Genomics of drug resistance in epilepsy

Andreja Avberšek

UCL

Queen Square Institute of Neurology

Department of Clinical and Experimental Epilepsy

Queen Square

London

WC1N 3BG

A thesis for submission to University College London for the degree of Doctor of Philosophy

DECLARATION

I, Andreja Avberšek, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed:

……..

Date:

ABSTRACT

Difficulties identifying drug-resistant epilepsy (DRE) at disease onset and complex temporal patterns of epilepsy represent challenges in research and clinical practice. A better understanding of the underlying mechanisms of DRE is needed to enable biomarker development, early diagnosis, and personalised treatments. This work explores the influence of genomic variation on DRE through genome-wide association (GWAS) and heritability analyses. It is part of a collaborative, European Commission funded project: EpiPGX (Epilepsy Pharmacogenomics: delivering biomarkers for clinical use).

Individuals with epilepsy were recruited from specialised clinical centres across Europe. Healthy controls were obtained from several publically available sources. To establish whether common genomic variants are associated with DRE, two GWAS were performed by the Author. The first analysis, comparing individuals with DRE and controls with drug-responsive epilepsy, did not reveal any variants with genome-wide significance. The second analysis, comparing individuals with DRE and healthy controls, revealed several loci with genome-wide significance. The top genome-wide association signal (rs75700350), located at 4q31.1, likely represents an artefact. Other findings include the signals at loci 5p13.2, and 11p13, pointing to potentially significant candidate genes, *SLC1A2* and *SLC1A3*, implicated in glutamate reuptake and excitotoxicity. Furthermore, one of these loci has been linked to an important epilepsy comorbidity, autism. The functional variants driving these signals may represent risk factors for drug resistance, epilepsy susceptibility, or variants affecting

pathophysiological pathways common to DRE and its comorbidities. The main limitations of these GWAS analyses were small sample sizes and the lack of replication.

To explore if drug resistance in epilepsy has a polygenic inheritance component, a single nucleotide polymorphism (SNP) heritability analysis was performed. This analysis yielded an estimate of DRE SNP heritability of 0.22, showing that drug resistance in epilepsy is heritable.

IMPACT STATEMENT

The findings of this thesis provide insight into the genetic background of drug resistance in common epilepsies. The heritability analysis presented in this work demonstrates that drug resistance in epilepsy has a heritable component, justifying further genome-wide research efforts in this field. The GWAS results provide preliminary evidence of potential functional variants underlying drug resistance across different epilepsy syndromes. Alternatively, they could also reflect the pathophysiological pathways common to DRE and its comorbidities.

Together, these findings may point to a future role of genomic variants as a screening tool to predict response to antiepileptic drugs (AEDs). If the results are confirmed in replication studies, this would increase our understanding of DRE mechanisms and facilitate the development of biomarkers and targeted therapies.

The extensive database of epilepsy cases with well-phenotyped long-term outcomes created in the course of this project is a rich resource for future research efforts seeking to understand epilepsy prognosis and treatment outcomes.

STATEMENT OF CONTRIBUTION

I completed this work as part of collaborative, European Commission 7th Framework funded project: EpiPGX (Epilepsy Pharmacogenomics: delivering biomarkers for clinical use). This was an extensive effort with contributions from numerous researchers (see full list below). My contribution is outlined in Table 1.

EpiPGX researchers:

Principal Investigators:

Sanjay M. Sisodiya (SMS), University College London (UCL) – Project Coordinator

Martin J. Brodie (MJB), University of Glasgow (UGLA)

John Craig (JC), Belfast Health and Social Care Trust (BHSCT)

Norman Delanty (ND), Royal College of Surgeons in Ireland (RCSI)

Chantal Depondt (CD), Université Libre de Bruxelles (ULB)

Andrés Ingason (AI), Islensk Erfdagreining EHF (deCODE)

Michael R. Johnson (MRJ), Imperial College London (Imperial)

Bobby P.C. Koeleman (BPCK), Universitar Medisch Centrum Utrecht (UMCU)

Roland Krause (RK), Université du Luxembourg (UL)

Holger Lerche (HL), Eberhard Karls Universitaet Tuebingen (EKUT)

Anthony G. Marson (AGM), University of Liverpool (ULIV)

Josemir W. Sander (JWS), Stichting Epilepsie Instellingen Nederland (SEIN)

Graeme J. Sills (GJS), ULIV

Federico Zara (FZ), Instituto Giannina Gaslini (IGG)

EpiPGX Clinical Fellows:

Pauls Auce (PA), ULIV

Andreja Avbersek (AA), UCL

Felicitas Becker (FB), EKUT

Bianca Berghius (BB), SEIN

Mojgan Borghei (MB), ULB

Antonietta Coppola (AC), IGG

Kristin Heggeli (KH), UCL

Costin Leu (CL), UCL

Narek Sargasyan (NS), UCL

Lisa Slattery (LS), RCSI

Douglas Speed (DS), UCL

William Stern (WS), UCL

Anna Tostevin (AT), UCL

Merel Wasenaar (MW), SEIN

Stefan Wolking (SW), EKUT

Table 1: Work completed by the Author and other EpiPGX Researchers

Task	Work performed by the Author	Work performed by the EpiPGX Consortium Collaborators
Consensus phenotype definitions	NA	EpiPGX Principal Investigators: JC , ND, CD, HL, AGM, JWS, SMS
Case record form design	Contributed	Led by CD
EpiPGX database design	Contributed	Led by RK
EpiPGX database data entry manual	Co-authored	Co-authored by CD, contributions from PA, FB, BB, MB, KH, RK, HL, AGM, NS, GJS, LS, WS, MW, SW
Training clinical fellows in phenotyping	Contributed	CD
Recruitment	Contributed to the recruitment of the UCL cohort	UCL cohort: AC, NS, WS, AT, Department of Clinical and Experimental Epilepsy (DCEE) Clinical Fellows and Consultants); Other cohorts: various contributors
Phenotyping	Phenotyped 2,707 out of 3,067 UCL cases Classified long-term outcomes for 500 ULIV cases Collected information on the reasons for undefined outcomes	KH, WS, NS collectively phenotyped 360 UCL cases; EpiPGX fellows phenotyped 6,560 cases from other sites
Preparation of anonymised cases for the cross-site inter-reliability assessment	Contributed	Led by CD, contributions from PA, FB, BB, MB, KH, NS, LS, WS, MW, SW
Inter-rater reliability assessment	Contributed	Led by CD, contributions from PA, FB, BB, MB, KH, NS, LS, WS, MW, SW, SMS
SNP genotyping	NA	Performed at DeCODE
Imputation	NA	Performed at DeCODE by AI
GWAS analyses	Performed GWAS1 and GWAS2 analyses with imputed data	Supervision CL and SMS; technical help and input from DS, LRG
Heritability analysis	Performed the analysis	Technical help DS

NA – not applicable

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ABC	ATP-binding cassette
ABCB1	ATP-binding cassette subfamily B member 1
ACTH	Adrenocorticotrophic hormone
ADHD	Attention deficit hyperactivity disorder
ADIPOQ	Adiponectin gene
AED	Antiepileptic drug
ADNFLE	Autosomal nocturnal frontal lobe epilepsy
AMPA	α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid
aCGH	Array comparative genomic hybridisation
APOE	Apolipoprotein E
ASD	Autistic spectrum disorder
ATP	Adenosine triphosphate
BBB	Blood-brain barrier
BCL11A	BAF chromatin remodeling complex subunit 11A
BHSCT	Belfast Health and Social Care Trust
BRD7	Bromodomain containing protein 7
CEPH	Centre d'Etude du Polymorphism Humain
CHD2	Chromodomain helicase DNA binding protein 2
Chr	Chromosome
CHRNA4	Neuronal nicotinic acethylcholine receptor subunit alpha 4
CMDh	Coordination Group for Mutual Recognition and Decentralised Procedures
CNS	Central nervous system
CNV	Copy number variation

CI	Confidence interval
cM	Centimorgan
CRF	Case record form
CYP	Cytochrome P450
DCEE	Department of Clinical and Experimental Epilepsy
DDD	Defined daily dose
deCODE	Islensk Erfdagreining EHF
DNA	Deoxyribonucleic acid
DRE	Drug-resistant epilepsy
EAAT1	Excitatory amino acid transporter 1
EAAT2	Excitatory amino acid transporter 2
EE	Epileptic encephalopathy
EEG	Electroencephalography
EKUT	Eberhard Karls Universitaet Tuebingen
EMA	European Medicines Agency
EMInet	Epilepsy and Migraine Integrated Network
EPHX1	Epoxide hydroxylase 1
Epi4K	Gene Discovery in 4,000 Genomes
EPIGEN	The Epilepsy Genetics Consortium
EPGP	The Epilepsy Phenome/Genome Project
eQTL	Expression quantitative trait loci
FANCL	Fanconi anaemia complementation group L
FaST-LMM	Factored spectrally transformed linear mixed models
FCD	Focal cortical dysplasia
FJX1	Four-jointed box kinase 1
FS+	Febrile seizures plus

FU	Follow-up
GABA	γ-aminobutyric acid
GABO:mi	GABO:MI Gesellschaft fur Ablauforganisation: Millarium MBH&Co KG GABO
GABRA1	Gamma-aminobutyric acid type A receptor alpha1 subunit
GABRA2	Gamma-aminobutyric acid type A receptor alpha2 subunit
GABRG2	Gamma-aminobutyric acid type A receptor gamma2 subunit
GAT-1	GABA transporter 1
GCTA	Genome-wide Complex Trait Analysis
GEFS+	Generalised epilepsy with febrile seizures plus
GGE	Genetic generalised epilepsy
GLUT-1	Glucose type 1 transporter
GOLIM4	Golgi integral membrane protein 4
GRC	The Genome Reference Consortium
GRCh37	Genome Reference Consortium human build 37
GRCh38	Genome Reference Consortium human build 38
GREML	Genomic restricted maximum likelihood method
GRIN2A	Glutamate NMDA receptor subunit 2A
GRM	Genetic relationships matrix
GWAS	Genome-wide association study
$\mathbf{h^2}$	Total heritability
$\mathbf{h^2_{GWS}}$	Genome-wide-significant-SNP-based heritability
$\mathbf{h^2_{SNP}}$	SNP-based heritability
HCN1	Potassium/sodium hyperpolarisation-activated cyclic nucleotide-gated channel 1
HEATR3	HEAT repeat containing protein 3
hg19	Human genome build 19

HGP	The Human Genome Project
HLA	Human leukocyte antigen
HMGB1	High-mobility group box 1
HWE	Hardy-Weinberg equilibrium
IBD	Identity by descent
IBS	Identity by state
IGG	Instituto Giannina Gaslini
ILAE	International League Against Epilepsy
IRASFS	The Insulin Resistance Atherosclerosis Family Study
IQR	Interquartile range
JME	Juvenile myoclonic epilepsy
KCNQ	Potassium voltage-gated channel subfamily Q
KCNQ2	Potassium voltage-gated channel subfamily Q member 2
KCNQ3	Potassium voltage-gated channel subfamily Q member 3
Kb	Kilobase
kDA	Kilodalton
Kv7.2	Voltage-gated potassium channel subunit 7.2
Kv7.3	Voltage-gated potassium channel subunit 7.3
LD	Linkage disequilibrium
LDAK	LD-Adjusted Kinships
LINC00433	Long intergenic non-protein coding RNA 433
lncRNAs	Long non-coding RNAs
MAF	Minor allele frequency
Mb	Megabase
MDS	Multidimensional scaling
MGARP	Mitochondria-localised glutamic acid rich protein

MGI	Mouse Genome Informatics
mRNA	Messenger ribonucleic acid
MTLE	Mesial temporal lobe epilepsy
MTLE-HS	Mesial temporal lobe epilepsy with hippocampal sclerosis
mTOR	Mammalian target of rapamycin
NAD	Nicotinamide adenine dinucleotide
NADK2	Nicotinamide adenine dinucleotide kinase 2
$Na_v1.1$	Sodium voltage-gated channel type I alpha subunit
NBS	National Blood Bank Service
Nc	Non-coding
NCDS	The National Child Development Study
NES	Non-epileptic seizures
NGS	Next generation sequencing
NHGRI-EBI	National Human Genome Research Institute and the European Bioinformatics Institute
NHNN	National Hospital for Neurology and Neurosurgery
NICE	National Institute for Health and Care Excellene
NIPBL	Nipped-B-like
OR	Odds ratio
PAMR1	Peptidase domain containing associated with muscle degeneration
PCA	Principal component analysis
PCDH7	Protocadherin 7
PET	Positron emission tomography
P-gp	P-glycoprotein
PWE	People with epilepsy
QC	Quality control

QQ	Quantile-quantile
RANBP3L	RAN binding protein 3-like
RCSI	Royal College of Surgeons in Ireland
REML	Restricted maximum likelihood
REST	Repressor element 1-silencing transcription factor
RNA	Ribonucleic acid
RR	Relative risk
SANAD	Standard and New Antiepileptic Drugs
SCN1A	Sodium voltage-gated channel type I alpha subunit
SCN2A	Sodium voltage-gated channel type II alpha subunit
SCN3A	Sodium voltage-gated channel type III alpha subunit
SCN8A	Sodium voltage-gated channel type VIII alpha subunit
SE	Status epilepticus
SEIN	Stichting Epilepsie Instellingen Nederland
SKP2	S-phase kinase-associated protein 2
SLC1A2	Solute carrier family 1 member 2
SLC1A3	Solute carrier family 1 member 3
SLC2A1	Solute carrier family 2 member 1
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variant
Sp1	Specificity protein 1
STXBP1	Syntaxin binding protein 1
SUDEP	Sudden unexpected death in epilepsy
SV	Structural variant
SYNGAP1	Synaptic Ras GTPase-activating protein 1
TNF-α	Tumor necrosis factor-α

TRIM44	Tripartite motif containing 44
TSPO	Translocator protein 18 kDA
TTC21B	Tetratricopeptide repeat domain 21B
UCL	University College London
UGLA	University of Glasgow
UKB	Universitaetsklinikum Bonn
UL	Université du Luxembourg
ULB	Université Libre de Bruxelles
ULIV	University of Liverpool
UMCU	Universitar Medisch Centrum Utrecht
VKORC1	Vitamin K epoxide reductase complex
VRK2	Vaccinia-related kinase 2
WES	Whole exome sequencing
WGS	Whole genome sequencing
WHO	World Health Organization
WTCCC	Wellcome Trust Case Control Consortium
WP	Work-package
XIST	X-inactive specific transcript

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisors, Sanjay Sisodiya, Ley Sander, and Rita Louro Guerreiro, for giving me the opportunity to have this academic experience. It has transformed my life and opened the door to career opportunities which would have otherwise been out of my reach.

I would like to thank the EpiPGX study participants for allowing us to use their medical and genetic data, and the clinicians, health care professionals, and researchers at all study sites, for their efforts recruiting, phenotyping, and managing the data.

A special thanks goes to Kristin Heggeli for her substantial phenotyping contribution, for the numerous occasions when when she helped me editing texts, and for being an amazing colleague and friend. I would also like to acknowledge the work of Anna Tostevin, Joe Willis, William Stern, Narek Sargasyan, and other DCEE clinical fellows. I am extremely grateful to Doug Speed for many insightful discussions and helpful advice regarding coding and genetic analyses, to Costin Leu for providing the study design and guidance, and to Roland Krause for informatics support. Last but not least, I would like to thank Caroline Selai, Daniela Warr Schori, Jane De Tisi, Juliet Solomon, Aidan Neligan, John Hardy, and Rafal Kaminski for their support during my studies.

This work is dedicated to individuals living with epilepsy and their families.

CHAPTER 1: BACKGROUND

1.1 Introduction

Epilepsy has a highly variable clinical course with outcomes ranging from response to the first anti-epileptic drug (AED) to treatment failure with numerous AED trials, as well as varying degrees of impact on cognitive function, mood, quality of life, and independence (1-3).

Drug resistance remains poorly understood, despite being the most significant contributor to disease burden in people with epilepsy (PWE). Drug-resistant epilepsy (DRE) is associated with higher frequencies of comorbid illnesses (4, 5), psychological dysfunction (6, 7), social stigmatisation (8), reduced quality of life, and a higher risk of premature mortality in comparison with drug-responsive epilepsy (4, 9-13). The mechanisms underlying drug resistance remain to be fully elucidated. There is general consensus that genetic factors likely contribute to variable AED response in humans; however, this is supported by limited scientific evidence (14, 15).

The present work explores the contribution of common genomic variation to DRE. Currently, the application of evidence-based management of epilepsy is limited by significant gaps in knowledge. Studying the genetic landscape of DRE is important to identify novel mechanisms underlying DRE and to establish genome-based biomarkers

for use in clinical practice and research. Early identification of DRE would decrease the burden of failed AED trials and shorten the periods of time individuals might spend taking AEDs without benefit. It would encourage better counseling, early consideration of rational combination therapy, and early referral for epilepsy surgery, resulting in better epilepsy outcomes (16-18). Predictive biomarkers of DRE have a significant potential for use in pharmaceutical drug development, enabling clinical trial designs where genotypes are used as eligibility or stratification criteria and not only in post-hoc analyses (19).

1.2 Definitions

1.2.1 Definition of drug resistance

Defining DRE appropriately and consistently is vital to the success of studying its epidemiological, clinical, genomic, and pharmacogenomic aspects. However, even this important first step remains a challenge and a subject of debate. Several definitions of DRE were proposed and used by various authors in the past decades (20-23). Some studies used strict definitions. For instance, Berg *et al*. suggested the following criteria in 1996: “uncontrolled seizures with an average frequency of at least one seizure per month for a period of at least two years” and “failure of at least three different AEDs, either because they did not control seizures or because of unacceptable side effects” (24). In contrast, other authors used much more lenient definitions, such as the one

proposed by Arts *et al*. in 1999: failure to attain at least three months' seizure-free period at six months after diagnosis (20). Significant differences in the way DRE has been defined pose a challenge for establishing unified treatment guidelines and comparisons between different studies.

To improve the medical care of PWE and to facilitate inter-study comparisons, the International League against Epilepsy (ILAE) proposed a consensus definition of DRE in 2010 (25). The ILAE definition framework encompasses two hierarchical levels. Level 1 outlines standard criteria for AED trial outcome classification into the following categories: "seizure-free", "treatment failure", or "undetermined". Categorising an AED trial outcome as "seizure-free" or "treatment failure" requires the AED trial to be "appropriate for the epilepsy and seizure type" and "adequate", i.e. "applied at adequate dosage for a sufficient length of time". Level 2 outlines the core definition of DRE as "failure of adequate trials of two tolerated and appropriately chosen and used AED schedules (whether as monotherapies or in combination) to achieve sustained seizure freedom". Seizure freedom refers to "freedom from all seizures, including auras" (25, 26). If appropriate, this core definition may be modified to fit specific purposes. It is important to note that DRE classification is not static as the course of epilepsy in an individual may fluctuate over time. As a consequence, an individual may be deemed drug-resistant at certain points in time and not others (26).

1.2.2 Definition of drug response

According to the ILAE definition Level 1 scheme, the outcome of an AED trial is categorised as “seizure-free” (Category 1 response) if the treatment results in “seizure freedom for 12 months, or for a minimum of three times the longest pre-intervention inter-seizure interval, whichever is longer” (26). This definition implements the “rule of three for calculating confidence intervals for zero events” (27). If the observed period of seizure freedom is “at least three times the longest pre-treatment inter-seizure interval”, then the certainty that there has been a therapeutic effect is 95%. Level 2 of the ILAE framework defines drug-responsive epilepsy as a Category 1 outcome of the current AED treatment, i.e. “freedom from seizures for a minimum of three times the longest pre-intervention inter-seizure interval, or 12 months, whichever is longer” (26).

Often, studies of long-term epilepsy outcomes describe drug-responsiveness as terminal remission. This is defined as the time from the last seizure to the end of the follow-up, on or off AEDs. One-year or five-year terminal remission is most commonly reported (17, 28, 29).

1.2.3 Undetermined response

In some cases, it may not be possible to classify the epilepsy as drug-resistant or drug-responsive. According to the ILAE consensus proposal, the response should be temporarily classified as “undetermined” in such circumstances (26). Some examples are listed below:

a) In an individual with newly diagnosed epilepsy where the seizures stop after introducing an AED, but not enough time has passed to categorise the outcome as seizure freedom.
b) If seizures continue, but the individual has had fewer than two informative AED trials. An AED trial may be uninformative because the drug is not appropriate for the seizure or epilepsy type, because it has not been administered at a sufficient dose, or for long enough due to adverse drug reactions or other reasons.
c) If there is an apparent change in drug responsiveness, for example, if there is a relapse of seizures after a period of remission. In such cases, "the epilepsy is no longer drug-responsive, but it can only be considered drug-resistant if it subsequently meets the criteria for drug resistance" (26).

1.3 Epidemiology

Establishing reliable epidemiologic estimates for a complex group of disorders such as epilepsy is challenging. Prevalence and incidence estimates vary widely across studies and are usually higher in developing countries (30). Two meta-analyses including 222 and 65 studies in adult and paediatric populations have attempted to assess the global prevalence rates of epilepsy (31, 32). The overall lifetime prevalence of epilepsy was estimated at 7.6/1,000 (31). The estimates ranged from 5.2 to 5.8/1,000 for developed and 8.8 to 15.4/1,000 for developing countries. The overall prevalence of active

epilepsy was estimated at 6.4/1,000. It ranged from 4.9 to 5.5/1,000 for developed and from 6.7 to 12.7 for developing countries. There was substantial variability between estimates from individual studies included in both meta-analyses, depending on the country income, age range of the population, and sample size. Other potential sources of variability included imbalances in the frequency of different aetiologies and differences in case definitions or case ascertainment (31, 32). Commenting on the relationship between the lifetime prevalence and prevalence of active epilepsy is difficult since the estimates come from different subsets of individual studies included in both meta-analyses, resulting in imbalances in cohort characteristics for both types of prevalence. In addition, studies reporting active epilepsy used a variety of case definitions which were not taken into account in the meta-analyses. Some examples include: an individual who is "currently taking medication for epilepsy or has had at least one seizure in the past year", "at least one epileptic seizure in the previous five years, regardless of AED treatment", and definitions using different time frames, such as two years (33, 34).

Two meta-analyses have provided global incidence rates of epilepsy. Overall incidence rates were estimated at 50 to 61/100,000/year. They ranged from 45 to 49/100,000/year for developed and from 82 to 139/100,000/year for developing countries (31, 35). The incidence rates were also higher in the paediatric and elderly age groups (31).

1.3.1 Long-term outcomes of epilepsy

1.3.1.1 Remission

The likelihood of achieving terminal remission is one of the ways to describe the long-term prognosis of epilepsy (36, 37).

There is general consensus that up to 70% of PWE eventually achieve long-term seizure freedom (38, 39). This is supported by several studies assessing the prognosis of epilepsy by following well-defined cohorts of newly diagnosed individuals prospectively for 5 to 40 years, as summarised in Table 1.1. The proportion of PWE in one-year terminal remission at the end of follow-up ranged from 59 to 84%, the median duration of follow-up being 5 to 11 years (17, 28, 40-43). The proportion of PWE in five-year terminal remission ranged from 54 to 73% (29, 37, 41, 42, 44-46). The median duration of follow-up in the studies reporting five-year terminal remission rates was 7.1 to 40 years.

Table 1.1: Studies of long-term outcomes in epilepsy

Study	Number of subjects Study design Duration of FU	Proportion of subjects with ≥ 1-year remission at any point during FU	Proportion of subjects with ≥ 5-year remission at any point during FU	Proportion of subjects with ≥ 1-year terminal remission	Proportion of subjects with ≥ 5-year terminal remission	Proportion of subjects not in terminal remission at the end of FU
Annegers *et al*., 1979 (44)	N = 457 Paediatric cohort Retrospective Minimum 5 years FU (72% of the cohort had ≥ 10 years FU and 30% had ≥ 20 years FU)	NA	65% within 10 years 76% within 20 years	NA	61% at 10 years 70% at 20 years	39% not in 5-year terminal remission at 10 years 30% not in 5-year terminal remission at 20 years
Bell *et al*., 2016 (45)	N = 354 Paediatric and adult Prospective Median FU 22.3 years (IQR 12.1-24.1)	NA	NA	NA	73%	27% not in 5-year terminal remission
Brodie *et al*., 2012* (28)	N = 1,098 Paediatric and adult Prospective Median FU 7.5 years (IQR 4.7-12)	75%	NA	68%	NA	32% not in 1-year terminal remission

Chen *et al*., 2018* (40)	N = 1,795 Paediatric and adult Prospective Median FU 11 years (IQR 7-16)	NA	NA	63.7%	NA	36.3% not in 1-year terminal remission
Cockerell *et al*., 1997 (41)	N = 564 Paediatric and adult Prospective Median FU 7.1 years (CI 5.7, 8.1)	96%	71%	84%	54%	26% not in 1-year terminal remission 46% not in 5-year terminal remission
Geerts *et al*., 2010 (29)	N = 413 Paediatric cohort Prospective Median FU 14.8 (range 11.6-17.5)	NA	NA	NA	70.9%	29.1% not in 5-year terminal remission
Giussani *et al*., 2016 (47)	N = 747 Paediatric and adult Retrospective Median FU 9.5 years (IQR 4.5-22.5)	NA	NA	NA	The probability of starting 5-year remission was 50.2% at 20 years	NA
Hanaoka *et al.*, 2017 (42)	N = 258 Paediatric Retrospective FU 10 years	NA	NA	71.7%	59.3%	28.3% not in 1-year terminal remission 40.7% not in 5-year terminal remission
Kwan and Brodie, 2000* (17)	N = 525 Paediatric and adult Prospective Median FU 5 years (range 2-12)	NA	NA	63%	NA	37% not in 1-year terminal remission

Lindsten *et al*., 2001 (48)	N = 107 Adult cohort Prospective Up to 12 years FU (80% of the cohort had ≥ 5 years FU, 35% ≥ 10 years FU)	68%	58%	NA	NA	NA
Mohanraj and Brodie, 2006* (43)	N = 780 Adult cohort Prospective Median FU 79 months (range 24-252)	64.4%	NA	59.2%	NA	40.8% not in 1-year terminal remission
Silanpää and Schmidt, 2006 (37)	N = 144 Paediatric cohort Prospective Median FU 40 years (range 11-42)	NA	NA	NA	67%	33% not in 5-year terminal remission
Wakamoto *et al*., 2000 (46)	N = 155 Paediatric cohort Retrospective Mean FU 18.9 +/- 5.3 years (range 6-37.5)	NA	NA	NA	62.8% at final FU Estimated 52% at 10 years from onset and 56% at 20 years from onset	37.2% not in 5-year terminal remission
Zhang *et al*., 2013 (49)	N = 180 Paediatric and adult Prospective Median FU 5 years (range 2-10)	80%	NA	60%	NA	40% not in 1-year terminal remission

CI = confidence interval, FU = follow-up, IQR = interquartile range, NA = not applicable

*There was an overlap of participants in these studies.

