective serotonin reuptake inhibitor treatment for major depressive disorder: genome-wide associations and functional genomics. Pharmacogenomics J. 2013;13(5):456-63.

736. Mas S, Gasso P, Lafuente A, Bioque M, Lobo A, Gonzalez-Pinto A, et al. Pharmacogenetic study of antipsychotic induced acute extrapyramidal symptoms in a first episode psychosis cohort: role of dopamine, serotonin and glutamate candidate genes. Pharmacogenomics J. 2016;16(5):439-45.

737. Viikki M, Huuhka K, Leinonen E, Illi A, Setala-Soikkeli E, Huuhka M, et al. Interaction between two HTR2A polymorphisms and gender is associated with treatment response in MDD. Neurosci Lett. 2011;501(1):20-4.

738. Cargnin S, Viana M, Sances G, Bianchi M, Ghiotto N, Tassorelli C, et al. Combined effect of common gene variants on response to drug withdrawal therapy in medication overuse headache. Eur J Clin Pharmacol. 2014;70(10):1195-202.

739. Mulder H, Franke B, van der-Beek van der AA, Arends J, Wilmink FW, Scheffer H, et al. The association between HTR2C gene polymorphisms and the metabolic syndrome in patients with schizophrenia. J Clin Psychopharmacol. 2007;27(4):338-43.

740. Prokunina-Olsson L, Muchmore B, Tang W, Pfeiffer RM, Park H, Dickensheets H, et al. A variant upstream of IFNL3 (IL28B) creating a new interferon gene IFNL4 is associated with impaired clearance of hepatitis C virus. Nat Genet. 2013;45(2):164-71.

741. Stenkvist J, Sonnerborg A, Weiland O. HCV RNA decline in chronic HCV genotype 2 and 3 during standard of care treatment according to IL28B polymorphism. J Viral Hepat. 2013;20(3):193-9.

742. Meissner EG, Bon D, Prokunina-Olsson L, Tang W, Masur H, O'Brien TR, et al. IFNL4-DeltaG genotype is associated with slower viral clearance in hepatitis C, genotype-1 patients treated with sofosbuvir and ribavirin. J Infect Dis. 2014;209(11):1700-4.

743. Mori N, Imamura M, Kawakami Y, Nagaoki Y, Kawaoka T, Tsuge M, et al. IFNL4 polymorphism effects on outcome of simeprevir, peginterferon, and ribavirin therapy for older patients with genotype 1 chronic hepatitis C. Hepatol Res. 2017;47(3):E5-E13.

744. Bajt ML, Ginsberg MH, Frelinger AL, 3rd, Berndt MC, Loftus JC. A spontaneous mutation of integrin alpha IIb beta 3 (platelet glycoprotein IIb-IIIa) helps define a ligand binding site. The Journal of biological chemistry. 1992;267(6):3789-94.

745. Chen YP, Djaffar I, Pidard D, Steiner B, Cieutat AM, Caen JP, et al. Ser-752-->Pro mutation in the cytoplasmic domain of integrin beta 3 subunit and defective activation of platelet integrin alpha IIb beta 3 (glycoprotein IIb-IIIa) in a variant of Glanzmann thrombasthenia. Proc Natl Acad Sci U S A. 1992;89(21):10169-73.

746. Gresele P, Falcinelli E, Giannini S, D'Adamo P, D'Eustacchio A, Corazzi T, et al. Dominant inheritance of a novel integrin beta3 mutation associated with a hereditary macrothrombocytopenia and platelet dysfunction in two Italian families. Haematologica. 2009;94(5):663-9.

747. Simon T, Verstuyft C, Mary-Krause M, Quteineh L, Drouet E, Meneveau N, et al. Genetic determinants of response to clopidogrel and cardiovascular events. N Engl J Med. 2009;360(4):363-75.

748. Dropinski J, Musial J, Sanak M, Wegrzyn W, Nizankowski R, Szczeklik A. Antithrombotic effects of aspirin based on PLA1/A2 glycoprotein IIIa polymorphism in patients with coronary artery disease. Thromb Res. 2007;119(3):301-3.