For most individuals who achieve seizure remission, this happens early in the disease course (28). Approximately 47 to 60% of people with newly diagnosed epilepsy achieve seizure freedom with the first AED, 13% to 17% with the second AED, and 3% with the third AED, with only 0.6% to 0.8% responding to further trials (17, 43, 49).

1.3.1.2 Drug resistance

While the prognosis in terms of AED response is favourable for the majority of PWE, some individuals do not achieve remission, despite treatment with several AEDs (17).

Several studies suggest drug resistance develops early in the epilepsy course (17, 21, 43). The work of Brodie and colleagues indicates that a large proportion of people who respond to their first AED never experience another seizure (78% in a recently published study) and that drug resistance can be predicted very early in the disease course, with most PWE responding to the first or second AED and a decreasing probability of responding to further trials (17, 28, 43). Lindsten *et al*. concluded that the majority of the drug-resistant population can be identified within two years following epilepsy diagnosis (48). In the study by Kwan and Brodie, only 11% of participants for whom the absence of response to the first AED was due to the lack of efficacy subsequently achieved seizure freedom (17). These results imply that the likelihood of a favourable prognosis is small once an individual is identified as resistant (17, 23). On this basis, the ILAE has defined DRE as ”failure of adequate trials of two tolerated and appropriately chosen and used AED schedules (whether as monotherapies or in combination) to achieve sustained seizure freedom“ (26).

DRE was variably defined by different authors prior to the ILAE definition and as a result the reported prevalence of DRE varies across studies. In hospital-based studies, using variable definitions of drug resistance, 25 to 35% of participants are reported to have DRE (20, 28, 43). Population-based studies report lower rates of drug resistance: 19.2% to 22.5% (50, 51), likely because less severe cases are usually not seen in tertiary centres. The study by Berg *et al*. was not strictly population-based; however, recruitment of PWE from academic centres, as well as private centres and community clinics, allowed for a more comprehensive inclusion of cases. The proportion of individuals with DRE in this study was 10% (21). One study reported the prevalence of DRE as 1.36/1,000 (50). To illustrate the impact of different definitions of DRE, Berg and Kelly (52) assessed the proportion of resistant cases in their cohort using six different definitions. The proportion of cases categorised as drug-resistant varied from 9 to 24%, depending on the definition used.

To my knowledge, only three studies have reported the incidence or prevalence of DRE using the ILAE definition (see Table 1.2). Giussani *et al*. studied the population of the Italian province Lecco; the prevalence of DRE was 0.73/1,000 population (15.6% of all people with active epilepsy) (47). The remaining two studies were tertiary-centre based and the proportion of people with DRE was 21.5% (53) and 17% (54).

Table 1.2: Studies of the incidence/prevalence of DRE using the ILAE definition

Study	Study design	Number of subjects Duration of FU	Definitions	Incidence/prevalence of DRE
Giussani *et al*., 2016 (47)	Retrospective, population-based study, paediatric and adult population	Total N = 1021 Incident cohort N = 342 Prevalent cohort N = 747 Median FU 9.5 years (IQR 4.5-22.5)	ILAE definition of DRE	Prevalence of DRE: 107/684 (15.6%) of all individuals with active epilepsy (0.73/1,000 local population of Lecco, Italy) Incidence of DRE: 10.5% during the FU period
Kong *et al*., 2014 (53)	Adult population Retrospective, tertiary referral centre based	N = 557 Mean FU 12 years (range 1-28)	Drug response defined as ”12 months of seizure freedom“ or ”at least three-times the pre-treatment seizure-free interval“	Prevalence of DRE: 21.5% of the entire cohort Prevalence of drug-responsive epilepsy: 40.9% Undefined response: 37.5%
Ramos-Lizana *et al*., 2012 (54)	Paediatric population Prospective, tertiary referral centre based	N = 508 Mean FU 90 months (SD 45, range 24-168)	ILAE definition of DRE Dug response defined as achievement of seizure freedom on the first or second AED regimen Undefined drug response - not fitting any of the above categories	Prevalence of DRE: 17% of the whole cohort, 19% of the treated individuals

FU = follow-up, IQR = interquartile range, SD = standard deviation.

1.3.2 Predictors of drug resistance

The course of epilepsy and the development of drug resistance are challenging to predict, irrespective of the epilepsy type (55). As already mentioned in Section 1.1, reliable prediction of DRE would significantly improve counselling of PWE regarding expected outcomes, facilitate timely consideration of alternative treatments, and potentially provide a useful means of stratification in clinical trials (16-18).

Studies of DRE predictors often produce conflicting results, which can partly be explained by different methodologies and heterogeneous study populations. Epilepsy cohorts often consist of individuals with different epilepsy types and the results can be skewed in favour of the most represented aetiology or seizure type (56). In addition, epilepsy oucomes may be defined differently across studies.

The most consistently reported predictors of drug resistance are:

Symptomatic aetiology

Individulas with epilepsy due to structural brain abnormalities are less likely to respond to AEDs than individuals without abnormal imaging findings. This has been showed in several studies for both focal and generalised epilepsies (17, 44, 57-60). Deficits on neurologic examination at disease onset and focal lesions on brain imaging which are indicators of a symptomatic cause (44, 61, 62) are also predictors of a poorer outcome, although some studies have not confirmed this (63). Among the localisation-related

epilepsies due to symptomatic causes, certain syndromes, for example mesial temporal lobe epilepsy with hippocampal sclerosis (MTLE-HS), are overrepresented in drug-resistant populations and in populations referred to specialist epilepsy services (57, 64).

Seizure type

Several studies have suggested that individuals with complex focal seizures are less likely to achieve seizure freedom in comparison with individuals experiencing generalised tonic-clonic and absence seizures (44, 65). Subgroup analysis of the data from the Standard and New Antiepileptic Drugs (SANAD) trial Arm A (focal epilepsy) showed that secondary generalised seizures were a risk factor for time to treatment failure and focal simple or focal complex seizures only were a protective factor (2). In the UK National General Practice Survey of Epilepsy (NGPSE), the seizure type was not an important predictor of prognosis (66). Individuals experiencing multiple seizure types have been shown to have a less favourable prognosis (29, 67-69). Recent research also suggests seizure clusters are associated with a poorer prognosis (70).

High seizure frequency at the onset of the condition

Several large studies have identified high seizure frequency or seizure density at onset as a predictor of drug resistance in multivariate analyses (2, 21, 36, 66, 71). In the study by Sillanpää and Schmidt, pre-treatment seizure frequency of less than one seizure per week was a predictor of the cumulative probability of one-year remission at any point during follow-up, one-year terminal remission, and continuous remission up to the end of follow-up in univariate analyses, but it only remained a significant

predictor in multivariate analysis for the latter outcome. The probability of uninterrupted remission in individuals with fewer than one seizure per week was 23%, compared with 14% in individuals with equal or more than one seizure per week. Weekly seizures prior to initiating treatment were associated with a higher likelihood of never entering one-year terminal remission (59). Kwan and Brodie have found a linear relationship between the proportion of individuals developing DRE and the seizure number prior to treatment start. Epilepsy was resistant to treatment in 51% of individuals reporting more than 20 seizures at diagnosis, compared to 29% of those who reported fewer than 20 seizures (17). On the other hand, a study in rural Kenya has shown a similar likelihood of remission in individuals who had been experiencing seizures for several years before starting treatment than in those starting treatment after a small number of seizures (72). Detailed analysis of data from several observational studies has suggested that the seizure frequency prior to treatment initiation is a predictor of drug resistance only in individuals with complex partial seizures (73). In the study by Sillanpää and Schmidt, a higher seizure frequency prior to treatment start predicted DRE only in individuals with symptomatic epilepsy (59). This would imply that the initial seizure frequency is a reflection of aetiology driving the response to AEDs (74). Differences in the outcomes of individual studies could be due to different methods, but also to different representation of certain aetiologies. Especially in paediatric populations, frequent seizures may be the initial presentation of an array of epileptic syndromes that vary widely in severity (59). It is also possible that frequent seizures induce changes leading to DRE, as Gowers stated in 1881: ”The tendency of the disease is toward self-perpetuation; each attack facilitates the occurrence of the next by increasing the instability of the nerve elements” (75).

Poor response to initial treatment

Initial response to AED treatment appears to be a significant predictor of long-term outcome. A significant reduction or complete seizure control observed in the first months of starting the first AED has been shown to predict the likelihood of subsequent remission (58). Continuing seizures despite an adequate dose of an appropriately selected AED are a predictor of a lower chance of remission, while this is not the case if an AED failure is due to the lack of tolerability (17). A study in a paediatric population showed that children continuing to experience a weekly seizure frequency within the first year of initiating treatment were eight-fold more likely to develop DRE and had twice the risk of never achieving a one-year terminal remission (59). If the first AED trial fails to control the seizures, the likelihood of success for the subsequent AED regimens gradually declines (17, 20, 23, 44, 58, 76-78). More recent studies have demonstrated remissions in a proportion of PWE who had failed more than five AED regimens, suggesting that AED failure does not indicate a uniformly poor prognosis (28). Many of these individuals with late remissions will, however, subsequently relapse (79). This is discussed in more detail in Section 2.5.3.5.

Factors that are less likely to be significant predictors of drug resistance or for which the evidence is inconclusive include:

Younger age at disease onset

Several studies in children suggest lower age at onset is a poor prognostic factor (2, 80, 81). However, age less than 12 years in the study by Camfield *et al*. (36) was a

predictor of remission, and in the study by Sillanpää and Schmidt, age at onset was not a predictor of outcome (59). The effect of age at onset on the outcomes was not shown to be consistent in studies including adult and paediatric populations (44, 66). Similarly, age at onset was not found to be an independent factor correlating with prognosis in multivariate analyses of prognostic factors (66). Of note, comparisons between studies are often difficult as different authors use different ways to categorise age at onset, e.g. younger or older than one year, 10 years, 12 years, or 70 years. Different categorisations may also affect the results. Taking all of this into account, it is likely that any effects of age on outcomes reflect the differences in the frequencies of epilepsy syndromes in different age categories (56).

Learning disability

Learning disability has been associated with DRE (36, 76) and seizure relapse after experiencing remission (82). Studies including all age groups or adults only, however, have not replicated these findings. Studies of the NGPSE and Glasgow cohorts did not show a significant effect of learning disability on prognosis (66, 71). It is possible that learning disability may not be an independent risk factor of DRE; however, it could be a reflection of an underlying neurological deficit and therefore a symptomatic or genetic cause (21, 83).

Despite the existing knowledge about the factors influencing the outcome of epilepsy, accurate prediction of drug resistance that could inform treatment decisions remains challenging (55). Few prognostic models enabling outcome prediction have been proposed.

1.4 Mechanisms of action of AEDs

Current AEDs suppress seizures, but fail to influence the underlying epileptogenic mechanisms (84). AEDs control seizures by either reducing neuronal excitation or enhancing neuronal inhibition directly, through interaction with specific ion channel subunits, or indirectly, through influencing the synthesis, metabolism, or reuptake of neurotransmitters that act on ionotropic receptors. AEDs are typically grouped according to their principal molecular target (see Table 1.3), although some AEDs are thought to exert their effects through more than one mechanism, and for many the precise mechanism of action is not known (85-87).

Table 1.3: Mechanisms of action of AEDs (adapted from Brodie and Sills, 2011 (87), and Sills, 2015 (86))

Mechanism of action	AED(s)
Sodium channel blockers	
Fast-inactivated state	Phenytoin, carbamazepine, lamotrigine, oxcarbazepine Eslicarbazepine
Slow-inactivated state	Lacosamide
Calcium channel blockers	
Low voltage activated channel	Ethosuximide
High voltage activated channel	Gabapentin, pregabalin
Voltage-gated potassium channel agonists	Retigabine
AMPA glutamatergic receptor antagonists	Perampanel
GABA antagonists	
$GABA_A$ receptor – prolongation of chloride channel opening	Barbiturates
$GABA_A$ receptor – increased frequency of chloride channel opening	Benzodiazepines
Inhibition of GABA transaminase	Vigabatrin
GABA reuptake inhibition (GAT-1)	Tiagabine
Modulation of synaptic vesicle protein 2A	Levetiracetam, brivaracetam
Carbonic anhydrase inhibition	Acetazolamide
Multiple pharmacological targets	Valproate, felbamate, topiramate, zonisamide, stiripentol
Unknown	Cannabidiol

AMPA = α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid; GABA = γ-aminobutyric acid; GAT-1 = GABA transporter 1.

Currently, we have an incomplete understanding of the spectrum of AED effects; the aforementioned mechanisms are merely those as of yet identified (85, 88). Traditionally, mechanisms of action of AEDs have been evaluated on isolated neurons

in vitro. Consequently, we have a good understanding of acute actions of AEDs on single neurons and a limited understanding of AED effects on neuronal interactions and neural network modifications (89). Since epilepsy is a disease of neuronal networks, methods focusing on single neurons might be limiting AED development for DRE. A further factor limiting AED development could be the models used in drug discovery. Most AEDs were discovered serendipitously or by testing a range of compounds in acute seizure models like the pentylenetetrazole or the maximal electroshock model (90). This could be limiting the development of AEDs with unique mechanisms of action. For instance, levetiracetam did not demonstrate efficacy in traditional seizure models used for drug screening, e.g. the pentylenetetrazole and maximal electroshock models, but it was shown to be efficacious in less conventional models such as the kindled and genetic animals (91).

Similarly, our understanding of the mechanisms of drug resistance and the overlap between drug resistance and disease aetiology is incomplete (92). Experimental and clinical data suggest that the basis of DRE is multifactorial (93). Most individuals with DRE are resistant to multiple AEDs targeting different pathways, suggesting that non-specific factors influence DRE or that several mechanisms act in an integrated manner to produce drug resistance (85, 93-96). Alternatively, there could be several specific mechanisms, none of which are addressed by current AEDs. Several hypotheses of drug resistance have been proposed, including the target hypothesis, the multidrug transporter hypothesis, and the intrinsic severity hypothesis (93).

1.5 Mechanisms of drug resistance

1.5.1 The target hypothesis

According to the target hypothesis, DRE results from the loss of target sensitivity due to "a structural or functional change at the site of action" (94). Alterations of the structure or function of ion channels and excitatory neurotransmitter receptors that are the principal targets of AEDs cause pharmacodynamic changes. These target alterations may develop in the course of epilepsy due to external factors, such as AEDs, or they may be genetically determined (96-98).

The target hypothesis originates from extensive research of carbamazepine effects on voltage-gated sodium channels in hippocampal neurons (99). Vreugdenhil and Wadman studied the effect of carbamazepine on sodium currents in hippocampal neurons in a rat kindling model of epilepsy. Carbamazepine selectively shifted the voltage threshold for steady-state inactivation to more hyperpolarised potentials. This shift was most pronounced shortly after kindling, recovering after a period without seizures, which suggests an epileptic activity-dependent block (100). Remy *et al.* studied changes in hippocampal tissue from people with TLE who underwent epilepsy surgery, comparing individuals who were resistant to carbamazepine with responders. They demonstrated that activity-dependent sodium channel block was lost in carbamazepine-resistant cases and preserved in responders. In the same paper, they also demonstrated the absence of the activity-dependent sodium channel block in

isolated dentate granule cells in pilocarpine-treated rats. Based on their experiments using human tissue and rat models, the authors proposed that seizure activity itself may cause a chronic change in sodium channel properties resulting in the loss of carbamazepine sensitivity (101). This mechanism could explain the development of carbamazepine resistance in the course of epilepsy, but it is uncertain whether the effect would extend to other AEDs (95). Furthermore, the loss of carbamazepine effect on sodium-channel properties demonstrated in the kindled rat model was transient and it is uncertain whether it is relevant in chronic epilepsy (100).

Another widely studied AED target are $GABA_A$ receptors, which exhibit significant alterations associated with seizures (95). The $GABA_A$ receptor is a ligand-gated chloride channel that is the main mediator of neuronal inhibition in the central nervous system (CNS). There are several $GABA_A$ receptor subtypes characterised by different subunit composition, regional and cellular distribution, affinity for GABA, ion channel properties, and desensitization rates. Changes in channel composition may affect the receptor sensitivity to benzodiazepines and barbiturates (99).

Mutations or variants affecting the genes encoding $GABA_A$ receptor subunits have been shown to underlie a spectrum of epilepsy syndromes, including idiopathic (genetic) generalised epilepsies, genetic epilepsy with febrile seizures, and epileptic encephalopathies (EE). Functional consequences of these mutations include $GABA_A$ receptor kinetics changes resulting in hyperexcitability (102-104). In addition to genetically determined alterations, there is evidence from clinical and preclinical studies indicating acquired $GABA_A$ alterations in status epilepticus (SE) and chronic epilepsy. The efficacy of benzodiazepines declines as the duration of seizures increases

(105-107). Individuals who experience SE are more likely to develop AED resistance (108). Studies in human epileptic tissue and animal models indicate that the mechanisms underlying the resistance to benzodiazepines could be $GABA_A$ receptor changes, including reduced number of receptors in the target tissue and altered subunit composition and function (109).

Drug-resistant individuals with MTLE-HS have been shown to have a reduction of benzodiazepine-binding sites in the hippocampus greater than what could be attributed to neuronal loss in sclerotic areas (110, 111). Furthermore, the affinity of benzodiazepine binding has been shown to change in drug-resistant focal epilepsy, indicating a biophysical change in $GABA_A$ receptors. [11C]flumazenil positron emission tomography (PET) imaging in drug-resistant focal epilepsy also suggests a loss of $GABA_A$ receptors (112, 113). These research findings indicate a deficit of benzodiazepine-sensitive receptors in some types of DRE, in line with the target hypothesis (109).

Studies in a rat pilocarpine model of epilepsy provide insights into the $GABA_A$ receptor function and pharmacology in the process of epileptogenesis. In this animal model, SE is induced using pilocarpine injections, and following a latent period animals begin to have spontaneous seizures of temporal lobe origin. During animal SE, there is development of rapid functional plasticity of $GABA_A$ receptors with profound reduction of benzodiazepine efficacy (114). Experiments in single dentate granule cells from rat brains combining patch-clamp recordings and single-cell messenger RNA (mRNA) methods show $GABA_A$ receptor subunit expression alterations preceding chronic seizure onset by weeks and correlating with significant changes in receptor

pharmacodynamics (increased sensitivity to zinc blockade and decreased sensitivity to benzodiazepines). These observations indicate that abnormalities in the expression and function of $GABA_A$ receptors represent an important mechanism underlying epileptogenesis (115). Importantly, $GABA_A$ receptor changes have an increased impact in the context of other changes, such as mossy fiber sprouting. $GABA_A$ receptor changes represent only one aspect of the complex process of DRE development (94, 115).

An important limitation of AED target studies in humans is accessing the brain tissue of drug responders as they generally do not undergo epilepsy surgery. This is reflected, for example, in the study by Remy *et al.* where the number of controls was much smaller than the number of cases, and they differed in pathology (101). A limitation of most studies using animal models is the lack of pre-selection of animals according to their *in vivo* AED response (99). Despite some very elegant studies of sodium channel and $GABA_A$ receptor changes in epilepsy, it is unlikely that these changes explain drug resistance to all drugs acting on these targets, let alone other AEDs or AEDs with multiple mechanisms of action. Most people with DRE are resistant to multiple AEDs with a broad range of targets, suggesting that non-specific mechanisms may contribute to drug resistance (116).

1.5.2 The multidrug transporter hypothesis

The multidrug transporter hypothesis postulates that DRE results from inadequate AED concentrations in target tissue as a result of increased efflux via the multidrug

transport proteins due to their modifications or increased expression. The multidrug transporter hypothesis is built on the following premises:

- Most AEDs are substrates of multidrug transport proteins
- The changes in multidrug transporter proteins primarily affect epileptogenic brain tissue, explaining the occurrence of CNS adverse effects in drug-resistant individuals
- Modifications or increased expression of multidrug transporter proteins are a feature of DRE and not epiphenomena of the underlying pathology (98)

Most multidrug transport proteins belong to the ABC (adenosine triphosphate (ATP)-binding cassette) superfamily and they are located on capillary endothelial cells forming the blood-brain barrier (BBB) (97). One of the best characterised transporters in several medical conditions is P-glycoprotein (P-gp), encoded by the *ABCB1* (ATP-binding cassette subfamily B member 1) gene (117, 118). Tishler *et al*. observed increased expression of P-gp mRNA in brain tissues of PWE undergoing surgery, suggesting that insufficient penetration of AEDs to the CNS may underlie drug resistance in epilepsy (119). This report and several subsequent studies demonstrating overexpression of P-gp in the affected brain tissue of individuals with DRE led to the multidrug-transporter hypothesis (95). For instance, Sisodiya *et al.* found increased P-gp expression in resected tissue from individuals undergoing epilepsy surgery due to focal cortical dysplasia (FCD), dysembryoplastic neuroepithelial tumours, and MTLE-HS (120). These findings were replicated in some animal models, suggesting that increased P-gp expression in the brain tissue is associated with decreased AED

concentrations. P-gp inhibitors have been shown to reverse this decrease (118). Observed differences in P-gp expression in humans may be due to genetic factors, exposure to certain drugs, or seizures themselves (121). The genetic factors, in particular *ABCB1* polymorphisms, were explored by Hoffmeyer *et al.* (122). They reported that the *ABCB1* exon 26 polymorphism C3435T affects P-gp expression and function. The TT genotype, present in 24% of the population, was associated with significantly lower expression of P-gp and higher tissue concentration of digoxin (which is a P-gp substrate) (122). In 2003, Siddiqui *et al.* reported a higher frequency of the C3435T CC genotype (associated with increased expression) in individuals with DRE in comparison with drug responders (123), but this has not been consistently reproduced in further studies (14). As a result, the association between variation in *ABCB1* and DRE has been challenged in the past decade. Genetic variation in *ABCB1* will be discussed in more detail in Section 1.8.2.2.

Other assumptions underlying the multidrug transporter hypothesis have also been challenged. Firstly, not all AEDs seem to be substrates of multidrug transporter proteins such as P-gp. Determining whether an AED is a substrate of P-gp is not straightforward since consensus criteria defining the P-gp substrate status have not been established. The available evidence is not equally reliable for all AEDs and the results from different models are often inconsistent. Furthermore, cell culture models used to determine whether a drug is a P-gp substrate may not fully represent the physiology *in vivo*. Taking all of this into account, phenytoin, phenobarbitone, oxcarbazepine and lamotrigine have consistently been shown to be P-gp substrates both *in vivo* and *in vitro*. Acetazolamide, carbamazepine, eslicarbazepine,

levetiracetam, and lacosamide are probable substrates, and valproate, topiramate, and gabapentin are possible substrates. Vigabatrin, zonisamide, and ethosuximide are unlikely substrates (118). As a consequence, the multidrug transporter hypothesis does not adequately explain broad AED resistance. Furthermore, the overexpression of P-gp in DRE may only be an epiphenomenon related to frequent seizures and/or underlying pathology, without affecting AED brain concentrations and efficacy (98).

1.5.3 The intrinsic severity hypothesis

According to the intrinsic severity hypothesis, "common neurobiological factors underlie both epilepsy severity and drug resistance" (85). Drug resistance is therefore an inherent property of the epilepsy related to its severity which can be described as a continuum ranging from mild to severe (85, 94, 96, 99). The intrinsic severity hypothesis originates from observations that high seizure frequency at disease onset is a predictor of DRE. Both the number of seizures at onset (2, 17, 21, 36, 66, 71) and immediately after introducing treatment have been shown to be important predictors of drug resistance (20, 44, 58). One possible explanation of these epidemiological data is that seizure frequency at disease onset reflects the inter-individual differences in inherent epilepsy severity, influencing the response to AEDs (85). An alternative explanation could be that frequent seizures at presentation trigger a pathophysiological process similar to kindling in animal models, resulting in DRE. This is inherently linked to the question whether 'seizures beget seizures' (75). However, symptomatic treatment following the first unprovoked seizure does not seem to

influence the long-term prognosis, despite controlling seizures in the short term (56, 124-126). It could be argued that such findings are not supportive of a kindling-like process, but could be in favour of the intrinsic hypothesis.

Experiments in acute seizure models show that suppression of epileptic activity by increasing doses of AEDs can be counteracted by intensifying the excitatory stimulus (127, 128). If the threshold for seizures is low, it may not be possible to suppress them with a non-toxic dose of any AED, in line with the hypothesis that the factors underlyng epilepsy severity and drug resistance may be shared (85).

While the link between high seizure frequency and drug resistance fits the intrinsic hypothesis well, the seizure patterns in some types of epilepsy are more difficult to explain by intrinsic severity, for example spontaneous remission in some childhood epilepsies that appear resistant at onset, progression over time in EE, and relapse after early remission in mesial temporal lobe sclerosis (85). In conclusion, the intrinsic hypothesis alone does not seem to be sufficient to explain DRE.

1.5.4 Other proposed mechanisms of drug resistance in epilepsy

More recent hypotheses that have received considerable attention are the network hypothesis and the methylation hypothesis (129). The network hypothesis is based on the premise that repeated seizures induce alterations in the brain, including neuronal death, gliosis, neurogenesis, axonal sprouting and synaptic reorganisation, resulting in remodelling of the network. This new, abnormal network is less sensitive to inhibitory

effects of endogenous antiepileptic systems and to the action of AEDs. However, plastic changes are not found only in DRE, but also in epilepsy cases responding well to AEDs. It is currently unclear if changes in DRE are different or more severe than in drug-responsive epilepsy (130).

According to the methylation hypothesis, DRE is a consequence of epigenetic modifications triggered by persistent seizures. Excessive depolarisation of the neuronal membrane may result in persistent epigenetic changes, including deoxyribonucleic acid (DNA) methylation patterns, histone modifications, and non-coding ribonucleic acid (RNA) changes. These epigenetic processes influence gene transcription, with the potential for persistent and dynamic regulation of neuronal gene expression. DNA methylation can suppress gene expression by directly or indirectly affecting the binding of sequence-specific transcription factors. Thus, epigenetic modifications may be mediated by recurrent seizures, resulting in DRE (131). Epigenetic changes could potentially explain observations such as discordance of monozygotic twins and fluctuating epilepsy course seen in some individuals (55). An example of epigenetic modifications that may potentially contribute to increased susceptibility to epileptic seizures following an episode of epileptic activity are changes in potassium channel subunit activity. Potassium channel subunits Kv7.2 and Kv7.3, encoded by the *KCNQ2* (potassium voltage-gated channel subfamily Q member 2) and *KCNQ3* (potassium voltage-gated channel subfamily Q member 3) genes play a crucial role in the repolarisation of neurons and raising the threshold for firing an action potential via non-activating potassium currents (M-currents), therefore regulating neuronal excitability (132, 133). Mutations in *KCNQ2* and *KCNQ3* cause benign familial neonatal

convulsions, an autosomal dominant epilepsy occurring in neonates (134-136), and early-onset EE (137, 138). Mucha *et al.* identified regulatory elements within the *KCNQ2* and *KCNQ3* and showed that the transcription of both genes is facilitated by transcription factor Sp1 (also known as specificity protein 1), and reduced by the repressor element 1-silencing transcription factor (REST). Repression of *KCNQ2* and *KCNQ3* expression by REST results in the inhibition of M-currents, causing neuronal hyperexcitability that may contribute to chronic epilepsy (139). Neuronal expression of REST increases as a consequence of kainate-induced seizures in animal models (140), cerebral ischaemia (141), and inflammatory mediators (139). Thus, long-term synaptic changes resulting from seizures may be mediated by the regulation of *KCNQ2* and *KCNQ3* transcription.

Albeit not formulated as a separate hypothesis, it has been suggested that inflammation is involved in the development of DRE as it can promote epileptic activity. It may significantly contribute to a high seizure frequency and intrinsic severity, but it could also be a consequence of persistent seizures (93, 142, 143). Accumulating evidence supports the role of inflammation in different types of epilepsies. For instance, anti-inflammatory treatments represent the standard-of-care in specific epilepsy syndromes such as infantile spasms where prednisolone and adrenocorticotrophic hormone (ACTH) are highly efficacious in controlling seizures. Specific epileptic disorders such as limbic encephalitis have been associated with the presence of antibodies directed against neuronal antigens (144). In addition, there are data supporting an inflammatory component of the epileptogenic process in focal epilepsies. For example, PET imaging has shown increased translocator protein 18 kDA

(TSPO; biomarker of neuroinflammation) in individuals with temporal lobe epilepsy, suggesting ongoing inflammation (145).

Inflammatory mediators released in the epileptic brain tissue have been reported to significantly contribute to neuronal hyperexcitability and drug resistance in experimental models (146). In addition to promoting local inflammation, cytokines such as tumor necrosis factor-α (TNF-α) can also mediate alterations in the expression and subunit composition of neurotransmitter receptors, affecting neuronal function and excitability (142, 143, 147, 148). For example, TNF-α has been shown to increase the release of glutamate from microglia (149). It upregulates AMPA receptors and causes exocytosis of AMPA receptors in hippocampal pyramidal neurons, thus increasing the number of AMPA receptors on the neuronal surface and enhancing glutamatergic transmission. Furthermore, it causes preferential exocytosis of glutamate receptor 2 (GluR2) subunit-lacking AMPA receptors with changed biophysical properties resulting in enhanced conductance and consequently increased excitatory synaptic strength. To further enhance neuronal excitability, TNF-α also induces endocytosis of $GABA_A$ receptors, resulting in a smaller number of surface $GABA_A$ receptors and a decrease in inhibitory synaptic strength (150). These observations suggest that TNF-α can regulate neuronal circuit homeostasis in a manner that may result in drug resistance. Other examples of immune mediators that may increase neuronal excitability and lower seizure threshold in epilepsy models include interleukin-1β (IL-1β), interleukin-6 (IL-6), and high-mobility group box 1 (HMGB1) (144).

1.5.5 Links between different hypotheses of DRE

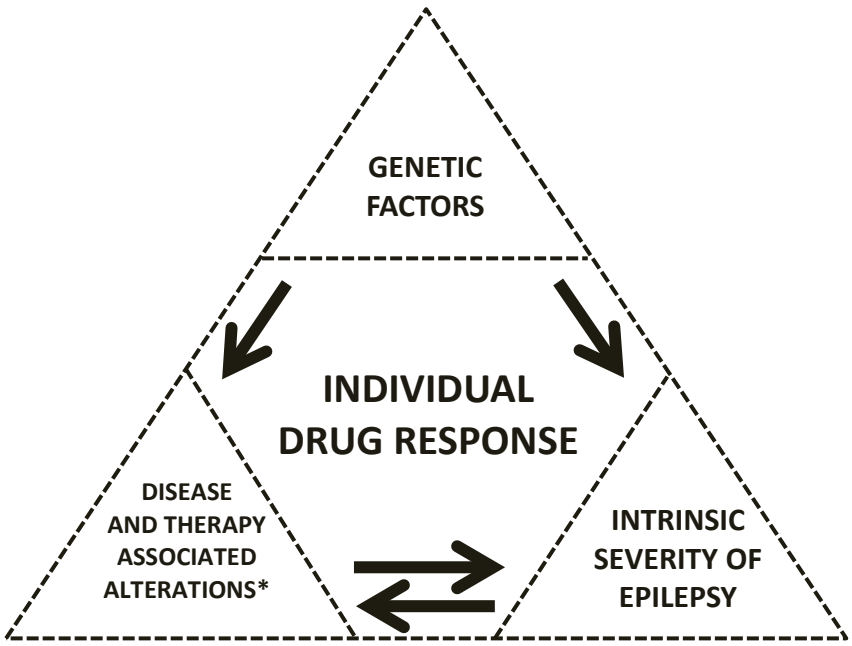
The aforementioned hypotheses of drug resistance are predicated on experimental and clinical studies showing correlations between various factors and drug response. However, associations alone do not permit causality to be ascribed. Studies to determine causality of different factors in DRE are difficult to design considering the intimate relationship between the causes and consequences of seizures (93).

The proponents of the intrinsic hypothesis argue a high seizure frequency results in target, network and BBB alterations, which are seen as merely a reflection of the high intrinsic severity. On the other hand, changes in the network, receptors, ion channels, or the BBB can contribute to epileptogenesis and to intrinsic severity once epilepsy has developed (93).

1.5.6 The role of genetic factors

Genetic factors may affect the intrinsic severity of epilepsy as well as the disease and therapy associated alterations (Figure 1.1). Pharmacogenomics of drug resistance of epilepsy will be discussed in section 1.8.2.

Figure 1.1: Factors influencing individual drug response (modified from Potschka, 2013 (93))



Genetic factors may underlie the intrinsic severity of epilepsy, as well as the disease-associated alterations. *Target and network changes, inflammation, changes to the BBB, AED metabolism or distribution.

1.5.7 Drug resistance – an inherent property of epilepsy or a separate condition?

The multidrug transporter hypothesis and the target hypothesis define drug resistance in terms of neurobiological factors distinct from the epilepsy itself (96). For instance, epilepsy may be caused by different underlying pathologies (including head trauma, structural brain abnormalities, or ion channel mutations), but the associated AED resistance may be a result of independent (e.g. genetic) factors (151). This view implies that individuals with DRE and individuals who respond to AEDs constitute two separate

categories (74) and that specific pharmacoresistance factors determine the lack of response to AEDs (152, 153). Treatments reversing resistance mechanisms, rather than reducing excitatory activity, could therefore lead to treatment response (96).

The intrinsic severity hypothesis on the other hand considers drug resistance as an inherent feature of the epilepsy, rather than a consequence of specific pharmacoresistance factors. It holds that there is a continuum of epilepsy severity, with severe epilepsy syndromes being more challenging to treat. The development of efficacious treatment approaches may benefit from a better understanding of the underlying mechanisms of epilepsy severity variations (96).

1.6 Definition of genomics and pharmacogenomics

Genetics and genomics

The terms “genetic” and “genomic” are often used interchangeably; however, the term genetic usually refers to the study of specific, individual genes and their role, and the term genomic refers to the entire genetic material. The term genome was conceived by Hans Winkler in 1920, combining the words “gene” and “chromosome” to designate the complete genetic makeup of an organism (154). The concept of genome changed and evolved in the decades that followed, from referring to “the haploid chromosomal number”, “the total chromosomal complement”, “the set of chromosomal genes composed of DNA”, and more recently “the totality of the DNA”

(155). When using the term “genomic” in this work, I refer to the latter meaning. As this work focuses on genome-wide approaches, I mostly use the term “genomic” in further text (unless the term “genetic” is more appropriate).

Pharmacogenomics

Pharmacogenomics is the study of the relationship between genomic variation and drug response. The mission of pharmacogenomics is to identify DNA sequence variations associated with the variability of drug pharmacokinetics, pharmacodynamics, and adverse effects in the population. Two important objectives of epilepsy pharmacogenomics are to expand our understanding of biological mechanisms underlying DRE, potentially identifying new drug targets, and to find biomarkers that could predict drug response, eventually serving to guide epilepsy treatment (156). Biomarkers allowing stratification of PWE into potential responders or non-responders, or those at risk of developing adverse events would transform epilepsy clinical practice by allowing a personalised medicine approach to the investigation and management of PWE. Identifying reliable genomic biomarkers of DRE could lead to the implementation of genotyping at an individual level to guide pharmacotherapy (157, 158).

1.7 The human genome sequence and variation

1.7.1 The reference genome

The reference genome is an electronic database that contains the human DNA sequence and serves as an approximation of the DNA of any individual and can be used for comparison of sequencing data and other research efforts. The first draft of the human genome was the product of The Human Genome Project (HGP). Running from 1990 to 2003, HGP mapped nearly all DNA bases (over 3 billion) using Sanger sequencing. The human genome sequence released in 2004 was 99% complete. The final HGP publication predicted between 20,000 and 25,000 protein-coding genes in the human reference genome (159). However, a recent examination of seven extensive proteomic studies estimated that the number of protein-coding genes may be closer to 19,000 (160).

Since the completion of HGP, the reference human genome has been curated and regularly updated by The Genome Reference Consortium (GRC). The Genome Reference Consortium Human Build 38 (GRCh38), released in 2014, is the most recent version. Its predecessor, GRCh37, was released in 2009 (161, 162).

1.7.2 Genomic variation in humans

In a typical individual genome, there are 4.1 million to 5.0 million sites that differ from the reference human genome (163). Genomic variation occurs at several levels; from gross chromosomal abnormalities, structural variations, and indels (insertions or deletions of bases), to single nucleotide polymorphisms (SNPs). Depending on the minor allele frequency (MAF) in the population, the genetic variants can be classified as rare and common. Variants with MAF < 0.5% or 1% are considered rare (164, 165). Most variants found in an individual's genome are common in the population. Only 40,000 to 200,000 variants in a typical individual genome (1 to 4%) occur with a frequency of less than 0.5%. However, on the population level the proportions of both types of variants are the reverse from this, with most observed variants being rare. The final 1000 Genomes data set contained approximately 64 million autosomal variants with a frequency of less than 0.5%, approximately 12 million with a frequency between 0.5% and 5%, and approximately 8 million with a frequency of less than 5% (163).

Single nucleotide polymorphisms (SNPs)

SNPs are single nucleotide (base-pair) changes in the DNA sequence compared with reference. SNPs are by far the most common form of variation, estimated to account for over 95% of all sequence variation (163). The term SNV (single nucleotide variant) is sometimes used interchangeably with SNP. Both terms denote a difference in one base pair between individuals within a population. The term SNP implies that a variant is common in a population, usually defined as MAF ≥ 1% (166). With SNV, there is no implication about frequency.

Indels

Indels describe deletions and insertions less than 50 base pairs in size (167). Indels are estimated to account for approximately 4% of all sequence variation (163).

Copy number variations (CNVs)

CNVs are defined as structural rearrangements increasing or decreasing the DNA content. The size of CNVs is typically defined as larger than 50 base pairs (167). Although SNPs and short indels represent more than 99.9% of the genomic variation, CNVs affect a larger number of bases. Approximately 2,100 to 2,500 CNVs can be found in a typical individual genome, affecting approximately 20 million bases of DNA (163).

Distinction between SNPs and mutations

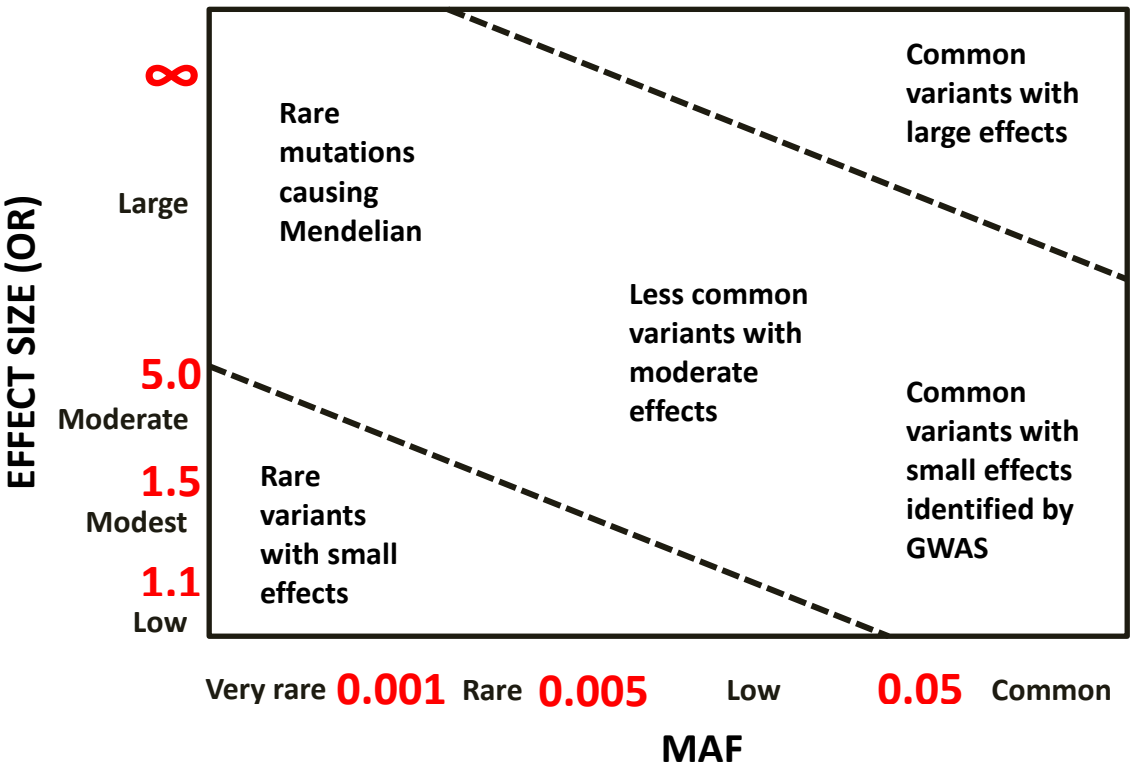
Mutations and SNPs can be structurally equivalent; however, the term SNP is generally used to describe common single base-pair changes, i.e. those found in at least 1% of the population, and the term mutation refers to rare genetic variants. The arbitrary threshold of 1% was established in preparation of the human genome sequence to differentiate common variants (SNPs) and rare variants (mutations). The assumption underlying this definition is that higher prevalence in the population indicates that SNPs have neutral or beneficial effects (168-170). Since the emergence of next generation sequencing (NGS) technologies, it has become evident that genomic variation in the population is more complex than previously thought. The distinction between SNPs and mutations is often not straightforward based on their frequency or disease-causing potential alone. For example, some pathogenic mutations have been

found in more than 1% of the population (171). SNPs can cause amino acid changes, affect mRNA stability and the affinity for transcription factor binding, resulting in functional consequences (172). This might explain why some SNPs are associated with complex diseases. Conversely, many rare variants that meet the mutation definition are not associated with disease, and some mutations are exclusively associated with disease in certain populations (173).

1.7.3 Genomic variation and complex traits

The majority of genetic variation in the human genome is not thought to have pathologic consequences. The challenge is to identify variants increasing risk of disease or acting as disease modifiers. The extent to which genetic variants influence complex traits depends on their MAF and effect size, i.e. the increase (or decrease) in risk carried by a given variant (170). Commonly occurring variants usually have small effect sizes, typically representing a 1.2 to 1.5-fold increase in the relative risk (RR). Genetic variants with high effect sizes are typically very rare (examples are mutations causing Mendelian disorders). Figure 1.2 illustrates the relationship between MAF and the effect of genetic variants on disease risk.

Figure 1.2: Allelic architecture of complex traits (modified from Manolio *et al.*, 2009 (174); Bush and Moore, 2012 (170))



OR = odds ratio; MAF = minor allele frequency. Most of the genetic variants that have been associated with diseases in humans lie on the diagonal between the two dashed lines. Most variants identified in GWAS are common SNPs with small effect sizes (lower right). There are also a few examples of common variants with large effect (upper right) influencing common diseases, usually those that occur in old age. Mendelian mutations with large effect sizes (upper left) are often discovered by linkage studies that focus on familial segregation. Rare variants of small effect (lower left) are very difficult to identify by current genetic methods.

1.7.4 Genetic methods to detect genetic loci contributing to diseases

Common traits have a different underlying genetic architecture compared to rare disorders and as a consequence, the appropriate choice of genetic methods to study specific traits may differ. Linkage and association studies are the predominant

strategies used to identify contributory genetic loci influencing complex phenotypes. Linkage studies focus on familial aggregation and are suitable for identifying rare and highly penetrant variants causing Mendelian diseases. Conversely, association studies seek disease risk alleles at the population level. They can be conducted using a candidate gene approach or genome-wide approach. While both linkage and association studies have their strengths, association analyses are particularly informative in pharmacogenomics. Whilst the candidate gene approach has proven especially valuable in studying the genes known to influence drug pharmacokinetics, genome-wide approaches have the potential to discover variants affecting drug pharmacodynamics (175, 176). Examples of GWAS discoveries expanding our understanding of inter-individual variability of drug response have been the GWAS linking SNPs in *VKORC1* (vitamin K epoxide reductase complex), the warfarin drug target, and cytochrome P450 (CYP) genes *CYP2C9* and *CYP4F2* with warfarin response, and the association of *CYP2C19* genotype with the effect on platelet aggregation and clinical response to clopidogrel (177, 178).

1.7.4.1 Linkage studies

Linkage studies are suitable for discovering rare variants with large effect sizes and may require genotyping a relatively small number of genetic markers. Before genome-wide approaches were available, linkage mapping of large families, with several affected members, facilitated the discovery of numerous genes causing Mendelian diseases (170). An example of a successful linkage study in epilepsy was the

identification of *SCN1A* (coding for the sodium voltage-gated channel type I alpha subunit; $Na_v1.1$) as the causal gene for generalised epilepsy with febrile seizures plus (GEFS+) (179, 180). Following this discovery, *SCN1A* was also linked to Dravet syndrome (181). Generally, linkage mapping is unsuitable for common traits with polygenic inheritance where multiple variants contribute to overall genetic risk, as the effect size of each risk allele is not sufficient to be detected this way (182). However, there are a few exceptions, including the identification of the *APOE* (apolipoprotein E) locus in early onset Alzheimer's disease (183, 184). Over the past few years, there has been a resurgence of interest in family-based linkage analyses coupled with NGS. Several research groups have used linkage in combination with whole-exome, whole-genome, or targeted NGS to identify new disease susceptibility genes (185). For example, Bowden *et al.* recognised that the linkage signal in the Insulin Resistance Atherosclerosis Family Study (IRASFS) originated from only a few families. They performed targeted sequencing of the chromosome interval restricted to the lineage peak, identifying a rare c.133G>C (p.Ala45Arg) mutation in the adiponectin gene (*ADIPOQ*) (186). In the future, linkage analysis of NGS data is expected to be even more widely used.

1.7.4.2 Candidate gene studies

Candidate gene studies focus on associations of selected gene variants and phenotypes of interest. Identifying suitable candidate genes depends on the pre-existing knowledge of their biological, pathophysiological, or functional link with the

disorder. If the understanding of biological mechanisms for a trait are limited, this approach will be biased. This is the case in epilepsy, where the poor understanding of drug resistance mechanisms represents a major limitation for candidate gene studies (187). Nevertheless, this approach has revealed the association between the human leukocyte antigen *HLA-B*1502* and carbamazepine-induced Stevens-Johnson syndrome in people of South Asian origin and *HLA-A*3101* in Caucasians (188, 189).

1.7.4.3 Genome-wide association studies (GWAS)

GWAS is a powerful approach for assessing known DNA variants across the genome to identify genetic risk factors for common diseases or traits (170). Unlike candidate gene analyses, GWAS do not require *a priori* assumptions for involved genes and consequently have the potential for discovery of novel pathophysiological mechanisms underlying a phenotype (190). Despite having significant limitations, such as case selection biases, genotyping errors, and/or the potential for false-positive and false-negative results, GWAS remain a powerful research tool for examining the genome and providing insights into the biological pathways involved in common traits (191).

The concept of GWAS is based on the common disease-common variant hypothesis (192, 193). The common disease-common variant hypothesis is built around the premise that common traits are caused by genetic variants common in the population, each causing a small increase in the disease risk. The disease risk (or effect size) carried by an individual variant must be small in comparison to that found in Mendelian diseases, otherwise these diseases would affect a much larger proportion of the

population. Consequently, a single variant can explain only a small proportion of the total phenotypic variability attributed to genetic factors (170).

GWAS were conceptualised by Risch and Merikangas in 1996 as a technique suitable to identify risk variants with small effects that cannot be detected with linkage analysis (194). However, the implementation of GWAS in practice followed a decade later, when the microarray technology, required to genotype hundreds of thousands of SNPs in large cohorts became available (195). Another important factor enabling GWAS was the completion of the International HapMap project. The HapMap Project, initiated in 2002, identified the location and density of common SNPs and it provided information on the haplotypes, i.e. sets of associated SNP alleles in a region of a chromosome that are inherited together, and the SNPs that tag those regions (170). Population-specific differences in genetic variations were catalogued. Phase I data were published in 2005, Phase II in October 2007, and Phase III in 2010 (196-198).

The 1000 Genomes project that followed was designed to provide a reference set of markers to allow the imputation of genotypes. The 1000 genomes project final analysis, published in 2015, lists "88.3 million variants, including 84.4 million bi-allelic SNVs, 3.4 million bi-allelic indels, and 60,000 structural variants (SVs) consisting of large insertions, deletions, inversions, and multi-allelic CNVs. The final release also included approximately 475,000 multi-allelic SNVs and indels" (164).

The haplotype information provided by the HapMap and 1000 Genomes projects has been successfully used in GWAS (170). In most chromosomal regions there are only a few common haplotypes accounting for most of the inter-individual variation in a

population (196). By genotyping only the SNPs that are representative of each haplotype (tagging SNPs), utilising the concept of linkage disequilibrium (LD; ”the degree to which an allele of one SNP is inherited or correlated with an allele of another SNP within a population” (170)), it is possible to capture most of the information on the pattern of genetic variation in the region. For example, by using 250,000 to 500,000 tagging SNPs that are representative of each LD block and predicting most of the remaining SNPs, it is possible to capture almost as much mapping information as with several million SNPs (170, 199-201).

When the GWAS methodology first became available, scientists got the opportunity to interrogate the entire human genome without predefined hypotheses and at levels of resolution that had previously been unattainable (191). The first GWAS was published in 2005, identifying the complement factor H gene as a risk factor for age-related macular degeneration (202). In 2007, a landmark GWAS study with approximately 17,000 participants identified 24 independent association signals for seven common diseases: bipolar disorder, coronary artery disease, Chron’s disease, hypertension, rheumatoid arthritis, type 1 diabetes, and type 2 diabetes (195). This was followed by several other GWAS successes in prevalent diseases with complex phenotypes. As of August 2018, the National Human Genome Research Institute and the European Bioinformatics Institute (NHGRI-EBI) GWAS catalog contains 3,541 unique GWAS studies with 69,969 unique SNV-trait associations (https://www.ebi.ac.uk/gwas/).

The GWAS field is advancing rapidly, with an emphasis on larger sample sizes and advancing technology, for example high-density genotyping arrays and NGS (203, 204). Over the past decade, several consortia have formed to meet the large sample size

requirements that are beyond the reach of single research groups. Increasing sample sizes and large-scale meta-analyses will likely continue to lead to new genetic discoveries, increasing the list of variants and genes associated with diseases and allowing the evaluation of disease heterogeneity. GWAS using SNP array data are increasingly being replaced with GWAS using NGS data (205). GWAS can also be conducted using mRNA data in expression quantitative trait loci (eQTL) studies. Additionally, GWAS results can be integrated with gene expression data, methylation/acetylation, and protein-protein interaction data (186, 204, 206, 207).

1.7.4.4 Estimating heritability from genome-wide data

The purpose of GWAS is to identify individual loci associated with a phenotype. In GWAS of complex traits, it is common practice to estimate how much of the phenotypic variability in the studied population can be explained by each associated SNP, by all associated SNPs combined, and by all SNPs in the GWAS. This is then compared with the total heritability of the phenotype estimated from pedigrees (208).

Narrow sense heritability

Heritability in the narrow sense (h^2) is defined as the proportion of the total phenotypic variance between individuals in a population due to additive effects of all causative genetic variants (209). Heritability can range from zero (no genetic contribution) to 100% (all phenotypic variance in the population is due to genetic variation). Traditionally, heritability was estimated by studying pedigrees, comparing the observed and expected phenotypic resemblance between parents and children,

siblings, or monozygotic and dizygotic twin pairs. When estimating heritability by comparing relatives, shared environmental factors can act as a confounder (210), hence the need for an analysis model that specifies how much of the expected phenotypic similarity is due to the genetic and environmental factors. For example, in analyses of monozygotic and dizygotic twins, the phenotypic similarity attributed to shared environment is assumed to be the same (211). By utilising genetic marker data, more recent methods enable heritability estimation without the requirement for such assumptions (212).

Missing heritability

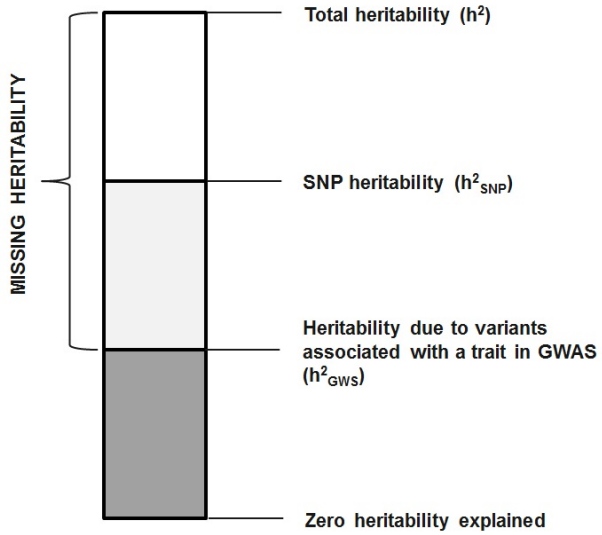
For the majority of common diseases studied with GWAS, the proportion of variance explained by genome-wide significant SNPs (h^2_{GWS}) is much less than the estimated total heritability of the trait (h^2). The gap between h^2 and h^2_{GWS} is referred to as the missing heritability (174, 204, 213). It has been posited that in the case of complex traits, heritability is hidden rather than missing (214). Possible explanations include rare variants with large effects, structural variation, or a large number of common variants with individual effect sizes insufficient to reach genome-wide significance (174, 215). The missing heritability issue triggered the development of methodology to estimate the amount of phenotypic variance accounted for by all SNPs used in a GWAS in conventionally unrelated individuals (see below) (209).