749. Sakura H, Bond C, Warren-Perry M, Horsley S, Kearney L, Tucker S, et al. Characterization and variation of a human inwardly-rectifying-K-channel gene (KCNJ6): a putative ATP-sensitive K-channel subunit. FEBS Lett. 1995;367(2):193-7.

750. Yoshida K, Nishizawa D, Ichinomiya T, Ichinohe T, Hayashida M, Fukuda K, et al. Prediction formulas for individual opioid analgesic requirements based on genetic polymorphism analyses. PloS one. 2015;10(1):e0116885.

751. Nishizawa D, Nagashima M, Katoh R, Satoh Y, Tagami M, Kasai S, et al. Association between KCNJ6 (GIRK2) gene polymorphisms and postoperative analgesic requirements after major abdominal surgery. PloS one. 2009;4(9):e7060.

752. Li Y, Sabatine MS, Tong CH, Ford I, Kirchgessner TG, Packard CJ, et al. Genetic variants in the KIF6 region and coronary event reduction from statin therapy. Human genetics. 2011;129(1):17-23.

753. Borges L, Hsu ML, Fanger N, Kubin M, Cosman D. A family of human lymphoid and myeloid Ig-like receptors, some of which bind to MHC class I molecules. J Immunol. 1997;159(11):5192-6.

754. K Siddiqui M, Maroteau C, Veluchamy A, Tornio A, Tavendale R, Carr F, et al. A common missense variant of LILRB5 is associated with statin intolerance and myalgia. Eur Heart J. 2017;38(48):3569-75.

755. Nordestgaard BG, Chapman MJ, Ray K, Boren J, Andreotti F, Watts GF, et al. Lipoprotein(a) as a cardiovascular risk factor: current status. Eur Heart J. 2010;31(23):2844-53.

756. Donnelly LA, van Zuydam NR, Zhou K, Tavendale R, Carr F, Maitland-van der Zee AH, et al. Robust association of the LPA locus with low-density lipoprotein cholesterol lowering response to statin treatment in a meta-analysis of 30 467 individuals from both randomized control trials and observational studies and association with coronary artery disease outcome during statin treatment. Pharmacogenet Genomics. 2013;23(10):518-25.

757. Goyette P, Sumner JS, Milos R, Duncan AM, Rosenblatt DS, Matthews RG, et al. Human methylenetetrahydrofolate reductase: isolation of cDNA, mapping and mutation identification. Nat Genet. 1994;7(2):195-200.

758. Cui LH, Yu Z, Zhang TT, Shin MH, Kim HN, Choi JS. Influence of polymorphisms in MTHFR 677 C-->T, TYMS 3R-->2R and MTR 2756 A-->G on NSCLC risk and response to platinum-based chemotherapy in advanced NSCLC. Pharmacogenomics. 2011;12(6):797-808.

759. Kao AC, Rojnic Kuzman M, Tiwari AK, Zivkovic MV, Chowdhury NI, Medved V, et al. Methylenetetrahydrofolate reductase gene variants and antipsychotic-induced weight gain and metabolic disturbances. J Psychiatr Res. 2014;54:36-42.

760. Arking DE, Pfeufer A, Post W, Kao WH, Newton-Cheh C, Ikeda M, et al. A common genetic variant in the NOS1 regulator NOS1AP modulates cardiac repolarization. Nat Genet. 2006;38(6):644-51.

761. Jamshidi Y, Nolte IM, Dalageorgou C, Zheng D, Johnson T, Bastiaenen R, et al. Common variation in the NOS1AP gene is associated with drug-induced QT prolongation and ventricular arrhythmia. J Am Coll Cardiol. 2012;60(9):841-50.

762. van Noord C, Aarnoudse AJ, Eijgelsheim M, Sturkenboom MC, Straus SM, Hofman A, et al. Calcium channel blockers, NOS1AP, and heart-rate-corrected QT prolongation. Pharmacogenet Genomics. 2009;19(4):260-6.

763. Forstermann U, Munzel T. Endothelial nitric oxide synthase in vascular disease: from marvel to menace. Circulation. 2006;113(13):1708-14.

764. Casadei Gardini A, Marisi G, Faloppi L, Scarpi E, Foschi FG, Iavarone M, et al. eNOS polymorphisms and clinical outcome in advanced HCC patients receiving sorafenib: final results of the ePHAS study. Oncotarget. 2016;7(19):27988-99.