SNP-based heritability

SNP-based heritability (h^2_{SNP}) is the proportion of phenotypic variance explained by a defined set of SNPs; this can be either all SNPs used in a GWAS (genotyped and

imputed) or all genomic variants from NGS (216, 217). In complex traits caused by a large number of variants with small effect sizes, consistent with a model of polygenic inheritance, more of the heritability can be explained by estimating heritability from the entire GWAS data set than just the significantly associated SNPs (h^2_{GWS}) (208).

Figure 1.3: SNP heritability explains some of the missing heritability (modified from Witte *et al*., 2014 (208))



Methods to estimate h^2_{SNP} are typically applied to conventionally unrelated individuals (216). Any random pair of unrelated individuals has common distant ancestors and consequently shares a small amount of genomic variation. These matching variants occur at random, i.e. they are independent. It is possible to make estimates of heritability by comparing these genetic similarities. If individuals sharing the same phenotype also tend to have genetic similarities, this represents evidence that heritability is non-zero. Estimating heritability from data of unrelated individuals has two key advantages. Firstly, as unrelated individuals are unlikely to share common

environment, the estimates are unlikely to be confounded by common environmental factors. Secondly, the sample sizes in GWAS analyses are significantly larger than in pedigree studies. This approach has shed some light on the missing heritability debate. The classic example is human height. The heritability of human height has been estimated at 80% from pedigree studies (218, 219). By 2008, several GWAS studies had identified approximately 50 SNPs associated with height, jointly explaining only approximately 5% of phenotypic variance (i.e. h^2_{GWS} is 5%). Increasing the sample size represented the opportunity to discover additional associated variants using the GWAS design (209). Indeed, a larger 2014 study by the GIANT consortium increased the number of loci to over 400, increasing the h^2_{GWS} estimate to 16% (220). In 2010, Yang *et al.* introduced a mixed linear model approach to estimate h^2_{SNP} using GWAS data and showed that a set of SNPs captured on a genotyping array explained 45% of variance in height, which is a significant proportion, but still less than h^2 (80%) (216). A smaller h^2_{SNP} in comparison with h^2 is expected because not all causal variants can be perfectly tagged by SNPs on a genotyping array, and because the SNPs on the array may not be in perfect linkage diseuqilibrium (LD) with the causal variants. More of the variance can be explained by increasing the number of SNPs tested or by using NGS data. However, when working with GWAS data, even the latest genotyping arrays do not include all SNPs. Most microarrays contain only common variants (present in > 1% population), so h^2_{SNP} does not reflect the contribution of rare SNPs and other types of genomic variation, such as structural variants (e.g. CNVs), epigenetic effects, and a whole host of 'omics' (proteomics, lipidomics, transcriptomics, metabolomics, etc.) (209, 217).

1.8 Genomic variation and drug resistance in epilepsy

1.8.1 Evidence for the genetic basis of epilepsy

There is no consensus estimate of the heritability of epilepsy, reflecting the heterogeneity of the condition and the variability of study designs and statistical approaches used to measure it (221). Twin studies of epilepsy susceptibility have been performed since the 1960s when Lennox first showed that monozygotic twins are more likely to be concordant for epilepsy more often than expected by chance (222). More recent epidemiological studies show significantly higher concordance rates in monozygotic compared with dizygotic twins, indicating that genetic factors are important in the pathogenesis of IGE (GGE), but also have a contribution in focal epilepsies (223-227). Heritability estimates from twin studies range from 8 to 88% (228-230). Potential reasons for the variable estimates can be attributed to the inclusion of different epilepsy syndromes, as well as different terminology and methodology across the studies (230).

A significant proportion of epilepsies are thought to have a genetic background. This includes monogenic causes (familial or *de novo*) and epilepsies with complex or polygenic inheritance where multiple genes, susceptibility alleles, and genetic modifiers play a role (231, 232). A polygenic genetic architecture has generally been associated with common epilepsies with complex patterns of inheritance, although the actual underlying genetic risk factors have proven difficult to discover (233, 234). Due

to the gaps in knowledge and the fact that genetic testing is not routinely applied in epilepsy management, we do not have a reliable estimate of the proportion of epilepsies that have a genetic underlying aetiology (235, 236). However, epidemiology studies quantifying the genetic contribution for epilepsies have provided valuable insights on the population level. Reliable estimates of epilepsy prevalence in relatives of PWE have been established by the Rochester Epidemiology Project. The overall risk of being diagnosed with epilepsy before the age of 40 is increased approximately threefold in relatives of PWE compared with the general population. The increase in risk is higher for idiopathic (genetic) generalised epilepsies (IGE; GGE) than for focal epilepsies (221). In addition, we have a good understanding of the genetic structure of epilepsy subpopulations where routine genetic screening is commonly used, for example EE, and preliminary estimates for the frequency of some monogenic epilepsies in the overall epilepsy population. A recent epidemiology study including close to 6,000 adults and children reported the prevalence of Dravet syndrome and tuberous sclerosis in the epilepsy population as 2.1% and 1.2%, respectively. The frequency of chromosomal/monogenic causes of epilepsy (excluding Dravet syndrome and tuberous sclerosis) was reported as 3.9% (236). There is substantial evidence that IGE (GGE) have a significant genetic contribution. This genetic basis is thought to be complex in most cases and monogenic in a minority. Pathogenic variants of *SLC2A1* are estimated to account for up to 1% and CNVs for approximately 3% of cases (237).

A number of genes causing or contributing to individual epilepsy syndromes have been discovered over the past two decades (238). Wang *et al.* recently performed an

extensive search of online databases and identified 977 genes associated with epilepsy. They categorised them as follows:

- 84 epilepsy genes causing syndromes with epilepsy as the only or the most prominent symptom
- 73 genes causing abnormalities of brain development and epilepsy
- 536 epilepsy-related genes causing syndromes with abnormalities of several organ systems where epilepsy is one of the features
- 284 potential epilepsy genes (232)

As there is continuous progress in the identification of novel epilepsy genes, this list may not be comprehensive.

This large base of genetic knowledge started with the discovery of a number of gene mutations causing rare Mendelian forms of epilepsy using linkage analysis. The first one was *CHRNA4*, coding for the neuronal nicotinic acethylcholine receptor subunit alpha 4, identified in 1995 in a large Australian kindred with autosomal dominant nocturnal frontal lobe epilepsy (ADNFLE) (239). Multiple other epilepsy gene discoveries in monogenic familial epilepsies followed, mostly coding for ion channels, including *KCNQ2* and *KCNQ3* in families with benign familial neonatal convulsions (134, 135), and *SCN1A* in GEFS+ (180). *SCN1A* mutations were later shown to have a wide phenotypic spectrum, ranging from milder epilepsy disorders such as GEFS+, febrile seizures plus (FS+), and simple febrile seizures to severe myoclonic epilepsy of infancy or Dravet syndrome (181, 240).

NGS technologies have accelerated the identification of disease-causing mutations through targeted gene panels, whole exome sequencing (WES), and whole genome sequencing (WGS). Most notably, NGS has facilitated the discovery of *de novo* causative mutations in sporadic diseases. A common method to achieve this is the comparison of exome or genome sequences of the affected individual and both (unaffected) parents (i.e. trio analysis). The application of these methods in the epilepsy field has helped identify pathogenic mutations in a wide spectrum of genes as causes of EE, expanding the list of proteins and pathways involved in epilepsy pathogenesis beyond ion channels. Some examples of epilepsy genes identified with NGS include: *de novo SCN8A* (sodium voltage-gated channel alpha subunit 8) mutations in EE and in milder phenotypes such as benign infantile seizures (241, 242), *KCNQ* (potassium voltage-gated channel subfamily), previously shown to cause benign familial neonatal convulsions (135), as a cause of early-onset EE (137, 138), *de novo* mutations in *HCN1* (potassium/sodium hyperpolarisation-activated cyclic nucleotide-gated channel 1) in individuals with early infantile EE resembling Dravet syndrome (243), *GABRA1* (GABA type A receptor alpha1 subunit) and *STXBP1* (syntaxin binding protein 1) in individuals with *SCN1A*-negative Dravet syndrome (244), mutations in *SYNGAP1* (synaptic Ras GTPase-activating protein 1) in individuals with EE, absences and myoclonic seizures, and mutations in *CHD2* (chromodomain helicase DNA binding protein 2) in EE associated with myoclonic seizures and photosensitivity (245-248).

These discoveries have significantly improved the clinical management of individuals with specific epilepsy syndromes and provided the basis to develop targeted treatments. The numerous rare variants in epilepsy genes that can explain epilepsy in

selected families, however, constitute only a fraction of the heritability of common epilepsies. As of now, the molecular basis of the majority of common epilepsies remains elusive (232).

Gross chromosomal abnormalities, detectable by karyotyping, can also cause epilepsy. One such example is the ring chromosome 20 syndrome (249). The frequency of CNVs is higher in PWE, especially when epilepsy is associated with learning disability, autism, or psychiatric disorders such as schizophrenia (237, 250). CNVs have been established as a risk factor for EE, generalised and focal epilepsies using array comparative genomic hybridisation (aCGH) or high-density SNP and CNV arrays, providing key insights into the genetic structure of common epilepsies. The first CNV linked with epilepsy was the 15q13.3 microdeletion, reported in individuals with idiopathic (genetic generalised) epilepsy (GGE) (251). Microdeletions of 15q11.2 and 16p13.11 are two additional examples of CNVs that are well established risk factors for epilepsy (252-255).

These observations indicate that rare CNVs carrying intermediate risks may represent a sizeable fraction of the heritability of common epilepsies. Pathogenic CNVs are thought to be the underlying cause of epilepsy in 3 to 5% of EE cases (256, 257).

Epilepsy GWAS

Several large GWAS studies have explored the genetic susceptibility loci of common epilepsies. The GWAS efforts have resulted in the establishment of several international consortia: EPIGEN (The Epilepsy Genetics Consortium), EPICURE, EMInet (Epilepsy and Migraine Integrated Network), The ILAE Consortium on Complex Epilepsies, and EpiPGX.

The first large-scale GWAS of GGE identified two genome-wide significant loci: 2p16.1 and 17q21.32. Analyses of subgroups revealed significant associations of the 2q22.3 locus with absence epilepsy and 1q43 with juvenile myoclonic epilepsy (JME). In addition, suggestive evidence for an association was detected at 2q24.3, close to the *SCN1A* gene, suggesting its role in common epilepsies (258). Another GWAS linked MTLE-HS with febrile seizures to common variants within the locus containing *SCN1A* (259). These findings were remarkable because they expanded the genetic spectrum of *SCN1A*-related epilepsies from rare syndromes caused by mutations in the gene to common epilepsies caused by common variants representing risk factors (260).

The GWAS efforts in epilepsy reached a new level with the establishment of the ILAE Consortium on Complex Epilepsies, bringing together several consortia to investigate common genomic variation in epilepsy. The ILAE Consortium has published two GWAS meta-analyses. The original ILAE meta-analysis included 8,696 individuals with epilepsy and 26,157 controls and utilised the phenotypic and genotypic data from 12 cohorts. To allow combining the phenotypic and genome-wide association data from five sites, investigators used standardised imputation and quality control (QC) to enable the joint

analysis of data acquired on different genotyping platforms. Meta-analysis including all epilepsy cases identified loci at 2q24.3 (rs6732655, c.2590-424T>A), implicating *SCN1A*, and at 4p15.1, harbouring *PCDH7* which encodes protocadherin 7. The association with *SCN1A* further consolidated the importance of its variants as risk factors for common epilepsies. Due to the significant overlap of samples, this GWAS is not considered a formal replication of *SCN1A* associations identified in previous studies (260). In contrast, protocadherin 7 had previously not been associated with epilepsy. Meta-analysis of the GGE subgroup identified a single signal at 2p16.1, implicating vaccinia-related kinase 2 (*VRK2*) or Fanconi anaemia complementation group L (*FANCL*). In the sub-analysis analysis of focal epilepsy, no SNP reached genome-wide significance. However, a SNP at 2q24.3, harbouring *SCN1A* (rs12987787, c.4284+591A>G) was close to reaching genome-wide significance (261). rs12987787 is in high LD with the *SCN1A* variant (rs7587026, c.-142+4684G>T) reported in the GWAS of mTLE-HS and febrile seizures (259).

The recently published extended ILAE meta-analysis included 15,212 individuals with epilepsy and 29,677 controls, combining the original ILAE cohort with additional 6,516 cases and 3,460 controls. As a consequence, the extended analysis does not represent a formal replication of the original ILAE analysis. Meta-analysis including all epilepsy cases confirmed two previously identified associations at 2p16.1, harbouring *FANCL*, and 2q24.3, harbouring *SCN1A*. It also led to the identification of a new genome-wide significant locus at 16q12.1, harbouring two potential candidate genes: HEAT repeat containing protein 3 *(HEATR3)* and bromodomain containing protein 7 *(BRD7)*. In contrast with the original ILAE analysis, the locus at 2q24.3 also reached genome-wide

significance in the analysis of focal epilepsy. In addition, the extended ILAE meta-analysis lead to the identification of 14 genome-wide significant loci associated with GGE, 8 of which were novel, and 2 novel loci associated with MTLE-HS. In detailed analyses of the associated loci, some signals were only significant in a single epilepsy subtype, whereas others were suggestive of pleiotropic effects. However, since the sample sizes in these epilepsy subtype analyses were small, the results should not be overinterpreted. The loci with genome-wide significance from all analyses corresponded to 146 genes in total. A scoring system with multiple criteria was used to establish the putative risk genes for epilepsy. These were defined as the genes with the highest score at each locus, including "seven ion-channel genes *(SCN1A, SCN2A, SCN3A, GABRA2, KCNN2, KCNAB1,* and *GRIK1)*, three transcription factors *(ZEB2, STAT4,* and *BCL11A)*, the histone modification gene *BRD7*, the synaptic transmission gene *STX1B*, and the pyridoxine metabolism gene *PNPO"* (262).

The ILAE GWAS demonstrated the value of pooling all types of epilepsy in one sample, as well as analysing sub-phenotypes. Larger sample sizes would have enabled further analyses of epilepsy sub-phenotypes.

How to improve the yield of epilepsy GWAS?

Epilepsy GWAS performed up to date have uncovered a relatively small number of risk variants compared to studies in other neurologic conditions. This most likely reflects a highly heterogeneous genetic background and likely a large number of common susceptibility alleles carrying very small risks, requiring large cohorts of tens of

thousands of cases for definitive GWAS studies (260). Although the ILAE Consortium meta-analysis included the largest cohort of PWE published to date, the sample size is modest compared to those in some other complex disorders. For instance, GWAS of schizophrenia and migraine have included cohorts of over fifty thousand cases and identified several dozen independent risk loci (263). The recently completed GWAS of insomnia has included in excess of 1.3 million cases and has identified 956 genes linked to the sleep disorder (264). Importantly, epilepsies are a heterogenous group of disorders. Analysing them as a group may have limitations in terms of discovering genetic risks. Sub-phenotyping and performing GWAS limited to sub-syndromes may prove more successful in the future. This approach has already proven beneficial in the ILAE Consortium on Complex Epilepsies GWAS meta-analysis where GGE and focal epilepsies were analysed as sub-phenotypes (261). A further dissection of epilepsy sub-phenotypes may be possible in the future as the epilepsy consortia accumulate sufficiently large cohorts with well-defined phenotypes.

1.8.2 Evidence for the genetic basis of drug resistance in epilepsy

Genomic variation is thought to be at least partly responsible for differences in response to AEDs among PWE; however, the heritability (i.e. extent of genetic contribution) of drug resistance in epilepsy has not been quantified yet. Using family studies to define the heritable component of phenotypic variance is much more challenging when applied to drug response than disease susceptibility. It is usually difficult to find pedigrees with multiple members affected with the same disease and

well-defined drug-response phenotypes; consequently, the contribution of heritability to the variability in drug response may be unknown (19). Epilepsy is not an exception. Whilst there is abundant evidence for the heritability of epilepsies from twin studies, candidate gene association studies, GWAS, and NGS approaches (223, 225, 226, 229, 230, 265), little is published on the genetic contribution to DRE. The available evidence for the heritability of DRE comes from animal studies, twin studies, and candidate gene association studies. Few studies have attempted to identify the genetic component of DRE in humans as finding suitable pedigrees or twin pairs affected by epilepsy is difficult. Generating heritability estimates in animal models is easier as it allows the elimination of some of the variability by using the same strain, aetiology, and controlling for environmental factors. In contrast with the small number of studies exploring the heritability of common epilepsies in humans, there is an abundance of candidate gene association studies and studies elucidating drug resistance in selected epilepsy syndromes. Whole genome approaches have not been extensively utilised for this purpose yet.

1.8.2.1 Evidence from heritability studies

One study has assessed the heritability of drug response in 37 twin pairs concordant for epilepsy (27 monozygotic, 10 dizygotic). Twin pairs were concordant for epilepsy type and had no antecedent environmental factors (such as brain trauma) that could potentially contribute to their epilepsy, allowing the assumption that their shared genetic predisposition was most likely the underlying cause of epilepsy. Clinical

outcomes of epilepsy were defined as seizure presence or absence in the year preceding the study assessment and epilepsy severity was assessed by a neurologist on a scale of 1 to 6, where 1 was remission and 6 was severe DRE. A high correlation for clinical outcome and severity of epilepsy was observed in twin pairs (0.62 for all twin pairs combined; 95% CI 0.42, 0.80), with no difference between the monozygotic and dizygotic twin pairs. The high correlation in clinical outcome was consistent across the different epilepsy syndromes and concordant outcomes were observed even with discordant treatments. This study thus failed to demonstrate that the clinical outcome is determined by genetic factors distinct from those determining susceptibility to the disease. Considering the predominance of GGE in the sample, it is possible that the epilepsy genetic susceptibility shared by twins contributed substantially to the clinical outcome (266). Nevertheless, the study was small and its results do not mean a role for genetic factors in drug resistance can be discounted. Experiments in animal models of focal symptomatic epilepsy indicate DRE could have a genetic component. Löscher *et al.* discovered that it was possible to select phenytoin responders and non-responders from a population of amygdala-kindled outbred Wistar rats. In contrast to responders, non-responders did not show any significant increase of after-discharge threshold following an intraperitoneal application of phenytoin. They showed that the response to phenytoin did not depend on the plasma level differences, kindling parameters, or electrode placement, but appeared to be an inherent property of the selected rats (267). An attempt to reproduce these results in Sprague-Dawley rats was unsuccessful (268). Cramer *et al.* compared five different inbred rat strains. Even though inbred strains are more genetically homogenous than outbred strains, they

demonstrated within-strain differences in response to phenytoin (269). Overall, animal data provide limited support for a genetic contribution to DRE.

1.8.2.2 Evidence from studying individual genes/syndromes

An extensive body of epilepsy pharmacogenomics studies has been published to date, mostly utilising candidate gene approaches. A large number of variants in different genes have been proposed to contribute to DRE. In a recent review, Balestrini and Sisodiya proposed a working classification of the genetic factors influencing drug resistance according to the mechanisms affected by genetic variation: pharmacokinetics, pharmacodynamics, and gene mutations capable of causing epilepsy (14).

i) Genetic factors influencing AED pharmacokinetics

Pharmacokinetics refers to the absorption, bioavailability, distribution, metabolism, and excretion of drugs. The onset, duration, and intensity of the effect of a drug are determined by its pharmacokinetics (270). Two processes have received considerable attention in relation to AEDs: metabolism with microsomal CYP enzymes and transport across the BBB which can both affect the response to AEDs in the population. The oxidation of several AEDs is catalysed by one or more of the CYP enzymes. The variability in the metabolism is the result of the presence of many CYP isoenzymes and polymorphisms in their coding genes (271). The first study

showing an association of genetic polymorphisms with metabolism of an AED was published in 1998 by Mamiya *et al*. They showed that polymorphisms in genes encoding drug-metabolising enzymes *CYP2C9* and *CYP2C19* influenced phenytoin doses used clinically (272). Since then it has been firmly established that *CYP2C9* polymorphisms are a significant determinant of the phenytoin metabolism rate (273, 274). The CYP2C9 enzyme accounts for approximately 90% of phenytoin metabolism (273, 275, 276). Two *CYP2C9* haplotypes are associated with poor metabolism of phenytoin: *CYP2C9*2* (rs1799853, c.430C>T, p.Arg144Cys) and *CYP2C9*3* (rs1057910, c.42614A>C, p.Ile359Leu). Individuals with these variants have a significantly higher risk of experiencing adverse drug reactions with phenytoin therapy (277-279). It has been shown that the *CYP2C9*3* allele is associated with an approximately 50 mg lower maximum phenytoin dose compared with wild-type homozygotes in regular usage in a series of individuals with epilepsy (274).

There are preliminary data showing that the metabolism of the following drugs is influenced by polymorphisms of genes coding for their corresponding major metabolising enzymes:

- carbamazepine - *CYP3A4* gene and epoxide hydroxylase 1 gene (*EPHX1*) polymorphisms
- clobazam - *CYP2C19* gene polymorphisms
- lamotrigine - uridine diphosphateglucuronosyl transferase gene polymorphisms

- phenobarbital - *CYP2C19* gene polymorphisms

- valproate - *CYP2C9*, *CYP2A6* and *CYP2B6* gene polymorphisms

- zonisamide - *CYP2C19* gene polymorphisms (14, 280).

Most of the aforementioned associations were established by studying drug levels and toxicity, with only a few exceptions where the association with drug response was determined. For instance, while the association of *CYP2C9* haplotypes with phenytoin toxicity has been well established, much less research has been done on the association of *CYP2C9* variants with drug resistance. Two studies have suggested a role of *CYP2C9* polymorphisms in DRE (281, 282); however, such results need to be interpreted with caution. A recent study showed that *EPHX1* polymorphisms influence plasma carbamazepine levels, but no association with drug response to carbamazepine was shown. The plasma level of carbamazepine was not associated with carbamazepine response (280).

In summary, the relationship between the variation in genes encoding the main AED metabolising enzymes and the metabolism, levels, toxicity, and response to individual AEDs is complex. Since AEDs are metabolised via a number of different enzymes, it is unlikely that any single CYP or other variant would completely explain DRE.

Another important aspect of pharmacokinetics is AED transport across the BBB via ATP-dependent transport proteins (multidrug transport proteins); located on capillary endothelial cells forming the BBB. Multidrug transport

proteins are members of the ABC superfamily. According to the multidrug transporter hypothesis, drug resistance results from inadequate AED concentration in the target tissue as a result of increased efflux via the multidrug transport proteins due to their modifications or increased expression (98). The most extensively studied transporter across medical conditions is P-gp, encoded by the *ABCB1* gene (117). Tishler *et al.* found an increased expression of P-gp mRNA in resected brain tissue of 11 out of 19 PWE who underwent epilepsy surgery. The variability in P-gp expression may be due to genetic variation, environmental factors, or both (119). There is a high number of polymorphisms in the *ABCB1* gene, with over 800 variants described to date, some of which alter the expression of the multidrug transporter (122, 283-285). The c.3435C>T (rs1045642) polymorphism is the most widely investigated. Siddiqui *et al.* observed increased expression of P-gp in individuals with the rs1045642 CC genotype, associated with DRE (123). While some studies replicated this finding (286-289), others either found a reverse association (290, 291), or no association (292-294). Contrary findings prompted several meta-analyses, mostly with negative results (295, 296). However, two recent meta-analyses including 8,604 cases from 30 studies and 8,331 cases from 23 studies, respectively, identified a significant correlation between the *ABCB1* rs1045642 CC genotype and DRE in Caucasian, but not Asian populations (297, 298). Another meta-analysis including 734 cases from 13 studies in Caucasian and Asian populations found an association between the TT genotype and drug resistance (299). Meta-analyses highlighted several difficulties when

comparing different studies, including different definitions of DRE. Some studies used stratification by epilepsy syndrome or seizure type, whereas others did not. Various studies included different uses of AEDs (monotherapies, combination therapies). Some AEDs, for example valproate, are not P-gp substrates. The use of valproate in a significant proportion of individuals in studies may have confounded the results of some meta-analyses, making it difficult to accurately determine whether the rs1045642 polymorphism is truly correlated with DRE (298, 299). In addition, the effect of a single polymorphism could be modified by variants in other genes and the environment. In conclusion, various analyses have resulted in contradictory conclusions regarding the relationship between C3435T and other *ABCB1* polymorphisms and DRE. The link between *ABCB1* polymorphisms and DRE thus remains unertain (14).

ii) Genetic factors influencing AED pharmacodynamics

Pharmacodynamics refers to the target binding, effects on the target, and chemical interactions of a drug at its site of action. Structural of functional changes at the site of action, resulting in altered pharmacodynamics, can be the cause of drug resistance. Alterations of AED targets, including ion channels and neurotransmitter receptors, can be genetically determined (300). To my knowledge, all genetic variants influencing pharmacodynamics are also genetic causes of epilepsy; I will thus discuss them in the next

section.

iii) Gene mutations capable of causing epilepsy

There are several examples of rare genetic forms of epilepsy where the causative mutation also determines drug resistance. Perhaps the most well-known example are mutations in *SCN1A* gene encoding $Na_v1.1$ which is a key player in action potential generation and propagation in neurons. Missense mutations in *SCN1A* commonly result in a mild phenotype (e.g. GEFS+) with seizures that are easily controlled by AEDs. In contrast, truncation mutations or missense mutations affecting the channel pore region cause a severe disturbance in protein function resulting in Dravet syndrome (severe myoclonic epilepsy of infancy), characterised by drug resistance as one of its main features. Loss of function of $Na_v1.1$. in Dravet syndrome results in reduced sodium current and consequently reduced neuronal excitability, so an epilepsy phenotype would not be expected (240, 301). Studies in a mouse model of Dravet syndrome have provided initial insights into the neurobiological changes associated with the loss of one copy of *Scn1a*. They helped establish that the mutant sodium channel subunits are expressed predominantly in GABA-ergic inhibitory interneurons. Consequently, reduced sodium current density and impaired neuronal excitability is observed in hippocampal inhibitory interneurons, but not excitatory neurons, leading to the prevailing hypothesis that impaired GABA-mediated inhibition leads to network hyperexcitablity and

seizures in Dravet syndrome (302, 303). Sodium channel blockers (which bind to the α-subunit of the voltage-gated sodium channel) decrease the sodium current in the inhibitory interneurons even further, resulting in seizure worsening (304). Importantly, some studies in human-derived induced pluripotent stem cells have not provided supporting evidence for an interneuron-specific decrease in excitability, suggesting that the mechanisms underlying seizures and drug resistance are more complex than previously thought (305). Changes in sodium currents and neuronal excitability may depend on the specific *SCN1A* mutation and several modifying factors such as the individual's genetic background and developmental stage (306). Nevertheless, understanding the underlying mechanism of seizures has enabled better clinical management which includes avoiding AEDs that exacerbate seizures, such as carbamazepine, and prioritising clobazam, valproate, stiripentol, topiramate, and the ketogenic diet (307). The new insights have also facilitated the exploration of targeted treatments such as cannabidiol (approved by the FDA in 2018 (308)) and fenfluramine which has recently shown clinically meaningful seizure improvements in open label studies and case series (309-311). Fenfluramine is not yet available on the market. A Phase 3 study comparing fenfluramine with placebo has recently been completed, but the results have not been published yet (312).