765. Choi JY, Barlow WE, Albain KS, Hong CC, Blanco JG, Livingston RB, et al. Nitric oxide synthase variants and disease-free survival among treated and untreated breast cancer patients in a Southwest Oncology Group clinical trial. Clin Cancer Res. 2009;15(16):5258-66.

766. Jordan BA, Devi LA. G-protein-coupled receptor heterodimerization modulates receptor function. Nature. 1999;399(6737):697-700.

767. Crist RC, Clarke TK, Ang A, Ambrose-Lanci LM, Lohoff FW, Saxon AJ, et al. An intronic variant in OPRD1 predicts treatment outcome for opioid dependence in African-Americans. Neuropsychopharmacology. 2013;38(10):2003-10.

768. Beer B, Erb R, Pavlic M, Ulmer H, Giacomuzzi S, Riemer Y, et al. Association of polymorphisms in pharmacogenetic candidate genes (OPRD1, GAL, ABCB1, OPRM1) with opioid dependence in European population: a case-control study. PloS one. 2013;8(9):e75359.

769. Clarke TK, Crist RC, Ang A, Ambrose-Lanci LM, Lohoff FW, Saxon AJ, et al. Genetic variation in OPRD1 and the response to treatment for opioid dependence with buprenorphine in European-American females. Pharmacogenomics J. 2014;14(3):303-8.

770. Campa D, Gioia A, Tomei A, Poli P, Barale R. Association of ABCB1/MDR1 and OPRM1 gene polymorphisms with morphine pain relief. Clin Pharmacol Ther. 2008;83(4):559-66.

771. Crystal HA, Hamon S, Randesi M, Cook J, Anastos K, Lazar J, et al. A C17T polymorphism in the mu opiate receptor is associated with quantitative measures of drug use in African American women. Addict Biol. 2012;17(1):181-91.

772. Van Goethem G, Schwartz M, Lofgren A, Dermaut B, Van Broeckhoven C, Vissing J. Novel POLG mutations in progressive external ophthalmoplegia mimicking mitochondrial neurogastrointestinal encephalomyopathy. Eur J Hum Genet. 2003;11(7):547-9.

773. Naviaux RK, Nguyen KV. POLG mutations associated with Alpers' syndrome and mitochondrial DNA depletion. Ann Neurol. 2004;55(5):706-12.

774. Van Goethem G, Dermaut B, Lofgren A, Martin JJ, Van Broeckhoven C. Mutation of POLG is associated with progressive external ophthalmoplegia characterized by mtDNA deletions. Nat Genet. 2001;28(3):211-2.

775. Stewart JD, Horvath R, Baruffini E, Ferrero I, Bulst S, Watkins PB, et al. Polymerase gamma gene POLG determines the risk of sodium valproate-induced liver toxicity. Hepatology. 2010;52(5):1791-6.

776. Fluck CE, Tajima T, Pandey AV, Arlt W, Okuhara K, Verge CF, et al. Mutant P450 oxidoreductase causes disordered steroidogenesis with and without Antley-Bixler syndrome. Nat Genet. 2004;36(3):228-30.

777. Elens L, Hesselink DA, Bouamar R, Budde K, de Fijter JW, De Meyer M, et al. Impact of POR*28 on the pharmacokinetics of tacrolimus and cyclosporine A in renal transplant patients. Ther Drug Monit. 2014;36(1):71-9.

778. Cal S, Peinado JR, Llamazares M, Quesada V, Moncada-Pazos A, Garabaya C, et al. Identification and characterization of human polyserase-3, a novel protein with tandem serine-protease domains in the same polypeptide chain. BMC Biochem. 2006;7:9.

779. Schelleman H, Brensinger CM, Chen J, Finkelman BS, Rieder MJ, Kimmel SE. New genetic variant that might improve warfarin dose prediction in African Americans. Br J Clin Pharmacol. 2010;70(3):393-9.

780. Ramamoorthy S, Bauman AL, Moore KR, Han H, Yang-Feng T, Chang AS, et al. Antidepressant- and cocaine-sensitive human serotonin transporter: molecular cloning, expression, and chromosomal localization. Proc Natl Acad Sci U S A. 1993;90(6):2542-6.

781. Strohmaier J, Wust S, Uher R, Henigsberg N, Mors O, Hauser J, et al. Sexual dysfunction during treatment with serotonergic and noradrenergic antidepressants: clinical description and the role of the 5-HTTLPR. World J Biol Psychiatry. 2011;12(7):528-38.

782. Whale R, Quested DJ, Laver D, Harrison PJ, Cowen PJ. Serotonin transporter (5-HTT) promoter genotype may influence the prolactin response to clomipramine. Psychopharmacology (Berl). 2000;150(1):120-2.

783. Kim DK, Lim SW, Lee S, Sohn SE, Kim S, Hahn CG, et al. Serotonin transporter gene polymorphism and antidepressant response. Neuroreport. 2000;11(1):215-9.

784. Tukey RH, Strassburg CP. Human UDP-glucuronosyltransferases: metabolism, expression, and disease. Annu Rev Pharmacol Toxicol. 2000;40:581-616.

785. Mei S, Feng W, Zhu L, Yu Y, Yang W, Gao B, et al. Genetic polymorphisms and valproic acid plasma concentration in children with epilepsy on valproic acid monotherapy. Seizure. 2017;51:22-6.

786. Gulcebi MI, Ozkaynakci A, Goren MZ, Aker RG, Ozkara C, Onat FY. The relationship between UGT1A4 polymorphism and serum concentration of lamotrigine in patients with epilepsy. Epilepsy Res. 2011;95(1-2):1-8.

787. Riviere JB, Verlaan DJ, Shekarabi M, Lafreniere RG, Benard M, Der Kaloustian VM, et al. A mutation in the HSN2 gene causes sensory neuropathy type II in a Lebanese family. Ann Neurol. 2004;56(4):572-5.

788. Wilson FH, Disse-Nicodeme S, Choate KA, Ishikawa K, Nelson-Williams C, Desitter I, et al. Human hypertension caused by mutations in WNK kinases. Science (New York, NY). 2001;293(5532):1107-12.

789. Manunta P, Lavery G, Lanzani C, Braund PS, Simonini M, Bodycote C, et al. Physiological interaction between alpha-adducin and WNK1-NEDD4L pathways on sodium-related blood pressure regulation. Hypertension. 2008;52(2):366-72.

790. Fischer U, Heckel D, Michel A, Janka M, Hulsebos T, Meese E. Cloning of a novel transcription factor-like gene amplified in human glioma including astrocytoma grade I. Hum Mol Genet. 1997;6(11):1817-22.

791. Duarte JD, Turner ST, Tran B, Chapman AB, Bailey KR, Gong Y, et al. Association of chromosome 12 locus with antihypertensive response to hydrochlorothiazide may involve differential YEATS4 expression. Pharmacogenomics J. 2013;13(3):257-63.

792. Pengelly RJ, Gibson J, Andreoletti G, Collins A, Mattocks CJ, Ennis S. A SNP profiling panel for sample tracking in whole-exome sequencing studies. Genome Med. 2013;5(9):89.

793. Choo SY. The HLA system: genetics, immunology, clinical testing, and clinical implications. Yonsei Med J. 2007;48(1):11-23.

794. Yang W, Wu G, Broeckel U, Smith CA, Turner V, Haidar CE, et al. Comparison of genome sequencing and clinical genotyping for pharmacogenes. Clin Pharmacol Ther. 2016;100(4):380-8.

795. Söderbäck E, Zackrisson A-L, Lindblom B, Alderborn A. Determination of CYP2D6 Gene Copy Number by Pyrosequencing. Clinical Chemistry. 2005;51(3):522-31.

796. Hosono N, Kato M, Kiyotani K, Mushiroda T, Takata S, Sato H, et al. CYP2D6 genotyping for functional-gene dosage analysis by allele copy number detection. Clin Chem. 2009;55(8):1546-54.

797. Yang Y, Botton MR, Scott ER, Scott SA. Sequencing the CYP2D6 gene: from variant allele discovery to clinical pharmacogenetic testing. Pharmacogenomics. 2017;18(7):673-85.

798. EMA. European Medicines Agency Good Pharmacogenomic Practice [updated September 2019. Available from: https://www.ema.europa.eu/documents/scientific-guideline/guideline-good-pharmacogenomic-practice-first-version_en.pdf.