Since common variants around *SCN1A* have been identified as risk factors for a wide range of common epilepsies (259, 261), it is possible that

variation in the *SCN1A* gene could also be associated with drug resistance in common epilepsies. Considering that multiple AEDs target sodium channels, investigating the *SCN1A* gene variation in relation with the clinical response could lead to important insights in DRE (313). Several studies have explored the association of *SCN1A* SNPs with drug resistance. Tate *et al.* explored the haplotype-tagging SNPs in *SCN1A* and reported the association of the rs3812718 (c.603-91G>A) genotype with the maximum dose of phenytoin and carbamazepine in regular usage in PWE (274). This SNP is located in a splice donor site which determines the alternative splicing of exon 5 encoding a voltage sensor domain of the sodium channel (274, 314). The association of serum phenytoin levels at maintenance dose and rs3812718 genotype was detected in a second study in a Chinese population (315). However, a further study in Austrian individuals did not replicate these findings (316). One study in a Japanese population has found an association of the AA genotype with the lack of response to carbamazepine (317); whereas several other studies in a range of populations did not succeed to replicate the association between the rs3812718 SNP and drug resistance to carbamazepine, oxcarbazepine, and multiple AEDs (280, 318-322). One study found a marginally statistically significant association between a different *SCN1A* SNP, rs10188577 (c.265-699A>G), genotype and drug resistance (322). Margari *et al.* explored the association of 14 *SCN1A* SNPs (located both in exons and non-coding areas) with drug resistance in a paediatric epilepsy cohort. Three intronic SNPs (but none of the exonic SNPs) were statistically significantly associated with treatment outcomes

(rs6730344, c.3430-9985G>T; rs6732655, c.2590-424T>A; rs10167228, c.2044-168A>C) (313). Of note, rs6732655 was also found to be an epilepsy risk factor in the ILAE GWAS of genetic determinants of common epilepsies (261).

While the mechanism of drug resistance in monogenic epilepsy syndromes caused by *SCN1A* mutations is at least to some extent explained, the relationship between *SCN1A* polymorphisms and drug resistance in common epilepsies appears to be more complex. There is limited support for the role of rs6730344, rs10167228, rs3812718, and rs6732655 in DRE, at least in some populations. These SNPs are intronic, potentially resulting in splice site alterations and changes in the protein (313, 320). More research is needed to establish whether the variation around *SCN1A* affects drug resistance in the general epilepsy population.

Another extensively studied example of a gene causing epilepsy and drug resistance is *SLC2A1* (solute carrier family 2 member 1). Mutations of the gene result in glucose type 1 transporter (GLUT-1) deficiency and consequently impaired transport of glucose across the BBB which affects brain development and function (323). Individuals with GLUT-1 deficiency have low CSF and brain glucose presenting with a variety of phenotypes including myoclonic-astatic epilepsy, early childhood refractory absence epilepsy, rarely juvenile absence epilepsy, and paroxysmal exercise-induced dyskinesia (324, 325). The severity of the phenotype depends on the extent

of the reduction of the GLUT-1 transporter function (326, 327); however, genotype-phenotype correlations are not clean-cut or easy to establish. Individuals with the same mutation may exhibit phenotypic heterogeneity in terms of the range and severity of clinical symptoms, allowing the possibility of genetic modifiers influencing the phenotype (328). Epilepsy associated with GLUT-1 deficiency is resistant to classic AEDs, but amenable to treatment with the ketogenic diet, which generates ketones replacing glucose and restoring brain energy metabolism (329). GLUT-1 deficiency syndrome is a great example of how understanding the underlying cause enables personalised treatment.

There are a few other examples of epilepsy syndromes where the underlying gene abnormality at leaset partially explains drug resistance to conventional AEDs and where precision therapy might be implemented to treat epilepsy. These include the use of mTOR (mammalian target of rapamycin) inhibitors in epilepsy associated with tuberous sclerosis (330) and the potential use of memantine in individuals with EE due to *GRIN2A* (glutamate NMDA receptor subunit 2A) mutations (331).

Mutations in the above listed genes explain resistance to individual or multiple AEDs in a few rare genetic epilepsy syndromes. In these syndromes, the mutations causing epilepsy also likely contribute to drug resistance, which would be consistent with the intrinsic hypothesis. However, the phenotypic heterogeneity frequently seen in

monogenic epilepsy syndromes allows for the possibility of genetic or other modifiers affecting both the seizure severity and drug resistance (96). These modifiers could be either syndrome-specific or acting across a broader spectrum of epilepsies.

1.8.2.3 Evidence from genome-wide approaches

Severe mutations in genes like *SCN1A* are rare, but it is possible that common variation in these genes could partially explain drug resistance in common epilepsies (92). GWAS have the potential to help elucidate this. GWAS studies in the field of epilepsy have mainly focused on identifying common susceptibility loci conferring an increased epilepsy risk (332). With increased collaboration among scientific groups enabling larger sample sizes, better exploration of drug response with GWAS approaches is becoming feasible. However, as of now, only one GWAS focussing on drug response has been published. Speed *et al.* reported a GWAS comparing responders and non-responders in a prospective cohort of newly-diagnosed PWE. Suggestive evidence for association with response to AEDs was found for the following loci: 6p12.2, 9p23 and 15q13.2. The authors argued that any individual common variant is unlikely to explain more than 4.4% of the variation in the treatment outcomes in newly-diagnosed epilepsy (333). There are no published GWAS on drug resistance in chronic epilepsy.

1.9 Rationale to study the genetic basis of drug resistance

Genomic medicine is transforming the field of epilepsy. Large-scale international collaborations and progress in sequencing technology have enabled the exploration of the genetics of common and rare epilepsies, explaining the aetiology in an increasing proportion of PWE (334). So far, the field has been most successful in establishing the genetic causes of EE, but there have also been novel gene discoveries in common epilepsies. For example, NGS have recently expanded our knowledge of the genetic causes of familial focal epilepsies and IGE (335). GWAS of common epilepsies suggest a complex genetic architecture, with several loci, each carrying a low disease susceptibility risk (261). There is still potential for novel discoveries in all epilepsies using both approaches, GWAS as well as NGS. Both approaches will continue to benefit from large international collaborative efforts, allowing pooled analyses of large cohorts with well-defined phenotypes.

The rationale to study the genetic basis of drug resistance in general epilepsies is supported by the following existing knowledge:

- Drug resistance is a major unmet medical need in epilepsy, representing a problem that is both difficult to predict in an individual and difficult to manage (156).

- DRE occurs across the whole spectrum of AEDs and epilepsies, regardless of the underlying aetiology, implying that it could be driven by a non-specific mechanism (116).
- Albeit limited, combined evidence from animal models and studies in individuals with epilepsy provides sufficient support for a genetic basis of DRE. Increased risk for DRE may be the consequence of genetic variation in genes coding for proteins involved in drug pharmacokinetics, pharmacodynamics, and genes that (when mutated) cause epilepsy. So far, efforts to establish specific pharmacogenomic markers of DRE have produced less robust or conflicting findings needing further validation, for example variation in genes encoding drug transporters and AED targets (14).
- There is a significant unmet need for genomic markers of DRE to guide epilepsy management.

GWAS is a valid method to explore the genetic basis of DRE for the following reasons:

- The GWAS experimental design is well suited to explore common genomic variation.
- It is reasonable to assume that DRE could potentially be caused by common genomic variation, considering that drug resistance is a common trait in the epilepsy population, affecting approximately 30% of PWE (20, 28, 43). Furthermore, since variants in genes mediating drug response are not expected

to be subject to selection pressure, functionally important variants may be common in the population (19).

- In the field of epilepsy, the GWAS approach has already resulted in discoveries impacting patient care, most notably the risk alleles associated with severe adverse drug reactions to AEDs (336). Similarly, the GWAS approach has the potential to provide insights into the prognosis of epilepsy.
- The unbiased nature of GWAS allows the exploration of DRE without the limitation to known pathways which have been extensively studied using candidate approaches.

Potential positive GWAS outcomes might contribute to the understanding of the molecular basis of DRE. Better understanding of DRE could facilitate targeted or precision therapies to address the neurochemical deficits leading to resistant seizures in an important subset of PWE (337-339). Rare successful examples of targeted molecular therapies to treat extremely resistant epilepsy syndromes have already attracted significant interest and enthusiasm from the epilepsy community (311).

1.10 Aims and hypotheses

The aim of this work is to explore the genetic contribution to DRE and identify common genomic variants associated with drug resistance in epilepsy of any aetiology. The identification of such variants could have two potential implications:

i) Pointing to novel biological pathways, increasing our understanding of DRE
ii) A variant highly associated with DRE could be useful as a biomarker to predict the outcome early in the disease course

This work has one principal research question:
Does common genomic variation contribute to drug resistance in epilepsy, regardless of the epilepsy type?

The following hypotheses were tested:

Hypothesis 1:

Common genomic variants are associated with drug resistance in common epilepsies.

To test Hypothesis 1, I performed a GWAS analysis comparing individuals with DRE and individuals with drug-responsive epilepsy (GWAS1).

Hypothesis 2:

Identified variants are determinants of drug response rather than epilepsy susceptibility variants.

To test Hypothesis 2, I performed an additional GWAS analysis comparing individuals with DRE and healthy individuals (GWAS2). If Hypothesis 2 is accurate, the loci identified in GWAS1 will not be replicated in GWAS2.

Hypothesis 3:

Drug resistance in epilepsy has a polygenic inheritance component.

To test Hypothesis 3, I performed a SNP heritability analysis, using the data from GWAS1. The heritability analysis explores the joint contribution of all variants included in the GWAS analysis of drug resistance.

CHAPTER 2: ESTABLISHING A PHENOTYPIC DATABASE OF DRUG RESPONSE IN EPILEPSY

2.1 Introduction

As stated in the preface, I completed this work as part of collaborative, European Commission 7th Framework funded project: EpiPGX (Epilepsy Pharmacogenomics: delivering biomarkers for clinical use). This was an extensive effort with contributions from numerous researchers. The full list of researchers is provided in the preface and my contribution is outlined in Table 1.

In this chapter, I outline the EpiPGX organisation and scope of work, focussing on the procedures for case selection for the GWAS and heritability analyses. I describe the results of the phenotyping work and the cohorts available for analyses. Lastly, I discuss the relevance and the impact of the chosen case definitions.

2.1.1 DRE definitions

Appropriate phenotype definitions and classification of cases and controls are of key importance in GWAS, particularly when cohort sizes are restricted. The analysis should

involve a representative sample of cases and controls to increase the likelihood of detecting a relevant signal and to eliminate potential confounding factors (340).

Defining drug resistance in epilepsy has been challenging, reflecting the complex and diverse course of the disorder. Before the ILAE proposed a unified concept of DRE in 2010, no consensus definition existed and drug resistance was defined differently across studies, making comparisons difficult. Definitions used in large studies of epilepsy outcomes before 2010, as well as the ILAE definition, are summarised in Table 2.1.

Table 2.1: DRE definitions used in studies of epilepsy outcomes

Reference	Definition
Arts *et al.*, 1999 (20)	"Failure to be ≥ 3 months seizure-free at 6 months after diagnosis"
Berg *et al.*, 2001 (21)	"The failure or lack of seizure control with ≥ 2 first-line AEDs with an average seizure frequency ≥ 1 seizure/month for 18 months and no more than 3 consecutive months seizure-free during that interval"
Dlugos *et al.*, 2001 (23)	"Failure to be ≥ 6 months seizure-free at 2 years after diagnosis"
Camfield and Camfield, 2003 (22)	"An average of ≥ 1 seizures in each 2-month period during the last year of follow-up, despite treatment with at least 3 AEDs as monotherapy or polytherapy"
Kwan and Brodie, 2004 (84)	"DRE is defined as < 1 year of seizure freedom despite treatment with at least 2 AEDs"
ILAE definition Kwan *et al.*, 2010 (26)	"A failure of adequate trials of 2 tolerated, appropriately chosen and used AED schedules (whether as monotherapy or in combination) to achieve sustained seizure freedom"

Berg and Kelly applied the definitions mentioned above (apart from the ILAE) prospectively in a *de novo* cohort of 613 children. Absolute agreement between any two definitions was above 0.80, but there were discrepancies in the kappa statistics ("agreement after correcting for agreement expected by chance alone"). Importantly, all definitions correlated well with the likelihood of a two-year and five-year remission at the last follow-up (52). The ILAE definition has been validated against the Berg, Kwan and Brodie, and Camfield and Camfield definitions in a recent study by Tellez-Zenteno *et al*. The ILAE definition was found to be reliable (inter-observer kappa score 0.77) and correlated well with the three older DRE criteria which retain their clinical significance. The differences in the proportion of the population classified as having DRE using the four definitions were not statistically significant (341). In conclusion, different DRE definitions seem to capture the resistant population well. As stated by Berg *and* Kely in the conclusion of their work, having a choice of several definitions of DRE is not necessarily a negative thing as it is unlikely that a single definition of DRE will suit all purposes. Depending on the research purpose, one definition may be more appropriate than another, or sometimes a slight modification (which the ILAE framework allows) may be required (26, 52).

The main advantage of the ILAE definition, in addition to providing consensus, is that it allows an early diagnosis of drug resistance in clinical settings and it has been widely used in studies of epilepsy outcomes over the past years (341-345). However, it is less readily applicable in some types of studies, for example clinical trials of new investigational drugs and pharmacogenomic studies.

Early phase pharmaceutical trials of novel AEDs are almost without exception conducted in individuals with DRE; however, the way this population is defined is usually guided by the design and the ethical imperative to include only individuals who have the potential to benefit from experimental treatments. Typically, at least two to eight seizures per month over a period of a few weeks (often eight) is required, as well as having failed a minimum of two to three AEDs. Sometimes an additional criterion requiring "no continuous 21-day seizure-free periods" is added (330, 346-351). These criteria are tailored to the design that usually includes a few weeks of screening during which the baseline seizure frequency can be assessed.

Similarly, epilepsy pharmacogenetic and pharmacogenomic studies often tailor the DRE criteria to increase the probability of demonstrating efficacy, resulting in a multitude of definitions in the published literature. Often, these are based on the criteria proposed by Siddiqui *et al.*, requiring a minimum of four seizures over a period of 12 months despite treatment with more than three AEDs (123). The number of AED trials required by different authors may vary from two to four (123, 352, 353). Other definitions used in these studies may require higher seizure frequencies, for instance one seizure per month or ten seizures over a period of 12 months (352, 353). The use of the ILAE definition in pharmacogenomic studies published prior to 2015 was rare, but has been increasing in the past three years, although it is still less common than variations of the Siddiqui *et al.* definition (294, 313, 353-355). Drug responsiveness is almost without exception defined as one year of seizure freedom (123, 352, 353), in contrast with the ILAE proposal recommending either "seizure freedom for 12 months, or for a minimum of three times the longest pre-intervention inter-seizure interval,

whichever is longer" (26). This may be due to the often retrospective phenotypic data collection where the exact intervals between seizures may not be readily available.

2.1.1.1 Addressing pseudo-resistance

Uncontrolled seizures due to factors other than DRE are often referred to as pseudo-resistance. Before determining the outcome of epilepsy as drug resistance, it is important to consider potential causes of pseudo-resistance, including (but not limited to): poor adherence, other conditions with symptoms mimicking seizures (e.g. NES), inappropriate medication for the epilepsy or seizure type, and insufficient dose of the correct medication (356, 357). The latter category also includes situations where suboptimal seizure control is accepted to balance the benefits, risks, lifestyle issues, and personal preferences. In some individuals, complete seizure control cannot be achieved without unacceptable adverse effects (358). Rarely, treatment with the most efficacious AED for the condition may not be appropriate due to the potential risks. For example, valproate remains one of the most efficacious AEDs to treat some types of epilepsy, such as CAE and JME (359-361), but it is associated with higher rates of congenital malformations and developmental disorders as a result of intrauterine exposure than other AEDs (362-365). In 2014, the National Institute for Health and Care Excellene (NICE) published a guideline recommending to avoid valproate use in pregnancy (366). More recently, the CMDh (Coordination Group for Mutual Recognition and Decentralised Procedures), a medicines regulatory body representing the European Union, and the EMA (European Medicines Agency) endorsed more stringent measures. According to the new requirements, "valproate should not be used in pregnancy unless the woman has a form of epilepsy that is unresponsive to

other AEDs", and it "should not be prescribed to women of child-bearing potential who are not enrolled in a pregnancy prevention programme" (367). As for some women valproate may be the only AED that controls their seizures, switching to a different AED due to pregnancy planning or pregnancy may mean accepting suboptimal seizure control (368).

Other situations where drug response may be challenging to ascertain include provoked and reflex seizures. Provoked seizures are seizures that are associated with provoking external factors. These include fever, sleep deprivation, and menstrual cycle changes. In clinical practice it may be difficult to establish the causal association between the seizure and the external factor with certainty and thus determine whether a seizure was provoked or a result of AED failure. The same may be true for reflex seizures which are epileptic events precipitated by external stimuli (e.g. light flashes), internal mental processes, or both (369, 370).

According to the ILAE DRE definition, potential causes of pseudo-resistance have to be considered when determining the outcome of individual AED trials. For the trial to be informative, the AED has to be "appropriate for the epilepsy and seizure type" and "applied at adequate dosage". Consequently, seizures that occur in the context of non-adherence are not considered a sign of drug resistance. The ILAE framework does not specifically address situations where suboptimal seizure control is accepted by the affected individual due to lifestyle reasons, safety, or tolerability. Provoked seizures are interpreted as AED failure and reflex seizures are not mentioned (26).

2.1.2 Consistency of data entry and classification in multicentric studies

Agreeing on and carefully applying definitions is especially important in international multicentric studies. Regardless of the definitions used, there may be cross-centre diversity of practice and interpretation. Definitions of DRE usually do not include instructions on how to classify an AED trial as adequate, how to interpret the outcomes of AED trials after epilepsy surgery, when to consider an individual non-adherent to medication, etc. As a consequence, classification of cases can depend heavily on the interpretation. Such challenges should be expected and managed to allow reliable classification of phenotypes.

2.1.3 Aim of the phenotyping work

The aim of phenotyping within EpiPGX was to provide reliably classified cases for GWAS studies, including the drug resistance GWAS. Establishing a collection of DNA samples linked with a database of phenotypic data was a prerequisite to allow meaningful genomic analyses.

2.2 The EpiPGX Consortium

The EpiPGX Consortium (referred to as the Consortium in further text) was established in 2011, with the aim to identify genomic markers of epilepsy treatment response to improve and guide epilepsy treatment.

2.2.1 EpiPGX sites

The following centres are part of the Consortium: UCL, ULB, IGG, EKUT, SEIN, UKB, RCSI, HSCT, deCODE, UL, UMCU, ULIV, Imperial, UGLA, GABO:mi (full names are provided in the Statement of contribution section).

2.2.2 Scope of work

The Consortium addresses several aspects of epilepsy pharmacogenomics, including:

i) Early and late response to AEDs

ii) Drug resistance

iii) Adverse drug reactions of AEDs

iv) Developmental malformations caused by specific AEDs

The scientific work is organised into ten work packages (WP). Each WP has a leading centre and a varying number of collaborating parties. The work on each WP focusses either on one of the epilepsy pharmacogenomics aspects, or activities that feed into all other WP (see Table 2.2). The EC funding started in November 2011 and finished in

October 2015. The Consortium is still active and analyses of the data continue to this date.

Table 2.2: EpiPGX WP

WP	Scope of work	Leading site
WP01	Characterisation of pharmacogenomic phenotypes and phenotype quality assurance	ULB
WP02	Genome-based biomarkers of early treatment response in newly-diagnosed epilepsy	ULIV
WP03	Genome-based biomarker discovery for resistance to multiple AEDs	UCL
WP04	Genome-based biomarker discovery for late response to specific AEDs	EKUT
WP05	Genome-based biomarker discovery for specific ADRs	RCSI
WP06	Genome-based biomarker discovery for valproate teratogenesis	BHSCT
WP07	Core analytic and bioinformatic processing	UL
WP08	Development of diagnostic tests and in silico database	deCODE
WP09	Project Management	GABO:mi
WP10	Dissemination and Training	UCL

Shortly following its establishment, the Consortium generated detailed consensus phenotype definitions relating to different aspects of epilepsy pharmacogenomics and a corresponding project-specific case record form (CRF). The CRF formed the basis for a centralised electronic phenotypic database enabling direct entry and sharing of the data across the sites. Parallel to these efforts, there was ongoing recruitment of individuals with epilepsy to establish a collection of DNA samples linked with the phenotypic database. PWE were recruited across all clinical EpiPGX sites. DNA samples were shipped to DeCODE for SNP-genotyping and WES. GWAS and other analyses to identify biomarkers associated with pre-defined epilepsy pharmacogenomic

phenotypes started soon after the genotype data became available. The present work is part of WP03 which focuses on broad AED resistance. The methods for phenotypic collection in WP03 are described in detail below.

2.3 Methods

2.3.1 Ethics

All study participants provided written, informed consent for genetic analyses. Local institutional review boards reviewed and approved study protocols at each contributing site (see Appendix 5). This work was undertaken in accordance with national and international regulations listed in Appendix 5. The EpiPGX project was also guided by an Ethics Advisory Board consisting of three expert ethicists, representing legal, philosophical, and social aspects of ethical issues in genetics, one rotating scientific advisory board member, and one rotating EpiPGX WP leader. The Ethics Advisory Board ensured a high standard of ethics was maintained in the overall pursuit of the project.

2.3.2 Recruitment and inclusion criteria

Individuals with epilepsy were recruited from clinical centres in the UK, Ireland, Belgium, Netherlands, Germany, and Italy, as outlined in Table 2.3. They had to meet the following inclusion criteria:

1. Diagnosis of epilepsy
2. Written informed consent (or assent where appropriate) signed by the individual or their legal representative

Table 2.3: EpiPGX recruiting centres

EpiPGX site	Recruiting centre	Ancestry
UCL	Outpatient Epilepsy Clinics at the Department of Clinical and Experimental Epilepsy (DCEE), National Hospital for Neurology and Neurosurgery (NHNN), London	British
ULB	Epilepsy Clinics at UZ Gasthuisberg, Katholieke Universiteit Leuven, and Hôpital Erasme, Université Libre de Bruxelles, Brussels	Belgian
IGG	Outpatient Epilepsy Clinics at Instituto Giannina Gaslini, Genova	Italian
EKUT	Outpatient Epilepsy Clinics at the University Hospital Tübingen	German
SEIN	Tertiary Referral Center for Epilepsy at Stichting Epilepsie Instellingen Nederland, Heemstede	Dutch
UKB	Klinik für Epileptologie, Universität Bonn	German
RCSI	Specialised Epilepsy Clinic at Beaumont Hospital, Dublin	Irish
UMCU	Outpatient Epilepsy Clinics at University Medical Center Utrecht	Dutch
ULIV	Outpatient Epilepsy Clinics at The Walton Centre for Neurology & Neurosurgery, Liverpool	British
UGLA	Epilepsy Unit, University Department of Medicine and Therapeutics, Western Infirmary, Glasgow	British

Ethnically matched healthy controls were obtained from publically available sources, as outlined in Table 2.4.

Table 2.4: Sources of healthy controls

Contributor	Description	Ancestry
National Blood Bank Service (NBS)	Healthy blood donors from the UK Blood Service	British
Wellcome 1958 Birth Cohort	Individuals born in a single week in 1958 in England, Scotland, and Wales, participating in The National Child Development Study (NCDS)	British
Trinity Student Study	Healthy young adult volunteers of Irish ancestry	Irish
Belgian donors	Blood donors and healthy volunteers of Belgian ancestry	Belgian
KORA	Epidemiologically recruited cohort from the Northern region of Germany	German
PopGEN	Epidemiologically recruited cohort from the Southern region of Germany	German
Italian SP1 and SP5	Healthy employees from Epilepsy centres in Italy	Italian

2.3.3 Procedures to ensure uniform phenotyping across the sites

Several processes were put in place to ensure uniform phenotyping across the sites:

i) Consensus definitions

ii) Unified CRF and centralised phenotypic database

iii) Phenotyping manual

iv) Phenotyping workshops

v) Evaluation of inter-rater agreement across the sites

These processes are described in more detail in further text below.

2.3.4 Definitions of pharmacogenomic phenotypes

Seizures and epilepsy syndromes were classified according to the 1989 ILAE terminology (Commission on Classification and Terminology of the ILAE, 1989), taking into account clinical information, imaging, and electroencephalography (EEG) findings. Criteria for IGE (GGE) comprised: "tonic-clonic, absence, or myoclonic seizures with generalised spike-wave discharges on EEG and no evidence of an acquired cause" (261). Cases with GTCS or absence seizures alone and a non-informative EEG remained unclassified. Consensus definitions for the relevant pharmacogenetic categories were developed by the EpiPGX clinical principal investigators (SMS, JC, ND, CD, HL, AGM) based on clinical experience and evidence from the literature (Appendix 2). Phenotype definitions relevant for this work are listed below.

Definition of drug resistance used in this work

The consensus definition of DRE was designed to allow meaningful classification of cases using medical records as well as promoting harmonisation across several sites and nations. It is based on the ILAE consensus proposal, which defines drug resistance as "a failure of adequate trials of two tolerated and appropriately chosen and used AED schedules (whether as monotherapies or in combination) to achieve sustained seizure freedom" (26). According to the ILAE definition, individuals with very rare seizures (for example one seizure in 12 months) may be classified as drug resistant (74) which could be a disadvantage in the context of pharmacogenomics research where the intention is to identify biomarkers of clinically meaningful DRE phenotype. Hence, the clinical principal investigators agreed to use the threshold of four seizures per year

that is frequently used in pharmacogenetic and pharmacogenomic studies (123, 352, 353). As a result, the EpiPGX consensus definition of DRE was formulated as follows: "seizures recurring at a frequency of ≥ 4/year over the year preceding the latest data entry, despite adequate trials of ≥ 2 tolerated and appropriately chosen and used AED schedules, whether as monotherapies or in combination."