799. Ingelman-Sundberg M, Mkrtchian S, Zhou Y, Lauschke VM. Integrating rare genetic variants into pharmacogenetic drug response predictions. Hum Genomics. 2018;12(1):26.

800. Lauschke VM, Ingelman-Sundberg M. Requirements for comprehensive pharmacogenetic genotyping platforms. Pharmacogenomics. 2016;17(8):917-24.

801. Chen P, Lin JJ, Lu CS, Ong CT, Hsieh PF, Yang CC, et al. Carbamazepine-induced toxic effects and HLA-B*1502 screening in Taiwan. N Engl J Med. 2011;364(12):1126-33.

802. Kelly LE, Rieder M, van den Anker J, Malkin B, Ross C, Neely MN, et al. More codeine fatalities after tonsillectomy in North American children. Pediatrics. 2012;129(5):e1343-7.

803. van Schie RM, Wadelius MI, Kamali F, Daly AK, Manolopoulos VG, de Boer A, et al. Genotype-guided dosing of coumarin derivatives: the European pharmacogenetics of anticoagulant therapy (EU-PACT) trial design. Pharmacogenomics. 2009;10(10):1687-95.

804. Kimmel SE, French B, Anderson JL, Gage BF, Johnson JA, Rosenberg YD, et al. Rationale and design of the Clarification of Optimal Anticoagulation through Genetics trial. Am Heart J. 2013;166(3):435-41.

805. Cavallari LH. Time to revisit warfarin pharmacogenetics. Future Cardiol. 2017;13(6):511-3.

806. Gage BF, Bass AR, Lin H, Woller SC, Stevens SM, Al-Hammadi N, et al. Effect of Genotype-Guided Warfarin Dosing on Clinical Events and Anticoagulation Control Among Patients Undergoing Hip or Knee Arthroplasty: The GIFT Randomized Clinical Trial. JAMA. 2017;318(12):1115-24.

807. Lee YM, McKillip RP, Borden BA, Klammer CE, Ratain MJ, O'Donnell PH. Assessment of patient perceptions of genomic testing to inform pharmacogenomic implementation. Pharmacogenet Genomics. 2017;27(5):179-89.

808. McKillip RP, Borden BA, Galecki P, Ham SA, Patrick-Miller L, Hall JP, et al. Patient Perceptions of Care as Influenced by a Large Institutional Pharmacogenomic Implementation Program. Clin Pharmacol Ther. 2016.

809. Gibson ML, Hohmeier KC, Smith CT. Pharmacogenomics testing in a community pharmacy: patient perceptions and willingness-to-pay. Pharmacogenomics. 2017;18(3):227-33.

810. Howard RL, Avery AJ, Slavenburg S, Royal S, Pipe G, Lucassen P, et al. Which drugs cause preventable admissions to hospital? A systematic review. Br J Clin Pharmacol. 2007;63(2):136-47.

811. Shamliyan T. Adverse drug effects in hospitalized elderly: data from the healthcare cost and utilization project. Clin Pharmacol. 2010;2:41-63.

812. Department of Health NHS Reference costs, 2014-15. In: Health Do, editor.: H.M. Government; 2015.

813. Impicciatore P, Choonara I, Clarkson A, Provasi D, Pandolfini C, Bonati M. Incidence of adverse drug reactions in paediatric in/out-patients: a systematic review and meta-analysis of prospective studies. Br J Clin Pharmacol. 2001;52(1):77-83.

814. Benkhaial A, Kaltschmidt J, Weisshaar E, Diepgen TL, Haefeli WE. Prescribing errors in patients with documented drug allergies: comparison of ICD-10 coding and written patient notes. Pharm World Sci. 2009;31(4):464-72.

815. Berm EJ, Looff M, Wilffert B, Boersma C, Annemans L, Vegter S, et al. Economic Evaluations of Pharmacogenetic and Pharmacogenomic Screening Tests: A Systematic Review. Second Update of the Literature. PloS one. 2016;11(1):e0146262.

816. Verbelen M, Weale ME, Lewis CM. Cost-effectiveness of pharmacogenetic-guided treatment: are we there yet? Pharmacogenomics J. 2017;17(5):395-402.

817. Borse MS, Dong OM, Polasek MJ, Farley JF, Stouffer GA, Lee CR. CYP2C19-guided antiplatelet therapy: a cost-effectiveness analysis of 30-day and 1-year outcomes following percutaneous coronary intervention. Pharmacogenomics. 2017;18(12):1155-66.

818. Alagoz O, Durham D, Kasirajan K. Cost-effectiveness of one-time genetic testing to minimize lifetime adverse drug reactions. Pharmacogenomics J. 2016;16(2):129-36.

819. Plumpton CO, Roberts D, Pirmohamed M, Hughes DA. A Systematic Review of Economic Evaluations of Pharmacogenetic Testing for Prevention of Adverse Drug Reactions. Pharmacoeconomics. 2016;34(8):771-93.

820. Veenstra DL. The value of routine pharmacogenomic screening-Are we there yet? A perspective on the costs and benefits of routine screening-shouldn't everyone have this done? Clin Pharmacol Ther. 2016;99(2):164-6.

821. Schildcrout JS, Denny JC, Bowton E, Gregg W, Pulley JM, Basford MA, et al. Optimizing drug outcomes through pharmacogenetics: a case for preemptive genotyping. Clin Pharmacol Ther. 2012;92(2):235-42.

822. Lashley FR. Genetic testing, screening, and counseling issues in cardiovascular disease. J Cardiovasc Nurs. 1999;13(4):110-26.

823. Salkovskis PM, Rimes KA. Predictive genetic testing: psychological factors. J Psychosom Res. 1997;43(5):477-87.

824. Wu AH, White MJ, Oh S, Burchard E. The Hawaii clopidogrel lawsuit: the possible effect on clinical laboratory testing. Per Med. 2015;12(3):179-81.

825. Centers for Disease C, Prevention. Disparities in premature deaths from heart disease--50 States and the District of Columbia, 2001. MMWR Morb Mortal Wkly Rep. 2004;53(6):121-5.

826. GoDarts, Group UDPS, Wellcome Trust Case Control C, Zhou K, Bellenguez C, Spencer CC, et al. Common variants near ATM are associated with glycemic response to metformin in type 2 diabetes. Nat Genet. 2011;43(2):117-20.

827. Shuldiner AR, O'Connell JR, Bliden KP, Gandhi A, Ryan K, Horenstein RB, et al. Association of cytochrome P450 2C19 genotype with the antiplatelet effect and clinical efficacy of clopidogrel therapy. JAMA. 2009;302(8):849-57.

828. Heikkinen AT, Lignet F, Cutler P, Parrott N. The role of quantitative ADME proteomics to support construction of physiologically based pharmacokinetic models for use in small molecule drug development. Proteomics Clin Appl. 2015;9(7-8):732-44.

829. Nguyen L, Dang CC, Ballester PJ. Systematic assessment of multi-gene predictors of pan-cancer cell line sensitivity to drugs exploiting gene expression data. F1000Res. 2016;5.

830. Cardon LR, Harris T. Precision medicine, genomics and drug discovery. Hum Mol Genet. 2016;25(R2):R166-R72.

831. Evers R, Blanchard RL, Warner AW, Cutler D, Agrawal NG, Shaw PM. A question-based approach to adopting pharmacogenetics to understand risk for clinical variability in pharmacokinetics in early drug development. Clin Pharmacol Ther. 2014;96(3):291-5.

832. Ramsey BW, Davies J, McElvaney NG, Tullis E, Bell SC, Drevinek P, et al. A CFTR potentiator in patients with cystic fibrosis and the G551D mutation. N Engl J Med. 2011;365(18):1663-72.

833. Graham DM, Coyle VM, Kennedy RD, Wilson RH. Molecular Subtypes and Personalized Therapy in Metastatic Colorectal Cancer. Curr Colorectal Cancer Rep. 2016;12:141-50.

834. Mesbahi M, Shteinberg M, Wilschanski M, Hatton A, Nguyen-Khoa T, Friedman H, et al. Changes of CFTR functional measurements and clinical improvements in cystic fibrosis patients with non p.Gly551Asp gating mutations treated with ivacaftor. J Cyst Fibros. 2016.