An AED trial was classified as adequate if it was applied at an adequate dose for a sufficient amount of time. An AED trial was classified as appropriate if it had been previously shown to be efficacious, preferably in randomised controlled studies. Minimum therapeutic doses for an average adult person were defined by the EpiPGX principal investigators (SMS, JC, ND, CD, HL, AGM, JWS, GJS) based on clinical experience, the World Health Organization (WHO) defined daily doses (DDD) (371), and evidence from the literature (Appendix 2). It is important to note that the agreed appropriate AED daily doses only apply to monotherapy trials and that the list was used as a guidance rather than a set of strict rules. Clinical judgement was required to evaluate the adequacy of AED trials in the context of polytherapy, extreme low or high weight, and for AED trials taking place in an individual's childhood. Laboratory reports of AED levels were taken into account if available. If the AED levels were below the local reference range while the individual was taking a stable dose of the AED and there were no signs indicating CNS toxicity, the AED trial was considered inadequate.

Individuals with non-epileptic seizures (NES) and those known to be non-adherent were excluded. Individuals who had undergone epilepsy surgery were classified as drug-resistant, provided they fulfilled the criteria for drug resistance prior to surgery.

Definition of drug responsiveness

According to the EpiPGX consensus definition, drug responsiveness was defined as freedom from seizures for ≥12 months up to the latest recorded visit. This is a slight departure from the ILAE definition where the outcome of an AED trial is categorised as "seizure-free" (Category 1 response) if "the treatment results in seizure freedom for 12 months, or for a minimum of three times the longest pre-intervention inter-seizure interval, whichever is longer" (26). Considering the phenotypic data were collected retrospectively, not taking into account the pre-intervention inter-seizure intervals, the EpiPGX definition of remission was thought to be appropriate and in line with other epilepsy pharmacogenetic and pharmacogenomic studies published in the literature (352, 353). Inter-seizure intervals are not consistently documented in medical records and the lack of information could have prevented the classification of cases. The downside is that some individuals with very rare seizures could have been classified as responders.

Unclassified cases (undefined response)

An individual remained **unclassified** if he/she:

1. Was systematically non-adherent to medication
2. Had a history of NES
3. Had one to three seizures in the 12 months prior to the latest recorded visit
4. Had an unknown number of seizures in the 12 months prior to the latest recorded visit
5. Had only one or no adequate trials of tolerated and appropriately chosen and used AED schedules

6. Had newly-diagnosed epilepsy with seizures stopping after introducing an AED, but not enough time had passed to categorise the outcome as seizure-freedom
7. Was not treated with adequate doses of AEDs in the 12 months prior to the latest recorded visit (while continuing to experience seizures)
8. There was insufficient information in the medical records to allow classification

2.3.5 Phenotypic data collection

Medical records were accessed and relevant clinical information was extracted using the EpiPGX-specific CRF (Appendix 1). The CRF was designed to capture a broad spectrum of phenotypic data in PWE, including:

1. General information (demographic data, information on DNA collection)
2. Epilepsy diagnosis
3. Comorbidities
4. Neurological examination
5. Seizure types
6. Seizure frequency data and remissions data
7. Non-medical epilepsy treatment (surgery, vagal nerve stimulation)
8. Medical investigations (imaging, electroencephalography (EEG))

9. Detailed AED history data per AED (maximum dose, serum level if available, treatment duration, compliance, efficacy and detailed information on adverse drug reactions, reasons for discontinuation if applicable)
10. Pregnancy outcomes (congenital malformations)
11. Classification of outcomes (pharmacogenomic phenotypes)

Data were entered in the EpiPGX database, developed specifically for the purposes of the project to improve the homogeneity of phenotyping and data sharing within EpiPGX (Appendix 4). It is hosted on a server at deCODE and it can be accessed online by members of the Consortium.

Cases were classified into relevant pharmacogenomic categories (phenotypes) using consensus definitions (Appendix 2). The EpiPGX data entry manual provided decision algorithms and guidance on phenotyping and data entry (Appendix 3). Even with all of this in place, a degree of clinical judgment was required to accurately classify the cases. Case classifications were entered in a designated section of the EpiPGX database. This has enabled researchers across the Consortium to quickly identify informative cases for different analytical efforts.

2.3.6 Inter-rater agreement

Inter-rater reliability was checked across the Consortium and internally at UCL. To test cross-centre phenotyping consistency, trained fellows at seven EpiPGX sites independently phenotyped a set of ten anonymised medical records. Over 500

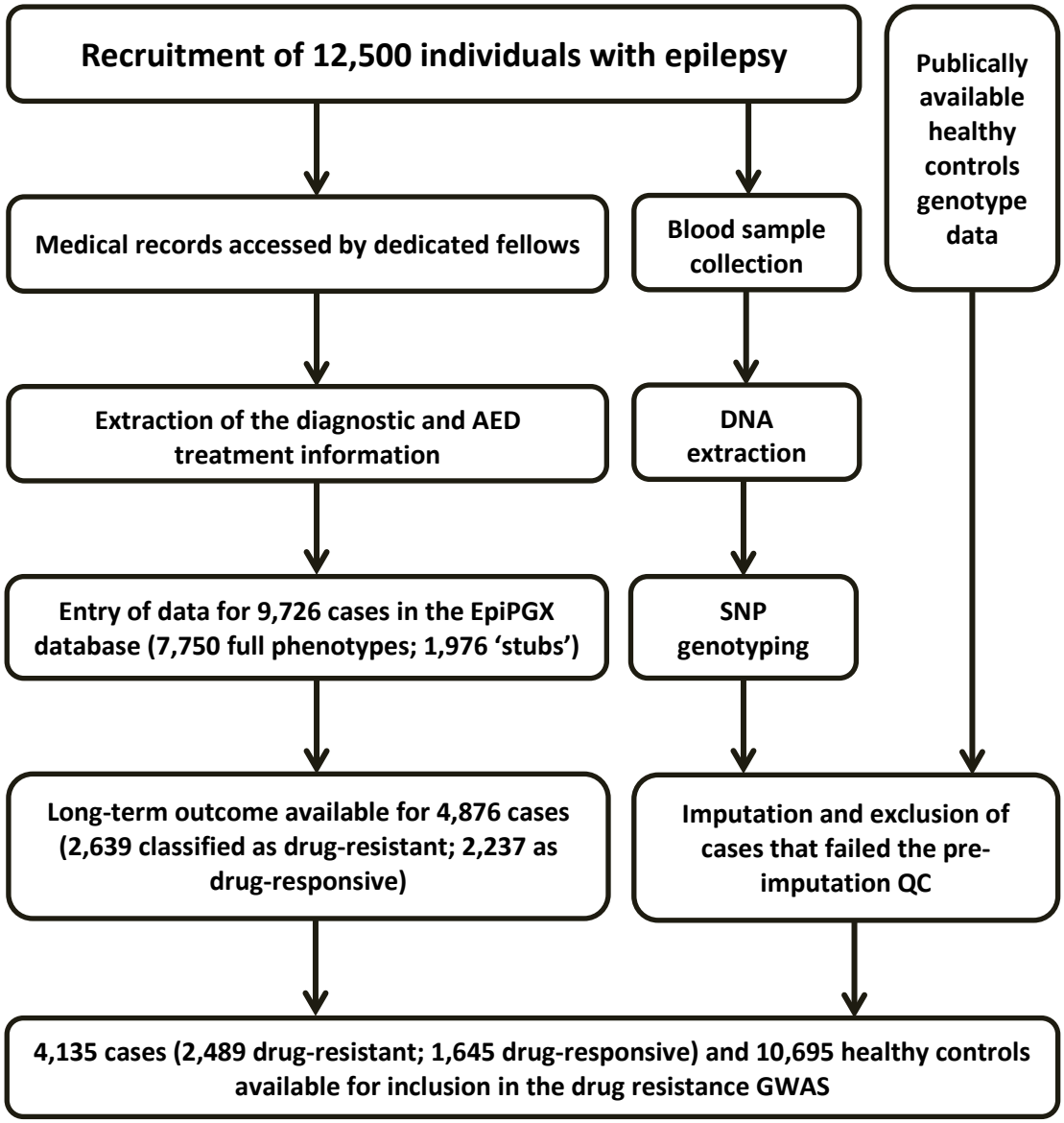
phenotype items were collected for earch set of medical records. Inter-rater agreement was checked for the following phenotype items: epilepsy syndrome diagnosis, seizure types, number of remissions, number of seizures before the first AED treatment, number of seizures in the last 12 months of follow-up, total number of appropriate and adequate AED trials, outcome of individual AED trials, number of failed AED trials, and reported adverse drug reactions to specific AEDs. In addition, we performed two focussed inter-rater agreement exercises to specifically test the classification of cases into drug-resistant, drug-responsive, and those for which the outcome was undefined. The first exercise was performed at the UCL and ULIV sites. Fifty consecutive cases from ULIV were independently classified by a clinical fellow from each site. The second exercise was an internal inter-rater reliability test at UCL. Thirty consecutive cases from each fellow working on the phenotyping task were independently classified as drug-resistant, drug-responsive, or undefined, by another fellow. Inter-rater agreement was expressed as percentage joint-probability agreement, defined as the proportion of the time the raters agree in a nominal or categorical rating system (372).

2.4 Results

2.4.1 Assembly of cohorts for inclusion in drug resistance GWAS

Over 12,500 individuals with epilepsy were recruited across the EpiPGX sites. Healthy controls were obtained from several publically available sources (See Section 2.3.2). As of October 2015, 9,726 out of the 12,500 recruited participants were phenotyped and their data were fully or partially entered in the EpiPGX database. The healthy controls were not entered in the EpiPGX database. Epilepsy outcome information was available for 4,876 epilepsy cases (2,639 were classified as drug-resistant and 2,237 as drug-responsive). After the review of the phenotypic data, SNP genotyping and imputation, 4,135 epilepsy cases (2,489 drug-resistant and 1,645 drug-responsive) were eventually available for inclusion in the drug resistance GWAS, in addition to 10,695 healthy controls. The overview of the cohort assembly for the drug resistance GWAS is shown in Figure 2.1 and individual steps are described in more detail in further text.

Figure 2.1: Cohort assembly overview



2.4.1.1 Accessing medical records and phenotyping

As mentioned above, 9,726 study participants were phenotyped as of October 2015. Medical records were accessed by dedicated clinical fellows and detailed disease and AED treatment information was extracted using the EpiPGX CRF. The data collected

prior to 2013 were entered in local databases. At UCL, basic demographic and diagnostic information was entered in an Excel database, while the AED treatment details were collected in paper format. In early 2013, the dedicated EpiPGX database was ready for use, enabling direct data entry. Data collected earlier were either automatically transferred from the local databases or manually entered in the EpiPGX database.

The phenotyping process proved to be more labour-intensive than anticipated, requiring several dedicated full-time fellows to perform this task. For example, a simple case (two AED trials, no ADRs, no non-medical treatments, pregnancies, or issues such as non-adherence) required the fellow to fill in 60 to 70 fields in the EpiPGX database, with an additional seven fields for the classification into the WP-specific phenotypes. The number of fields that had to be entered increased with every additional AED trial and in case of complications such as ADRs, teratogenicity, or non-medical treatments. Many individuals had had in excess of 20 AED trials and some medical records spanned several decades. For instance, the oldest clinical letter accessed at UCL was dated June 1933. Such complex cases required as many as 500 fields to be entered, taking several hours to complete.

In addition to extracting the relevant information from electronic or physical medical records and data entry, the fellows had to judge whether an individual case met the criteria to fit into any of the relevant pharmacogenomic categories (WP-specific phenotypes; see Appendix 2). In order to decide whether a case was drug-resistant or drug-responsive, the fellows followed an algorithm (see Appendix 3), taking into account the number of seizures in the 12 months prior to the latest recorded visit, the

number of adequate and appropriate AED trials, non-medical treatments, and potentially any issues such as non-adherence or NES. Classification of cases by the person with access to the full medical records ensured all the available information was integrated in the final judgement, including information that would have been difficult to capture electronically (e.g. AED dose changes over time, appropriateness of AEDs for certain epilepsy syndromes).

Since the available resources did not allow for deep phenotyping of the entire EpiPGX cohort, each centre devised a strategy to maximise the number of cases for inclusion in individual GWAS analyses. For instance, when medical records were difficult to access (i.e. not in electronic format), often only basic demographic and disease information was entered. The Consortium agreed to consider a case fully phenotyped if there was complete information on at least one AED trial. If this was not true, the entry was considered a 'stub'. Out the 9,726 cases entered in the EpiPGX database as of October 2015, 7,750 cases were considered fully phenotyped, and the remaining cases were entered as 'stubs'.

2.4.1.2 Classification of epilepsy outcomes for the drug resistance GWAS

The long-term outcome (drug resistance, drug responsiveness) information was available for 4,876 out of the 9,726 cases entered in the EpiPGX database. Of those, 2,639 were classified as drug-resistant and 2,237 as drug-responsive. The outcome was undefined for 5,124 individuals which is a significant proportion of the entire sample. This was most commonly due to insufficient information in the EpiPGX database. Other

common reasons were NES, non-adherence, and unknown number of seizures in the 12 months prior to the latest recorded visit. The reasons why the outcome remained undefined were tracked and documented for the UCL sub-cohort (as outlined in Section 2.4.1.3), but not for the entire EpiPGX sample as this was not part of the agreed data collection.

2.4.1.3 Phenotyping of the UCL sub-cohort

At UCL, 3,131 individuals with epilepsy were recruited from the epilepsy clinics at NHNN and their medical records were accessed by dedicated clinical fellows. Participants for which the diagnosis of epilepsy was not confirmed (64 in total) were excluded from further review, and the remaining 3,067 participants were phenotyped. Prior to 2013, the phenotypic data were collected in paper format, with the exception of the basic demographic and diagnostic information which was entered in an Excel database and later transferred to the EpiPGX database. Data collected from early 2013 onwards were entered directly in the EpiPGX database. Out of the 3,067 phenotyped cases, 1,486 were classified as drug-resistant, 577 as drug-responsive, and for 1,004 the outcome was undefined (see Figure 2.2). The reasons why the outcome was undefined are summarised in Table 2.5.

Figure 2.2: Phenotyping of the UCL sub-cohort – overview

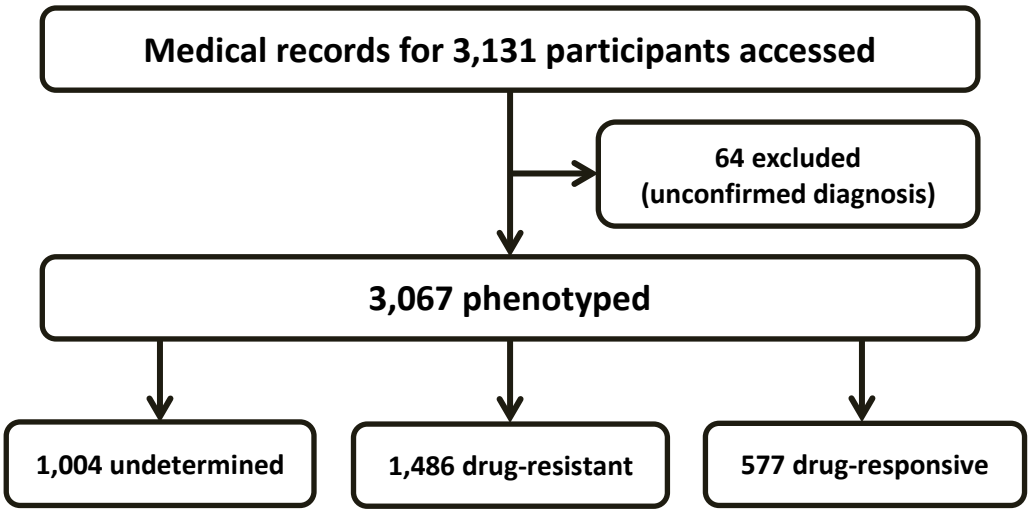


Table 2.5: Reasons why the outcome was undefined in 1,004 cases in the UCL sub-cohort

REASON	NUMBER (PROPORTION)
Non-adherence to medication	44 (4.4%)
History of NES	195 (19.4%)
1-3 seizures in the 12 months prior to the latest recorded visit	150 (14.9%)
Unknown number of seizures in the 12 months prior to the latest recorded visit	221 (22.0%)
Only one or no adequate trials of tolerated and appropriately chosen and used AEDs	51 (5.1%)
Not treated with adequate doses of AEDs in the 12 months prior to the latest recorded visit (while continuing to experience seizures)	74 (7.4%)
Newly diagnosed epilepsy where the seizures stopped after introducing an AED, but not enough time has passed to categorise the outcome as seizure freedom	11 (1.1%)
Insufficient information on epilepsy treatment in the medical records / physical medical records or parts of physical medical records inaccessible	271 (27.0%)
A combination of two or more of the above	13 (1.3%)

2.4.1.4 Inter-rater agreement

Inter-rater agreement was checked across the Consortium for key phenotype items relevant for several WP, as well as at UCL and ULIV for the classification of DRE and drug responsiveness. The cross-centre phenotyping exercise started in May 2013. Trained fellows from seven EpiPGX sites phenotyped ten sets of anonymised notes. Overall agreement for all phenotype items was 74%. Agreement for classification into drug-resistant and drug-responsive cases was 76%. Most of the disagreement resulted from differences in determining whether the outcome was undefined among the

fellows. An additional source of disagreement was differential classification of cases where seizure freedom was achieved as a result of epilepsy surgery. These issues were promptly addressed by the Consortium. Following the cross-centre validation exercise, a phenotyping workshop was organised in October 2013, providing additional training for the clinical fellows. In addition, the EpiPGX database entry manual was compiled to ensure the homogeneity of classifications and data entry.

In December 2013, we performed two inter-rater reliability exercises at the UCL and ULIV sites focussing on the classification of broad AED drug resistance and response. For the inter-rater agreement check between UCL and ULIV, fifty cases consecutively entered in the EpiPGX phenotypic database from ULIV were independently classified by the Author and the clinical fellow from ULIV (PA). Inter-rater agreement for classification of cases into drug-resistant, drug-responsive or undefined was 96%. In addition, we performed an internal inter-rater reliability test at UCL. Four fellows worked on the phenotyping task at UCL (AA, WS, KH, NS). Thirty consecutive cases from each fellow (120 cases in total) were independently classified as drug-resistant, drug-responsive, or undefined, by another fellow. The inter-rater agreement was 93% to 97%. All the discrepancies resulted from differences in understanding or interpretation of the recorded information. Importantly, there were no discordances in classifying the cases as drug-resistant or drug-responsive, but only in establishing the outcome as undefined. This is comparable with the report by Hao *et al.* which assessed the inter-rater reliability of the ILAE definition of DRE. The ILAE definition was applied to 150 individuals with epilepsy followed-up at a single centre by two independent raters. Inter-rater agreement for the classification of epilepsy as drug-resistant, drug-

responsive, or undefined was 94% (373). Even though the comparison with the work by Hao *et al.* is to some extent limited due to different definitions of DRE, it is valuable since the EpiPGX definition was based on the ILAE definition, with the same underlying principles.

2.4.2 Cases available for drug resistance GWAS

After SNP genotyping and imputation, 2,489 drug-resistant cases (out of 2,639), 1,645 drug-responsive cases (out of 2,237), and 10,695 healthy controls were available for inclusion in drug resistance GWAS (see Table 2.6).

Table 2.6: Cohorts available for inclusion in drug resistance GWAS

Country	**DRUG-RESISTANT**	**DRUG-RESPONSIVE**	**HEALTHY CONTROLS** **(publically available)**
UK - London	1,163	450	5,272
UK - Liverpool	388	622	
Ireland	310	100	2,223
Belgium	196	94	1,622
Italy	128	93	261
Germany	281	251	1,317
Netherlands	23	35	0*
Total	**2,489**	**1,645**	**10,695**

*Due to the geographic proximity and similar ancestry, Belgian and Dutch samples are considered to have similar haplotype structure (374, 375), justifying the use of Belgian healthy controls for both cohorts.

2.4.3 Demographic and epilepsy characteristics

Demographic and disease characteristics of individuals with DRE and drug-responsive epilepsy are summarised in Tables 2.7 and 2.8. All parameters are calculated for cases and controls remaining in the analysis after GWAS QC filtering.

Table 2.7: Demographic and disease characteristics of cases included in drug resistance GWAS

Parameter	DRUG-RESISTANT	DRUG-RESPONSIVE	P-value*
Male gender (percentage)	47.5%	49.3%	0.0118
Mean age at epilepsy onset (years)	16.3 (SD 13.8)	25.2 (SD 18.1)	P < 0.0001
Proportion of individuals who never experienced a 12-month remission	16.6%	NA	NA
Mean number of adequate AED trials	4.3 (SD 2.3)	1.9 (SD 1.3)	P < 0.0001

*Double sided t-test was used to compare means and chi-square test to compare proportions.

Table 2.8: Distribution of epilepsy phenotypes in both epilepsy cohorts

Epilepsy type	DRUG-RESISTANT	DRUG-RESPONSIVE
Focal*	2,077 (85.7%)	1,179 (72.5%)
Focal cryptogenic	1,020 (42.1%)	831 (34.3%)
Focal symptomatic	990 (40.9%)	321 (19.7%)
Generalised*	214 (8.8%)	274 (16.9%)
Generalised idiopathic (genetic)	152 (6.3%)	220 (13.5%)
Generalised symptomatic	40 (1.7%)	7 (0.4%)
Unknown whether focal or generalised*	132 (5.4%)	173 (10.6%)

*Broad epilepsy type (focal, generalised, unknown whether focal or generalised) was used as a covariate in GWAS1.

2.5 Discussion

2.5.1 Cohorts for inclusion in GWAS

As of October 2015, the EpiPGX cohort consisted of over 12,000 PWE. In addition, the Consortium had access to 10,695 publically-available healthy controls. Out of the over 12,000 epilepsy cases in the EpiPGX database, less than 50% had sufficiently well characterised outcomes to meet the critera for inclusion in the drug resistance GWAS (2,489 drug-resistant cases and 1,645 drug responders). This was mostly due to the demanding phenotyping procedures and insufficient information in the medical records, or the inability to access full medical records. The information on specific reasons why full medical records were not available was not collected; however, it is likely that this was often due to the fact that various research groups were using the notes for their studies and publications, excluding some of the more complex cases from the sample. Nevertheless, included individuals were representative of the typical population seen at tertiary centres, with a higher representation of drug-resistant individuals and fewer responders.

Individuals with DRE and drug responders differed in terms of gender distribution and some disease characteristics. There was a slightly higher proportion of females in the DRE cohort in comparison with drug responders. Similar trends have also been observed in population-based and specialist centre studies of DRE cohorts (79, 376, 377). The reasons for this difference could include substituting efficacious drugs (ike

valproate for GGE) with less optimal treatments due to family planning (366). Individuals with DRE had on average tried a higher number of AEDs and had a lower age at epilepsy onset. A possible reason for the difference in age at onset seen in the sample could have been recruitment from adult epilepsy clinics at tertiary care centres. Individuals with epilepsy onset in their childhood are more likely to be under follow-up at specialist epilepsy centres as adults if they are drug-resistant than if they are drug responsive. Drug responders are more likely to be either followed up by local neurologists, or no longer under follow-up and thus not accessible to recruitment within EpiPGX, skewing the mean age of the DRE cohort towards a lower value. Younger age at epilepsy onset has been associated with poor prognosis in some epidemiological studies in children (2, 80, 81), with other studies (especially in adults) showing either the opposite or no asssociation (36, 44, 56, 59, 66). Where the association was found, it was most likely driven by different aetiologies encountered in different age groups. Hence, I did not include the age at onset as a covariate in the association analysis (I did include epilepsy type).

Individuals with DRE and drug responders also differed in terms of epilepsy types. Focal symptomatic and focal cryptogenic epilepsy were more common in the drug-resistant group, and generalised idiopathic (genetic) in the drug responders group. As epilepsy type is a consistent predictor of drug response, with symptomatic causes predicting DRE (17, 44, 57-59), I included epilepsy type as a covariate in GWAS1.

Defining the phenotype is of central importance in GWAS of complex traits, yet this step often receives little attention (378, 379). Since the accuracy of phenotyping affects the likelihood of detecting an association, as well as the reproducibility of the

results, it is important to apply uniform phenotypic criteria to define cases and controls. As GWAS studies are often conducted in a multicentre setting, there is the potential for different interpretations of the criteria by individual researchers, leading to heterogeneity of case classification (380).

Significant efforts were made within EpiPGX to ensure homogeneity of classifications across the sites and to obtain good quality data despite the retrospective collection, including the use of consensus definitions, a unified CRF, and evaluation of inter-rater agreement.

Limitations of the phenotypic work were mostly linked to the retrospective data collection. In addition, classifying epilepsy cases into drug-resistant and drug-responsive is associated with other challenges which are much more difficult to overcome. Firstly, in the absence of objective methods, the assessment of seizure frequency relies on patient reports. The diagnosis and classification of epilepsies, seizure types, and drug response rest primarily on clinical observations, which may be subjective and sometimes poorly described in the medical records. Secondly, the epilepsy course may fluctuate between periods of resistance and remission in some individuals, posing an issue in the context of a genetic study.

Advantages and limitations of the EpiPGX phenotyping approach are discussed in more detail below.

2.5.2 Lessons learnt from phenotyping efforts in EpiPGX

2.5.2.1 Consensus definitions

Agreeing on the consensus definitions for long-term epilepsy outcomes (drug resistance, drug responsiveness, and undefined response) was the first step to enable the standardisation of case selection for drug-resistance GWAS.

According to the ILAE definition, one seizure per year is sufficient to classify a case as drug-resistant. However, the DRE definition used for the purpose of this work requires a minimum of four seizures in the 12 months prior to the latest documented visit, resulting in a more stringent drug-resistant phenotype. Extreme phenotypes are often used in GWAS to increase the chance of detecting a signal, although the value of this approach in different disorders can be difficult to predict (340). While the seizure frequency of four seizures over the course of 12 months certainly does not define an extreme phenotype, setting this limit was helpful to exclude individuals with very infrequent seizures for whom the AED response can be difficult to judge, especially in a retrospective setting. This has been illustrated by the ESPERA study which looked at the applicability of the ILAE definition of DRE. Misclassification of outcomes was much more likely in cases with fewer than four seizures per year (381). Individuals with infrequent seizures are more likely to fluctuate between periods of uncontrolled seizures and remissions (382). Consequently, increasing the threshold to four seizures per year is expected to result in the exclusion of some individuals with a fluctuating

course of epilepsy which could potentially be beneficial in the context of a pharmacogenomic study.

To further improve the selection of cases for this research, special care was taken when evaluating individuals who underwent epilepsy surgery. If seizure-freedom was achieved following epilepsy surgery, the individual was still classified as drug-resistant rather than drug-responsive, provided the DRE criteria had been met prior to the surgery. According to the ILAE proposal, these cases would have been classified as drug-responsive; however, this may not reflect the underlying biology as remission is achieved with surgical intervention rather than pharmacotherapy. Since the main goal of this study was to examine the genetic basis of AED resistance, it was important to eliminate the possibility of epilepsy surgery acting as a confounder.

Several measures were taken to make the assessment of adequate AED trials as uniform as possible. The ILAE definition framework does not provide guidance on the adequate doses and durations of treatment as this may vary in individuals due to a range of factors, such as age, weight, liver, and kidney function, as well as any interactions with concomitant medications (373). An individualised approach is needed in clinical practice; however, this may be difficult to translate into a set of rules and criteria for use in research. As a consequence, epilepsy studies rarely define what is considered an adequate trial (382). Even in pharmaceutical trials of novel AEDs where establishing drug resistance is extremely important, criteria for past AED trial failure are typically not defined and the judgement is left to investigators (383-386). In order to standardise the classification of adequate AED trials for the purpose of this study, the EpiPGX Consortium agreed on the appropriate minimum daily doses of individual

AEDs that would likely result in sufficient exposures in an average adult to provide guidance for the fellows who phenotped the cases.

Uncontrolled epilepsy does not always equal DRE (387). Pseudo-resistance due to misdiagnosis, inappropriate AED(s), inadequate AED dose, or non-adherence should be excluded before an individual can be considered drug-resistant (388, 389). Drug response is also very difficult to judge if an individual experiences NES, especially from retrospective medical records. There is considerable variability in the approach to pseudo-resistance across clinical studies in epilepsy. For example, the history of NES is a standard exclusion criterion in pharmaceutical trials (383-386), but this is not always the case in other types of studies. It is not unusual to encounter cohort studies exploring DRE that do not disclose how NES were handled, or whether they were taken into consideration at all (344, 345, 386, 390). Similar heterogeneity is seen with non-compliance and other causes of pseudo-resistance. Major causes of pseudo-resistance were taken into account in this work to improve case selection. However, since the phenotyping was done retrospectively, some causes of pseudo-resistance may not have been identified. This includes individuals accepting suboptimal seizure control due to the lifestyle (provided that they were taking sufficient doses) or women discontinuing valproate due to pregnancy planning and not achieving seizure control with other AEDs.

2.5.2.2 Uniform application of the phenotype definitions

Standardised application of phenotype definitions across the sites was achieved by using a common CRF and a centralised EpiPGX database, as well as inter-rater agreement checks and workshops. There was continuing cross-site communication and discussion of difficult cases.

One of the big learnings in EpiPGX was the extent of diversity of practice across European countries in evaluating epilepsy, from establishing the diagnosis to evaluating AED treatment outcomes. Standardisation of phenotyping in collaborative projects is key to overcome this challenge. Similar to EpiPGX, other international epilepsy genetics consortia (ILAE Consortium on Complex Epilepsies, Epi4k – Gene Discovery in 4,000 Genomes, EPGP – The Epilepsy Phenome/Genome Project) have established efficient phenotyping and informatics cores to standardise the documentation of epilepsy phenotypes across different sites (261, 334, 391). So far, the main focus of these consortia has been the discovery of causal epilepsy genes rather than AED treatment outcomes. As defining and documenting the latter in a consistent manner is more complex than simple disease descriptions, the EpiPGX Consortium has implemented more extensive and rigorous standardisation across a broad spectrum of phenotypes than what had been published in the literature before. For example, checking for inter-rater reliability has not been part of the methodology in the largest recently published epilepsy genetics studies (102, 261, 335, 392). However, this could change soon as several consortia are starting to explore epilepsy treatment outcomes. For example, the Epi4K Project 3 scope of work includes

exploring the effect of all variants identified as risk factors for epilepsy on the prognosis, as well as searching for variants affecting AED response independent of disease susceptibility (334).

2.5.3 Limitations of the EpiPGX phenotyping methods

2.5.3.1 Retrospective data collection

A case-control design with retrospective identification of cases and controls is the most widely used approach in GWAS investigating the genomic basis of complex traits (393), including the present work. Retrospective identification of cases enables a much faster assembly of large cohorts needed for this type of analyses than would be possible with prospective work. Executing a similar drug resistance GWAS using prospective case ascertainment would probably not be feasible in a three-year time frame and it would require significantly larger resources.

While the main advantage of retrospective data collection lies in the ability to identify large numbers of cases, the main disadvantage is potential selection bias. As PWE were recruited from specialist epilepsy clinics, difficult-to-treat cases were over-represented in the overall sample. This could have had a beneficial effect because of the possibility that using a more extreme phenotype may increase the likelihood of detecting a genetic association.

As the outcomes were determined from medical records, the allocation of the AED treatment outcome status relied on the accuracy and completeness of the records. Fellows involved in phenotyping (including myself) frequently encountered difficulties when judging the adequacy and response to past AED trials, especially with older medical records. At UCL (DCEE), there were noticeable quality differences between modern medical records and records from two or more decades ago. The most obvious difference was in the way seizure frequencies were documented, with seizure counts often missing in older letters. Another common issue, especially with paper records, was missing letters and reports. All of these issues hindered the classification of drug response, resulting in AED trials documented as unclassified. As a consequence, the average number of total and adequate AED trials per individual reported here is most likely an underestimate. Furthermore, it is likely that a substantial proportion of the 271 UCL cases with undefined response due to insufficient information on epilepsy treatment would have met the definition of DRE, had that information been available. The main consequence for the GWAS is reduced power.

2.5.3.2 Subjective reporting of seizure frequency

Some issues in epilepsy treatment practice cannot be overcome by careful phenotyping and standardisation of data entry. In the absence of objective markers of seizure frequency, medical judgement of treatment response relies on imperfect reporting of seizure frequency by PWE or carers. Many PWE struggle to maintain accurate records of seizures, especially if they are accompanied by impaired

awareness, or if they occur at night (394, 395). Eyewitness accounts often disagree on important details of seizure presentation (395, 396). According to a study comparing patient records of seizures and video-telemetry results, only approximately 40% of individuals were able to document all seizures accurately. There was low documentation accuracy for complex focal seizures and nocturnal seizures (395).

2.5.3.3 Manual database entry

Manual transcription of data from medical records may result in errors, either due to omission, or wrong value entry. It is well known that structured data entry improves the accuracy of manual documentation (397). Hence, a structured CRF and a unified database were used to improve the overall quality of the data. The EpiPGX electronic database generated alerts if inconsistent data were entered or if no values were entered in the key data fields, prompting the person entering the data to check their consistency. Despite the risk of human error, classification of cases into pre-defined pharmacological categories (phenotypes) by the person entering the data was thought to be more accurate than using computer algorithms. The reason behind this was that the person entering the data had full access to the medical records, including complex information that would have been difficult to capture in the CRF or potential algorithms.

2.5.3.4 Larger than anticipated proportion of cases with undefined outcomes

Another significant limitation of this work was the unexpectedly high proportion of cases for which the long-term treatment outcomes were undefined (52.7% of all cases entered in the EpiPGX database). The most common reason for this was missing information due to the time and resource constraints that did not allow for deep phenotyping of the entire EpiPGX cohort. Other reasons why the long-term outcomes remained undefined in a proportion of cases included unknown number of seizures in the last year of follow-up, NES, and inadequate AED trials in individuals continuing to experience seizures.

Complete information on the reasons for undefined outcomes was only documented for the UCL sub-cohort. The most common reason was insufficient information in the medical records (27% of unclassified cases). Electronic medical records at DCEE were introduced in 2007. As a consequence, physical records had to be accessed to obtain any information recorded prior to 2007. Since physical records are regularly used by several clinical and research groups, inconsistent availability and missing volumes of notes were a recurrent issue. Usually two attempts at requesting individual records were made before the outcome was registered as undefined due to insufficient information. Since DCEE is a tertiary centre, some PWE may only be referred there for a single consultation or for diagnostic assessment, after which they continue their follow-up with local neurologists. In such cases, the lack of follow-up information often made the classification of response impossible.

The second most common reason for undefined outcomes was unknown number of seizures in the last 12 months of follow-up (22%). Since the EpiPGX definition of DRE requires at least four seizures in the 12 months prior to the last recorded visit, quantitative (rather than just qualitative) information was needed to judge whether participants met the criterion. The number of seizures was not possible to assess if the medical records contained vague descriptors of seizure frequency such as: 'a few seizures', 'one cluster', 'unchanged seizure frequency', or when the consultant was unable to judge the nature of episodes. The latter was especially common with sensory auras. The outcome also remained undefined if an individual had only one to three seizures in the past year (approximately 15% of the UCL cohort).

History of NES was another common reason why drug response was difficult to judge (19%). Less common reasons included no or only one adequate AED trial (5.1%) and inadequate doses of current AEDs (7.4%). AED doses were often inadequate if the individual was in the process of titration, or if he or she did not follow the treating neurologist's recommendation to increase the AED dose.

Several published studies using the ILAE definition of DRE reported the proportion of cases with undefined outcome. Since the definition of DRE used in EpiPGX was modified, any comparisons with published data have limited value. In addition, EpiPGX was not an epidemiological study and PWE were not recruited consecutively. Two of the identified published studies using ILAE definition were epidemiological studies assessing the prevalence of DRE (53, 54). In addition, I identified one cross-sectional study (398), one retrospective study (399), and one study of inter-rater reliability of the ILAE classification (373). The latter assessed the inter-rater reliability of outcome

classification in 150 consecutive PWE seen at one centre. The proportion of cases with undefined response was 27.3% to 29.3% (373). The other four studies are summarised in Table 2.9. The proportion of cases with undefined response in these studies ranged from 7 to 37.5%, compared with 32.8% in the UCL sub-cohort in the present study. As expected, the proportion of cases with undefined response was larger in retrospective studies.

Only one of these studies reported the reasons underlying the undefined response category. In 40.2% of participants the response was undefined because they had failed only one AED; 36.8% were receiving inadequate AED doses for various reasons (refusal by individual, titration phase, non-adherence, absence of all required information). For 22.5% it was difficult to determine the outcome either because seizure-freedom was achieved, but the period was insufficient to qualify as remission, or because a new intervention had just been introduced. Missing information was not a major problem in this study, with only 0.5% cases for which the available medical information was insufficient. This is expected considering that only PWE under active follow-up at the time of the study were selected for the review. The entire cohort was under the follow-up of two neurologists, ensuring consistent medical documentation (53).

Table 2.9: Proportion of participants with undefined response in studies using the ILAE definition of DRE

Study	Design	Number of participants	Proportion of participants with undefined response
Ramos-Lizana *et al.*, 2012 (54)	Prospective, tertiary referral centre based study Paediatric population	508	49/508 (9.6%)
Kong *et al.*, 2014 (53)	Retrospective, tertiary referral centre based Adult population	557	209/557 (37.5%)
de Zélicourt *et al.*, 2014 (398) (ESPERA study)	Prospective, multicentre study Adults with focal epilepsy	405	28/405 (7%)
Gomez-Ibañez *et al.*, 2017 (399)	Retrospective, single centre-based Consecutive GGE cases (> 16 years at inclusion)	279	52/279 (18.6%)

ESPERA = European Observational study on PWE.

The ILAE and EpiPGX case definitions are based on the same principles, resulting in similar classification outcomes. Regardless of which definition is used, there is a significant proportion of cases that do not fit the drug resistance or drug responsiveness categories (undefined response). The main reasons for undefined response in the aforementioned publications are similar to my observations in the UCL sub-cohort. As expected, the proportion of cases with undefined response was higher in retrospective studies. However, direct comparison of the proportions of each specific cause may not be appropriate due to the definition differences, as well as different study designs and populations.

The disadvantage of having a high proportion of cases with undefined response was a smaller than expected number of cases suitable for inclusion in the GWAS analysis. On the positive side, exclusion of cases with less than clear outcomes is beneficial in the

context of a pharmacogenomic study. Exclusion of cases with very infrequent seizures maximises the chance to identify genomic markers of clinically meaningful drug resistance.

2.5.3.5 Fluctuating course of epilepsy – consequences for genomic studies

As mentioned in Section 1.2.1, one of the disadvantages of the ILAE definition and any ILAE-based definition is that it is only valid at the time of assessment. Even though a significant proportion of individuals respond early and remain in sustained remission, while others remain resistant, several recent reports suggest that the natural course of epilepsy is much more complex than expected and that there is often a dynamic relationship between drug resistance and remission. A proportion of patients have a fluctuating course, with periods of seizure freedom and relapses (400, 401). Associating unstable phenotypes with constant genotypes represents a challenge that is difficult to overcome. In a recent publication, Neligan *et al.* estimated that approximately 10% of the epilepsy population follows a fluctuating course (401). In the study by Brodie *et al.*, following participants for a median of 7.5 years, 16% had a fluctuating course (28). Several studies have documented (often transient) remissions of ≥ 12 months in patients who had met various criteria for DRE (79, 376, 377, 382, 402-406). In the study by Callaghan *et al.*, 246 individuals with DRE were followed for a median of 3.1 years. They estimated the probability to attain a 12-month remission as 5% per year. This percentage was only slightly smaller when only medical treatments were taken into account. Terminal remissions were mostly observed within the first

ten years after diagnosis; however, a small proportion of individuals entered a remission as late as 30 to 35 years into the disease course (403). In a retrospective study of 187 adults with DRE with a median follow-up of seven years, the estimated annual probability of achieving a ≥ 12 months remission was approximately 4% (404). Neligan *et al.* found that a substantial proportion of PWE who initially achieved remission subsequently relapsed when followed up for an extended period (406). In the study by Berg *et al.*, 68% of the individuals who achieved a ≥ 12 months remission relapsed by the end of follow-up (382). Repeated remissions and relapses were common. In the study by Callaghan *et al.*, there was a 70% risk of relapse after a ≥ 12 months remission at five years of follow-up (79).

The published literature suggests that remissions in DRE are often not enduring. In the study by Berg *et al.*, participants experiencing infrequent seizures had a higher chance to experience remissions, but they were also at a higher risk of a relapse afterwards. The authors speculated that in these PWE AEDs might have been partially efficacious, "reducing their break-through seizure rate to such a low point that they experienced 1-year seizure-free periods" (382). Assessing AED response in individuals with very infrequent seizures is challenging. Potentially, these PWE would need to be followed for several years to determine the outcome with confidence (26, 382).

The issue of fluctuating epilepsy course is not addressed by the ILAE framework or any other definition of DRE. All definitions used in research are only valid at the time when they are applied. It is probable that some of the individuals in our drug-resistant cohort would have achieved remission and that some of the drug-responsive individuals would have relapsed if they were assessed at a different time point. A solution to this

would be defining an extreme drug resistance phenotype for the purpose of whole-genome studies. It has been demonstrated that the probability of achieving remission gradually declines with the number of failed AED trials. After five to six unsuccessful trials, the probability of a remission is close to zero (49, 78, 79). Thus, one possible option how to define extreme drug resistance would be as failure of ever achieving a 12-month period of seizure-freedom despite treatment with at least five appropriate and adequate AED trials. Consensus would be needed on the minimum required follow-up duration in this context. From my experience with phenotyping, a very small proportion of cases would actually fulfill these criteria because remissions are relatively common. Retrospective data collection represents an issue, and potentially prospective collection would be beter. In our cohort, 421 PWE (10.2% of all epilepsy cases) had no remissions and only 185 (4.5%) had at least five documented appropriate and adequate AED trials in addition to no remissions. I have identified six studies reporting the proportion of PWE who never achieved a 12-month remission. The proportion was relatively small, ranging from 4% to 35.4% (Table 2.10).

Table 2.10: Proportion of participants with no 12-month remissions in cohort studies

Study	Number of participants Study design Duration of FU	Proportion of participants who never achieved a 12-month remission
Brodie *et al.*, 2012 (28)	N = 1,098 Paediatric and adult Prospective Median FU 7.5 years (IQR 4.7-12)	25%
Cockerell *et al.*, 1997 (41)	N = 564 Paediatric and adult Prospective Median FU 7.1 years (CI 5.7, 8.1)	4%
Geerts *et al.*, 2010 (29)	N = 413 Paediatric cohort Prospective Median FU 14.8 (range 11.6-17.5)	8.5% had no remissions exceeding 3 months
Lindsten *et al.*, 2001 (48)	N = 107 Adult cohort Prospective Up to 12 years FU (80% of the cohort had ≥ 5 years FU, 35% ≥ 10 years FU)	32%
Mohanraj and Brodie, 2006 (43)	N = 780 Adult cohort Median FU 79 months (range 24-252)	35.4%
Zhang *et al.*, 2013 (49)	N = 180 Paediatric and adult Prospective Median FU 5 years (range 2-10 years)	20% (13% tried ≥ 2 AEDs and 7% only tried 1 AED and could therefore not be classified as drug-resistant)

FU = follow-up; N = number of individuals in a cohort.

Exploring extreme drug resistance phenotypes in pharmacogenomic studies would be possible; however, assembling cohorts of sufficient size would represent a significant challenge, requiring collaborative efforts. Prospective data collection would provide higher phenotypic data quality.

2.6 Conclusions

Appropriate selection of cases and controls is the foundation of any pharmacogenomic research and a key factor influencing the outcomes. Considering the large sample sizes required for GWAS analyses, phenotyping is a major endeavour in this type of research, requiring substantial resources and time.

The phenotyping efforts within EpiPGX provided sufficiently large and well-characterised cohorts of individuals with DRE and drug responders to conduct drug resistance GWAS analyses, although the phenotyping methods had several limitations. Phenotyping long-term AED treatment outcomes proved to be significantly more challenging than anticipated due to the requirement to integrate complex information to assign the final phenotypic categories. Challenges were encountered on both levels: when judging individual AED outcomes, as well as long-term outcomes (resistance vs. response). The long-term outcomes remained undefined for a significant proportion of cases, mainly due to missing data in the medical records and the inability to deeply phenotype all cases.

Agreeing on EpiPGX consensus definitions of long-term outcomes was equally challenging and the Consortium leaders considered a range of possibilities before settling on the final definitions. Considering the complexity of AED response, it is not surprising that simplifying it to a dichotomous variable suitable for a case-control GWAS approach was challenging. Using quantitative rather than dichotomous outcomes may be worth exploring in the future.

The published literature provided little guidance on how to approach these issues in multicentric studies. The literature review I performed showed that the phenotyping methods in pharmacogenomics publications are usually not described in detail, so it is often unclear how the phenotypic categories were defined, how non-pharmacological interventions were taken into account, whether common causes of pseudoresistance were excluded, and how the adequacy and appropriateness of AED trials were judged. This could reflect the fact that implementing strict guidelines is difficult and potentially so restrictive it excludes high numbers of cases (as was the case in the present work), so authors may consider it more appropriate to rely on clinical judgement. Some studies mention AED levels, but these are not relevant for all AEDs (407). Considering the cost of pharmacogenomic studies and the impact of case selection on the quality of the results, phenotyping should receive more attention. A consensus on how to assess the adequacy of AED trials across research studies in epilepsy would improve transparency and allow for easy comparisons of studies. Further efforts to achieve consensus on how to define an adequate AED trial, considering multiple factors simultaneously, are needed. A better understanding of the relationship between the AED dosage and treatment duration on the clinical outcome would be helpful (389). According to the ILAE proposal, adequate means application of the AED "at adequate dosage for a sufficient length of time" (26). Defining the adequate clinical dose range is challenging, considering the significant inter-individual differences in the doses required to achieve seizure freedom (408). Some of the inter-individual variation can be attributed to factors such as age, the presence of hepatic or renal impairment, as well as the treatment regimen (as monotherapy or combination). For adults, the WHO's DDD guideline represents a helpful reference, defining the average

"maintenance dose per day for a drug used for its main indication" (409). In his commentary on the ILAE definition, Kwan argues that more work is needed and that referring to systems like the WHO DDD could be a way forward, but points out that this only applies to monotherapy (74). Some authors emphasise the importance of a documented attempt to titrate the AED to reach an adequate dose, especially when gradual titration is needed to achieve good tolerability (410).

The approach used in EpiPGX proved to be pragmatic and feasible to apply across multiple sites. Similar approaches could be considered in epilepsy research performed by multicentre consortia. Several improvements could be made in future studies, such as using community-based cohorts to access drug responders, and prospective design.

CHAPTER 3: GWAS

3.1 Introduction

I performed two GWAS analyses to explore the influence of common genomic variation on drug resistance in epilepsy. In the introductory part of this chapter, I outline the principles of GWAS design and explain the theoretical background of each step in a GWAS. The GWAS protocol and settings I used for the purpose of this work are outlined later on, in Section 3.3 (Methods). Lastly, I report and discuss the results of both GWAS analyses.

3.1.1 The principles of GWAS design

As explained in Section 1.7.4.3, GWAS examine a genome-wide set of variants to identify associations with diseases or traits and are especially suitable to detect associations of variants that are common in a population. A GWAS design includes the following steps: SNP genotyping, imputation, marker and sample QC, association analysis, and data interpretation.

3.1.1.1 SNP genotyping

The first step in GWAS is SNP genotyping using chip-based microarrays to determine the alleles of, typically, 500,000 to over 1,000,000 SNPs as a surrogate of the total genomic variation. SNPs represented on commercial platforms are selected using the following strategies: tagging SNPs (representative of LD blocks) or SNPs equally distributed through the genome (dispersed approach), or both, to ensure maximally informative sets (199, 200). The two most commonly used platforms are Illumina and Affymetrix. Both types of microarrays contain multiple oligonucleotide sequences flanking selected SNPs that are differentially labelled according to the SNP allele (200). The Affymetrix platform contains short DNA sequences to detect specific SNP alleles as spots on a quartz chip (microarray). In contrast, Illumina uses slightly longer DNA sequences recognising different alleles and a bead-based technology (170). When the sample DNA is hybridised to a microarray, it differentially binds to the probes corresponding to the specific SNP alleles present in that sample. The assay produces two-colour readouts (depending on the SNP allele) for each genomic locus. Intensity values of the two colours convey information about the ratio between the two alleles at that locus. Usually, a large number of samples are genotyped in a single study, allowing normalisation of the intensity values across all samples. When these values are appropriately normalised and plotted, samples with identical genotypes at an individual genomic locus aggregate in a cluster. If two different alleles (A and B) exist at a locus, the samples are expected to exhibit three separate clusters (AA, AB, and BB). Genotype calls are made by comparing this information with information from a standard cluster file derived from a representative sample set with known genotypes,

allowing genotype calls in the studied samples. Implementing an efficient and robust clustering algorithm is essential for accurate genotyping (411). SNP genotyping using commercial microarray platforms (Affimetrix, Illumina) is expected to have greater than 99% accuracy. Reproducibility, expressed as genotype concordance between replicate studies of the same sample, has been shown to range from 99.4 to 99.9% for the same laboratory and platform (412, 413).

3.1.1.2 Imputation

Imputation is the process of statistical prediction of genotypes not directly genotyped in a sample using the haplotype structure of an external high-density reference (414).

The purposes of imputation are:

1. Increased power

   By predicting the un-typed SNPs, the number of SNPs available for association testing increases substantially, therefore increasing the power.

2. Fine mapping

   Imputation provides a high-resolution overview of an association signal across a locus.

3. Combining data genotyped with different arrays

   GWAS data sets genotyped with different arrays (common when working with several batches, or across consortia) can be imputed to the same reference

panel to provide genotypes for a common set of SNPs across all experiments, enabling all results to be combined into joint analyses (414).

Commonly used reference panels are the HapMap and 1000 Genomes references. The most recent 1000 Genomes release includes genomes of 2,504 individuals from 26 populations genotyped for over 88 million variants, including short indels and structural variants (163, 414). To facilitate haplotype matching, the reference panel should include haplotypes from the same population as the reference panel. Several types of software are used for genotype imputation, the most frequently used being IMPUTE (415), MaCH (416, 417), and Beagle (418). Imputation involves several computational steps. As the quality of imputation depends on the study and reference data being on the same physical strand of DNA, aligning the strands is an important initial step. This is followed by haplotype estimation, also referred to as haplotype phasing. Haplotypes are estimated at SNPs genotyped (observed) in the study sample. Next, the phased haplotypes are modelled as a mosaic of those in the haplotype reference panel. As the reference haplotypes contain the genotypes of SNPs that were not genotyped in the study sample, this allows the inference of unobserved genotypes. As the study sample haplotypes may map to multiple reference haplotypes, the imputed genotypes are assigned a probability score based on the haplotype overlap. This score can be utilised in GWAS to account for the the uncertainty of imputation. For each individual, imputation provides a probability distribution of possible genotypes at each un-typed variant from the reference panel (414, 415, 419, 420).

3.1.1.3 Quality control (QC) procedures in GWAS

GWAS aim to correlate the phenotype differences with differences in SNP allele frequencies. The commonly used case-control approach assumes that any detected allele frequency differences actually relate to the studied phenotype, i.e. there are no unobserved confounding effects, either directly attributable to the causal marker or through another marker that is located nearby (157).

The potential to identify true genomic associations in a GWAS depends on the overall data quality. Association tests are compromised if the genome-wide SNP data have not passed appropriate QC, potentially leading to excessive type I or type II errors. Sample identity issues caused by sample handling errors can be detected by comparing genetic data with clinical and self-reported data (e.g. gender or relatedness). Confounding and false-positive associations in GWAS can also be a consequence of population stratification, relatedness, and batch effects (421).