835. Drummond M, de Pouvourville G, Jones E, Haig J, Saba G, Cawston H. A comparative analysis of two contrasting European approaches for rewarding the value added by drugs for cancer: England versus France. Pharmacoeconomics. 2014;32(5):509-20.

836. Zierhut HA, Campbell CA, Mitchell AG, Lemke AA, Mills R, Bishop JR. Collaborative Counseling Considerations for Pharmacogenomic Tests. Pharmacotherapy. 2017;37(9):990-9.

837. 23 and Me health reports 2016 [Available from: https://www.23andme.com/en-gb/health/reports/.

838. Newman WG, Murphy BF, Callard A, Payne K. A role for genetic counsellors and clinical geneticists in pharmacogenetics? Clin Genet. 2012;82(2):201-2; author reply 3.

839. Matthaei J, Brockmoller J, Tzvetkov MV, Sehrt D, Sachse-Seeboth C, Hjelmborg JB, et al. Heritability of metoprolol and torsemide pharmacokinetics. Clin Pharmacol Ther. 2015;98(6):611-21.

840. Cronin-Fenton DP, Damkier P. Tamoxifen and CYP2D6: A Controversy in Pharmacogenetics. Adv Pharmacol. 2018;83:65-91.

841. Matthaei J, Tzvetkov MV, Gal V, Sachse-Seeboth C, Sehrt D, Hjelmborg JB, et al. Low heritability in pharmacokinetics of talinolol: a pharmacogenetic twin study on the heritability of the pharmacokinetics of talinolol, a putative probe drug of MDR1 and other membrane transporters. Genome Med. 2016;8(1):119.

842. Singer E, Antonucci T, Van Hoewyk J. Racial and ethnic variations in knowledge and attitudes about genetic testing. Genet Test. 2004;8(1):31-43.

843. Donovan KA, Tucker DC. Knowledge about genetic risk for breast cancer and perceptions of genetic testing in a sociodemographically diverse sample. J Behav Med. 2000;23(1):15-36.

844. Suther S, Kiros GE. Barriers to the use of genetic testing: a study of racial and ethnic disparities. Genet Med. 2009;11(9):655-62.

845. Zimmerman RK, Tabbarah M, Nowalk MP, Raymund M, Jewell IK, Wilson SA, et al. Racial differences in beliefs about genetic screening among patients at inner-city neighborhood health centers. J Natl Med Assoc. 2006;98(3):370-7.

846. Shields AE. Ethical concerns related to developing pharmacogenomic treatment strategies for addiction. Addict Sci Clin Pract. 2011;6(1):32-43.

847. McGuire AL, Hamilton JA, Lunstroth R, McCullough LB, Goldman A. DNA data sharing: research participants' perspectives. Genet Med. 2008;10(1):46-53.

848. Gymrek M, McGuire AL, Golan D, Halperin E, Erlich Y. Identifying personal genomes by surname inference. Science (New York, NY). 2013;339(6117):321-4.

849. Tandy-Connor S, Guiltinan J, Krempely K, LaDuca H, Reineke P, Gutierrez S, et al. False-positive results released by direct-to-consumer genetic tests highlight the importance of clinical confirmation testing for appropriate patient care. Genetics In Medicine. 2018.

850. O'Neill P, Mestre-Ferrandiz J, Puig-Peiro R, Sussex J. Projecting expenditure on medicines in the UK NHS. Pharmacoeconomics. 2013;31(10):933-57.

851. Houwink EJ, Rigter T, Swen JJ, Cornel MC, Kienhuis A, Rodenburg W, et al. [Pharmacogenetics in primary health care: implementation and future expectations]. Ned Tijdschr Geneeskd. 2015;159:A9204.

852. Jia X, Han B, Onengut-Gumuscu S, Chen WM, Concannon PJ, Rich SS, et al. Imputing amino acid polymorphisms in human leukocyte antigens. PloS one. 2013;8(6):e64683.

853. Albert S, Garanto A, Sangermano R, Khan M, Bax NM, Hoyng CB, et al. Identification and Rescue of Splice Defects Caused by Two Neighboring Deep-Intronic ABCA4 Mutations Underlying Stargardt Disease. American journal of human genetics. 2018;102(4):517-27.

854. Forsythe DE. Personal communication- Ciliary phenotype of BBS patient fibroblasts. 2018.