3.1.1.3.1 Marker QC procedures

Marker QC procedures include removing SNPs with excessive missing genotype data, SNPs with a significant deviation from Hardy-Weinberg equilibrium (HWE), SNPs with significantly different missing genotype rates between cases and controls, and SNPs with a very low MAF (422). In analyses that include imputed data, SNPs with poor imputation quality measures are also excluded (414).

i) *Exclusion of SNPs with high missing genotype rates*

Poorly characterised markers need to be filtered out from GWAS as problems with missingness or different missingness rates in cases and controls may lead to spurious associations. SNPs with excessive missingness rates or significantly different proportions of missing genotypes in cases and controls are excluded. Different thresholds are used for this purpose. The most commonly used standard is to exclude SNPs with more than 5% missing data across the sample from further analyses. In addition, SNPs should be checked for significant differences in missingness rates between cases and controls (195, 422, 423).

ii) *Excluding SNPs with a significant deviation from Hardy-Weinberg equilibrium (HWE)*

Most human SNPs satisfy the HWE law according to which states that both the allele and genotype frequencies in a population stay constant over generations. A SNP satisfies HWE if the ratios of genotype frequencies in the population are in the right proportion. This can be tested using a chi-square goodness of fit test or the HWE exact test (424, 425). Possible reasons for a departure from HWE are analytical mistakes, natural selection, population admixture, inbreeding, and segmental duplications. Usually, SNPs that significantly deviate from HWE are excluded from downstream association analyses because this may be the result of genotyping or genotype calling errors (195, 422). On the other hand, deviations from HWE can be the consequence of selection associated with

the disease process, so they can be seen in case samples at loci associated with the disease and removing these SNPs from further analyses could be counter-productive. Thus, HWE deviations are only tested for in controls. It is common to exclude SNPs with the HWE exact test significance values (P_{HWE}) below 10^{-4} to 10^{-7} (423, 426).

iii) *Excluding SNPs with low MAF*

If a SNP has a low MAF, this will result in low genotype counts for at least one of the three possible genotypes in the study population, potentially confounding the calling process. Normally, SNPs with low MAF are excluded from downstream analyses as they are more likely to be affected by genotyping errors and are generally less informative in GWAS (423). Furthermore, sample sizes would have to be extremely large to identify associations of variants with very low MAF. Commonly used thresholds to exclude SNPs with low MAF are 0.01, 0.02, or 0.05, depending on the cohort size (195, 423, 427, 428).

iv) *Excluding imputed SNPs with poor quality metrics*

Imputed SNPs are statistical predictions, not actual observations as the genotyped SNPs. The certainty of imputation for each imputed SNP is described by several imputation quality metrics. These metrics are specific to the software used. Here I am focussing on IMPUTE (imputation software used in the present work – specifically version 2.3.0) and SNPTEST (association analysis software used for marker QC in the present work). The

main imputation quality metrics given by IMPUTE and SNPTEST include the *info* score, *average_maximum_posterior_call*, *concord_type0*, and *r2_type0*. IMPUTE assigns an internal "type" to each SNP. Imputed SNPs are referred to as "Type 0" and SNPs genotyped in the study data set are referred to as "Type 2" (415, 419).

The *info* score is a statistical information metrics capturing the certainty of imputation. It is based on the population allele frequency and it is calculated as a function of the observed statistical information on the allele dosage in the imputed sample and the expected allele dosage (429, 430). Values range from 0 to 1, with lower values signifying greater uncertainty in the imputed genotypes and values close to 1 indicating a high certainty. The genotyped SNPs have *info* values of approximately 1. According to the IMPUTE manual, negative *info* scores may be encountered "when the imputation is very uncertain", and the value -1 means the metric could not be calculated (419). The *info* metric is often used to exclude poorly imputed SNPs from downstream association analyses; however, there is no strong consensus on the exact cutoff value (431). There is precedence for taking forward only SNPs with *info* scores of at least 0.3 to 0.9 (428, 431-433).

IMPUTE reports the posterior probabilities of genotype calls. Posterior probability is the likelihood that an individual carries a SNP genotype, taking into account all the background (prior) and new (generated during the imputation) evidence. The values range from 0 to 1. The posterior

probability of observed (measured) genotypes is 1 as there is no uncertainty due to imputation (430). The *average_maximum_posterior_call* is an imputation confidence score generated by SNPTEST. It is the average of the maximum posterior probabilities of the imputed genotypes for a SNP across all samples in a study (434). As with the *info* metrics, there is no default threshold for excluding SNPs on the basis of *average_maximum_posterior_call* metrics. There is precedence for using 0.90 as the cutoff (435).

Imputation accuracy (performance) can be assessed using the r-squared correlation (IMPUTE *r2_type0*) and concordance rate (IMPUTE *concord_type0*) metrics. Both metrics are derived by masking a proportion of observed genotypes and treating them as imputed (436). The *r2_type0* metric is defined as the squared correlation between observed and masked genotypes at a SNP. Values between 0 and 1 can be encountered, with higher values indicating better imputation accuracy. There is no agreed threshold for filtering SNPs on the basis of the *r2_type0* metrics; however, cutoff values between 0.3 to 0.9 have been used (428, 437). *The concord_type0* metric is defined as the concordance between the observed and masked/imputed genotypes at a SNP. Values between 0 and 1 can be encountered, with higher values indicating better imputation accuracy. There is no agreed threshold for filtering SNPs on the basis of the *concord_type0* metrics; however, cutoff values between 0.7 and 0.9 have

been used (438, 439).

3.1.1.3.2 Sample QC procedures

Sample QC procedures include removing samples with potential identity problems, samples with excessive missing genotype data, heterozygosity, population stratification, or relatedness issues.

i) *Checking for potential sample identity problems*

Comparing genetic and clinical data can potentially uncover sample identity problems resulting from sample handling errors. Comparing the reported gender for each sample with the genotype data (chromosome X heterozygosity) can identify discrepancies, for example an individual's reported gender is female, but genetic data show homozygosity for all chromosome X markers. In this case, the study record forms and medical records should be revisited to determine whether there has been a sample handling error. Another way to check sample identity is by comparing genotype data with self-reported relatedness between individuals. This is done by estimating the degree of shared genetic ancestry between individuals, i.e. computing kinship estimates for every pair of individuals in

the study (421). This is described in more detail below.

ii) *Sample relatedness*

In population-based GWAS using case-control design, all the individuals should be unrelated. This means that the relatedness between any pair of samples should not exceed the relatedness between second-degree relatives. Inclusion of first-or second-degree relatives in the sample might result in bias as the genotypes shared by the related individuals will be over-represented, and thus the allele frequencies in the population may no longer be reflected in the sample (422). Duplicate and related samples should therefore be removed from the analysis. Using dense SNP genotype data (excluding sex chromosomes), it is possible to compute IBS (identity by state), a metric based on the average proportion of alleles in common at genotyped SNPs for each pair of individuals. Independent SNPs (i.e. those not in LD) are used for this method. Alleles at each locus are compared between the two individuals. If two different alleles are possible (A and B), there are three options: the two individuals have both alleles in common (e.g. AA and AA; BB and BB; AB and AB), one allele in common (e.g. AA and AB; AB and BB), or no alleles in common (e.g. AA and BB). The IBS data are then used to estimate IBD (identitiy by descent), the degree of recent shared ancestry for each pair of individuals (421, 426). Two alleles are IBD if they originate from the same ancestral allele. Theoretical values of IBD across all alleles in a study are 1 for monozygotic twins or duplicated samples, 0.5 for first-degree relatives, 0.25 for second-degree relatives, and

0.125 for third-degree relatives. Unrelated individuals share zero alleles IBD at each locus (421, 422). There is often some variability around these theoretical values, arising from genotyping errors, LD, and population structure. In GWAS analyses it is common to remove one out each pair of individuals with an IBD > 0.1875, which is midway between the expected IBD for third- and second-degree relatives. Similarly, an IBD > 0.98 indicates monozygotic twins or duplicated samples (422).

iii) *Exclusion of samples with high or low proportions of heterozygous SNPs*

Heterozygosity is the proportion of heterozygous genotypes across all autosomal SNPs of the sample (423). Excess heterozygosity may point to sample contamination and low heterozygosity may be an indicator of hybridisation problems or inbreeding. High heterozygosity outliers (more than 3 or 4 standard deviations from the sample mean) are usually excluded from the association analysis (195, 422, 423).

iv) *Exclusion of samples with high missing genotype rates*

The missing genotype rate is defined as the fraction of missing calls over the total number of SNPs per sample. A high proportion of missing data may indicate hybridisation problems, which can be caused by microarray malfunction or poor DNA quality (423, 440). It is common to exclude samples with > 0.05 missing call rates from the association analyses (195, 440).

v) *Population stratification*

The genetic composition (i.e. allele frequencies) varies within and between populations, irrespective of disease status. Population stratification refers to the presence of a systematic difference in allele frequencies due to the diverse ancestry of cases and controls (157). Population stratification is one of the main sources of confounding in GWAS that explore the association of SNP alleles with a trait of interest (422). In epidemiology studies, a confounder is defined as a factor that is associated with both the exposure variable and the trait, but is not a result of the exposure variable. In GWAS analyses, the exposure variables are SNP alleles. Confounding occurs when cases and controls have different allele frequencies attributable to different population substructure. In this situation, an association signal may arise because of different allele frequencies between the founder populations that are differentially represented in the case and control groups and not because of an association with the risk for the trait under investigation. Population stratification is probably the most frequently cited reason for non-replication of GWAS results (157). Since large sample sizes are required to identify common variants carrying risk for common traits, even a low degree of population stratification can negatively affect GWAS results (194).

There are various approaches to control for population stratification in GWAS. When selecting samples for GWAS, careful consideration should be given to matching the cases and controls according to ancestry, although

this is not always sufficient (441). The effect of population stratification can also be reduced by removing the population outliers (422).

The most commonly used methods for identifying and removing samples with divergent ancestry are principal component analysis (PCA) and multidimensional scaling (MDS). The relatedness or genetic distance between individuals in the sample is determined by measuring IBS defined as the genome-wide proportion of shared alleles (426). Differences in IBS between pairs of individuals can result from relatedness or from population-specific genotype frequencies (442). IBS is used to create a similarity matrix for PCA or MDS. PCA and MDS are multivariate statistics methods in which each individual is represented by a vector or point in an X-dimensional matrix and the pairwise inter-individual similarity is indicated by the distance between two points. The genetic background of an individual is expressed as a vector of coordinates reflecting the pairwise inter-individual genetic correlations (443, 444). The first vector represents the axis accounting for the greatest possible amount of genetic variation in the data; the second vector represents the axis accounting for the second greatest amount of genetic variation, etc. (444).

These vectors are commonly used as covariates in the association analysis to account for population stratification in GWAS, a method established by The Wellcome Trust Case Control Consortium (195).

The PCA or MDS output is commonly visualised on a plot where the first two PCA (MDS) components (coordinates) for each sample are represented on the X and Y axes. Any outlying samples may be removed at this point on the basis of visual inspection (413).

Another method of checking for population stratification in GWAS is to examine quantile-quantile (QQ) plots. In these plots, the observed test statistics (i.e. P-values generated for every SNP) in a GWAS are ranked from the lowest to the highest and plotted against a theoretical (such as the chi-square) distribution of values expected under the null hypothesis, i.e. if there were no SNPs associated with the trait. The plots are expected to show a diagonal line, also referred to as the identity diagonal line. Since the vast majority of SNPs tested in a GWAS are not expected to be associated with the trait, strong deviations from the diagonal line may indicate undetected sample duplications or familial relationships, population stratification, or systematic technical bias (191). The deviation (inflation) of observed statistics from the expected distribution is expressed as the genomic inflation factor (λ), defined as the ratio of the median of observed chi-square test statistics and the median of chi-square expected under the null distribution (431). Genomic control is a method that can be used to correct for this inflation. When implementing genomic control, the observed test statistics is divided by λ. There is no universally agreed threshold for the λ cutoff value. Values below 1.1 have been considered acceptable in large GWAS (195).

3.1.2 Association analysis

The most common approach to association analysis is comparison of two groups of samples: a case group (i.e. the group with a disease/trait under investigation) and a control group (i.e. the group without the disease/trait). For each SNP genotyped in both groups, a test is conducted to investigate if the allele frequency is significantly different between the case group and the control group. The effect size is usually reported as odds ratio (OR) of two odds: the odds of disease for carriers of a specific allele, and the odds of disease for non-carriers. In addition, P-value indicating the significance of the OR is reported. The commonly used statistical methods to examine dichotomous phenotypes (case-control) are the chi-square test, Fisher's exact test, and logistic regression. In logistic regression, the outcome of a linear regression model is transformed using a logistic function predicting the likelihood of having the disease phenotype given the genotype status. Often, logistic regression is the selected method because it allows the inclusion of covariates and can generate adjusted OR as a measure of effect size (170).

The genotype data used in case-control association analyses can be encoded in different ways which can influence the statistical power of a test. In allelic association tests a SNP allele is tested against the phenotype. In contrast, genotypic association tests use genotypes or genotype categories. In genotype association testing,

assumptions may be made about the underlying genetic model of inheritance, such as dominant, recessive, additive, or multiplicative (170, 445).

The bioinformatics software packages commonly used for association analyses are SNPTEST (415) and PLINK (426). Linear mixed models, such as Factored Spectrally Transformed Linear Mixed Models (FaST-LMM) algorithm, can also be used. Mixed model in this context is equivalent to a form of linear regression in which the genetic similarity between individuals in the study is used as a covariate. The genetic similarity between every pair of individuals in the study is derived from IBS (explained above) and captured in the genetic similarity matrix. The advantage of FaST-LMM is the correction of confounding by population stratification, cryptic relatedness, or close familial relationships (446).

3.1.3 Correction for multiple testing

It is important to consider multiple comparisons when determining statistical significance in GWAS. Since it is common to test more than a million SNPs in a single GWAS, the probability of Type I errors (false positive associations) with a statistical significance level typical of small studies (e.g. $\alpha = 0.005$ or 0.001) is high. The most commonly used solution in GWAS is applying the Bonferroni correction, which means dividing the conventional P-value threshold by the number of statistical tests performed in parallel. This approach is conservative and applicable only if the tested SNPs are independent. For SNPs located in regions with strong LD (LD blocks) that are not truly independent, it would result in over-correction, increasing the probability of

Type II errors (false negative results). Since the human genome in Europeans and North Americans has approximately one million independent chromosomal regions, the corrected thresholds for an assumed million independent SNPs are $\alpha < 5*10^{-8}$ (corresponding to $\alpha < 0.05$) or $1*10^{-8}$ (corresponding to $\alpha < 0.01$). Using these corrected thresholds may result in substantial Type II error rates if sample sizes are small, for example in studies of rare disorders (447).

3.1.4 Interpreting results

Association studies using case-control design can produce valid results, provided that good epidemiologic design practice is applied (191). However, interpretation of GWAS results can be challenging due to the limited resolution of the SNP genotyping arrays. The allele driving the association (i.e. index SNP) is not necessarily the functionally relevant allele. Often, other SNPs in high LD with the genome-wide associated variant have functional implications for the phenotype. Thus, an associated SNP may point to the functional allele in the region (haplotype block), or it may be used as a genomic biomarker for the trait (200). Additional work is usually required to bridge the gap between finding an associated SNP and understanding how a locus contributes to a trait. This work may combine fine-mapping of association signals and functional annotation of variants. Fine-mapping refers to dense genotyping, imputation, or sequencing of the loci of interest to identify a larger number or all variants in the candidate region, followed by an association analysis to identify the specific variant responsible for the association or secondary association signals at a locus. Functional

annotation refers to identifying SNPs within regulatory elements of the DNA or effects of SNP alleles on gene expression (448).

More than 90% of the SNPs found to be associated with diseases in GWAS are not located in protein-coding regions, indicating that they may increase the risk by altering regulatory elements of DNA elements resulting in gene expression changes (203, 449, 450). Variation in gene expression is an important contributor to disease risk and it has been shown to be highly heritable (449). Non-coding variants can influence gene expression in several ways. Variants within promoters or enhancers can influence transcription (451, 452). Intronic variants can potentially affect alternative splicing and mRNA stability. Furthermore, variants may affect the expression or function of non-coding RNAs (453). Identifying the genes affected by non-coding variants often requires considerable effort. Attributing the effect of non-coding variants identified by GWAS to the nearest gene, while common practice, may not necessarily be scientifically accurate (454, 455). In certain cases, the answer is provided by complementary methods, such as eQTL and tissue-specific expression patterns of local genes (456, 457). An eQTL is a locus that influences the expression level of a gene. Such a locus can be located close to the gene (cis-eQTL), or far away (trans-eQTL). Standard eQTL analysis combines a genome-wide analysis of transcript levels in the tissue(s) of interest with GWAS. Interpretation of eQTL data can be enriched by incorporating information on epigenetic modifications and analysis of regulatory networks (449).

As explained above, genomic variants found to be associated with traits of interest in GWAS are unlikely to be the functional variants (203). SNPs are markers of regions and

functional variants driving the association may be several thousands of base pairs away from the associated SNP(s). If the association of the genotyped SNP with a phenotype is shown to be biologically plausible, the phenotypic association is considered direct. If a tagging SNP that is in strong LD with the functionally significant SNP shows statistically significant assoiation with the phenotype, the association is indirect (170). Follow-up studies are important to explore all variants in complete or near-complete LD with the tagging SNP to identify potentially functional variants (170, 458).

3.2 Aim

The aim of the present GWAS work was to identify common genomic variants associated with drug resistance in epilepsy of any aetiology by studying individuals with DRE, drug responders, and healthy controls. As outlined in Section 1.10, I tested two hypotheses using GWAS:

Hypothesis 1:

Common genomic variants are associated with drug resistance in common epilepsies.

To test Hypothesis 1, I performed a GWAS analysis comparing individuals with DRE and individuals with drug-responsive epilepsy (GWAS1).

Hypothesis 2:

Identified variants are determinants of drug response rather than epilepsy susceptibility variants.

To test Hypothesis 2, I performed an additional GWAS analysis comparing individuals with DRE and healthy individuals (GWAS2). If Hypothesis 2 is accurate, the loci identified in GWAS1 will not be replicated in GWAS2.

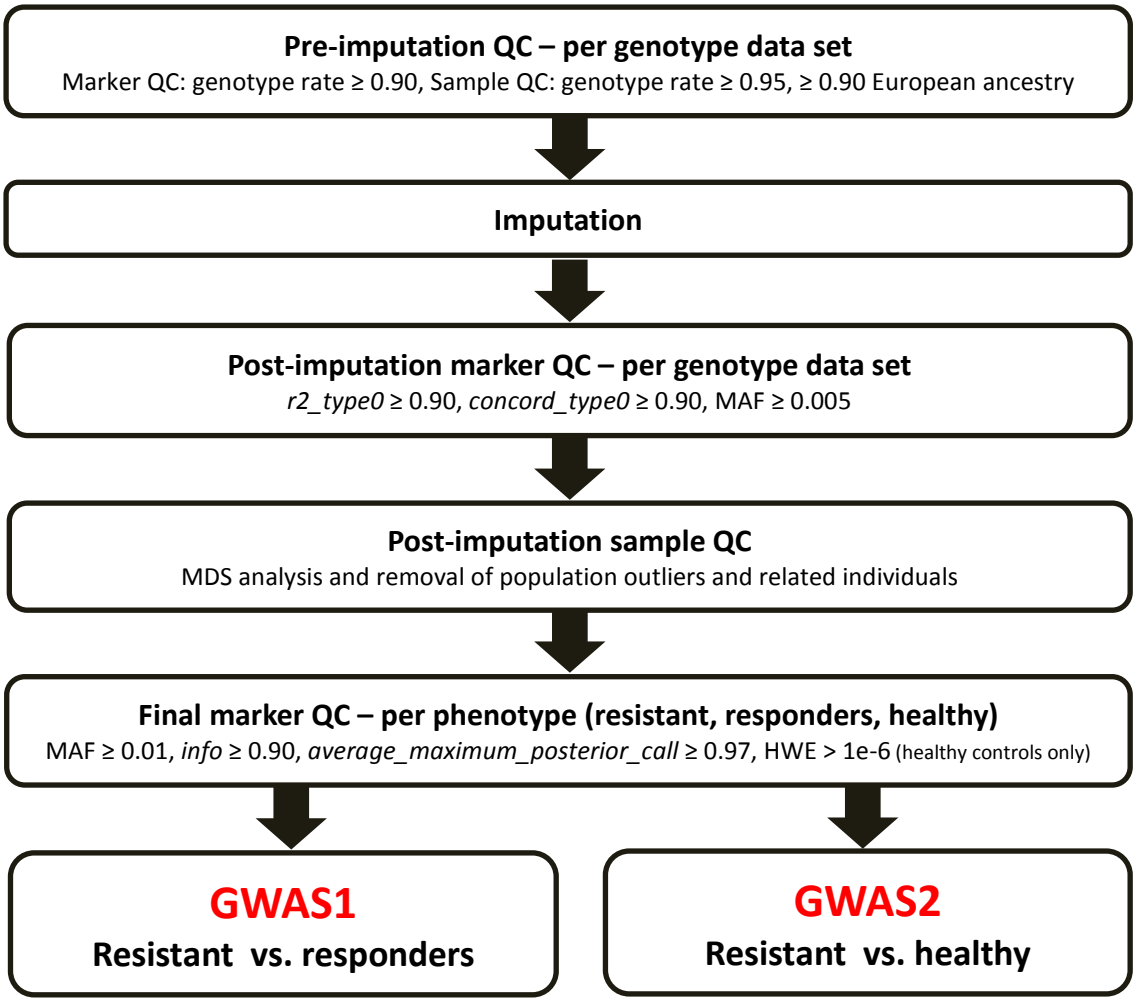
3.3 Methods

I performed two GWAS analyses:

a) GWAS1: comparison of individuals with DRE and individuals with drug-responsive epilepsy
b) GWAS2: comparison of individuals with DRE and healthy controls

I performed both analyses following the protocol outlined in Figure 3.1.

Figure 3.1: GWAS protocol



QC = quality control; MAF = minor allele frequency; HWE = Hardy-Weinberg equilibrium; MDS = multidimensional scaling. IMPUTE metrics include *r2_type0* and *concord_type0*. SNPTEST metrics include *average_maximum_posterior_call* and *info.*

3.3.1 Peripheral blood collection, DNA extraction, and SNP genotyping

Peripheral blood samples were collected from all participants. DNA was extracted using standard procedures and SNP genotyping was performed. The majority of samples were genotyped using the Illumina platform and a small number using the Affymetrix platform (for details see Tables 3.1 and 3.2).

Table 3.1: Genotyping platforms used for epilepsy samples

Country (Ancestry)	Number of epilepsy cases (drug resistant and drug responsive)	Genotyping platform(s)	Chip(s)
UK	2,623	Illumina	OmniExpress-12 v1.1 OmniExpress-24 v1.1 HumanOmniExpress Human610-Quad HumanHap550v3 HumanHap550-Quad Illumina 1.2M
Ireland	410	Illumina	OmniExpress-12 v1.1 OmniExpress-24 v1.1 Human610-Quad HumanHap550-Quad HumanOmni1-Quad
Belgium	290	Illumina	OmniExpress-12 v1.1 OmniExpress-24 v1.1 Human610-Quad HumanHap300v1 HumanHap300v2
Italy	221	Illumina	OmniExpress-12 v1.1 OmniExpress-24 v1.1
Germany	532	Illumina	OmniExpress-12 v1.1 OmniExpress-24 v1.1
		Affymetrix	Array 6.0
Netherlands	58	Illumina	OmniExpress-12 v1.1 OmniExpress-24 v1.1

Table 3.2: Genotyping platforms used for healthy control samples

Country (Ancestry)	Number of healthy controls	Genotyping platform(s)	Chip(s)
UK	5,272	Illumina	Illumina 1.2M
Ireland	2,223	Illumina	HumanOmni1-Quad
Belgium	1,662	Illumina	HumanHap300v1 HumanHap300v2
Italy	261	Affymetrix	Array 6.0
Germany	1,317	Affymetrix	Array 6.0

3.3.2 Imputation

All genotype data sets were imputed to the 1000 Genomes July 2011 reference using IMPUTE2 (459) at deCODE.

3.3.3 Sample and marker QC procedures

Pre-imputation QC:

Pre-imputation marker and sample QC were performed separately for each data set at deCODE using PLINK 1.9 (426) and included the following steps:

1. Removal of markers with < 0.90 genotype rate
2. Extraction of genotypes for 2,766 ethnicity-sensitive SNPs common to all Illumina SNP arrays
3. STRUCTURE 2.2 software (460) was used to derive European, Asian, and African ancestry probabilities, with the following Hapmap samples as respective reference populations: Yoruba in Ibadan, Nigeria; Japanese in Tokyo, Japan; Han Chinese in Beijing, China; and Utah residents with ancestry from Northern and Western Europe
4. Exclusion of samples with < 0.90 European ancestry in individual data sets
5. Exclusion of samples with > 0.05 missing genotype rates
6. Exclusion of samples where gender determined from the genotype did not match the reported gender

Array-specific maps retrieved from the Wellcome Trust website (http://www.well.ox.ac.uk/~wrayner/strand/) were used to update all marker positions and chromosome numbers to the Genome Reference Consortium Human Build 37 (GRCh37) and all A/T and C/G markers were removed to avoid strand issues. The genotypes were then split up according to chromosome arms. Phased haplotypes were created for every data set using SHAPEIT v2 (461), with the recommended effective population size setting (11,418) and the 1000 Genomes phase 1 (July 2011) integrated (v3) map files as the reference. Following the haplotype phasing, imputation was performed for every data set using IMPUTE version 2.3.0 (414, 415, 462), with the recommended effective population size setting (20,000) and 1000 Genomes phase 1 (July 2011) integrated (v3) genotypes as the reference (163).

I performed additional marker and sample QC as follows:

Marker QC:

I first performed a marker QC for each data set separately based on IMPUTE version 2.3.0 metrics, using the following thresholds:

- *r2_type0* ≥ 0.90

- *concord_type0* ≥ 0.90

- MAF ≥ 0.005

I then performed a further marker QC for each phenotype group separately (i.e. drug-resistant cases, drug-responsive cases, and healthy controls) using the thresholds listed

below. I used SNPTEST v2 (434) to generate the following metrics: *info*, *average_maximum_posterior_call,* and *controls_hwe.*

1. Drug-resistant cases:

   - MAF ≥ 0.01

   - *info* ≥ 0.90

   - *average_maximum_posterior_call* ≥ 0.97

2. Drug-responsive cases:

   - MAF ≥ 0.01

   - *info* ≥ 0.90

   - *average_maximum_posterior_call* ≥ 0.97

3. Healthy controls:

   - MAF ≥ 0.01

   - *info* ≥ 0.90

   - *controls_hwe* > 1e-6 (healthy controls only)

   - *average_maximum_posterior_call* ≥ 0.97

Sample QC

As all samples with < 0.95 genotype rate had already been removed at the time of imputation, I did not perform any additional genotype rate QC.

Using PLINK 1.9 (426), I created a subset of markers independent of each other with respect to LD using a window size of 150 markers, shifting by 10 markers at a time and removing one half of every SNP pair with genotypic r2 > 0.4 (PLINK command: *--indep-*

pairwise 150 10 0.4). Using this subset of markers, I calculated heterozygosity (PLINK command: --*het*) and IBS/IBD (PLINK command: --*genome*). I then removed:

1. All samples with outlying heterozygosity values (> 4 standard deviations from the median of the whole sample)
2. Duplicate and related individuals, by excluding one sample out of each pair of samples with Pi-hat (proportion of IBD) > 0.125

To infer the population structure and detect population outliers, I performed MDS analysis with PLINK 1.9. I visualised the results in R (463) using the first two principal coordinates for plotting. I excluded the outliers using visual inspection. In addition, I performed MDS analysis using EpiPGX samples against the HapMap 3 samples. The HapMap 3 collection includes 1,301 samples from 11 populations: African ancestry in Southwest USA, Utah residents with Northern and Western European ancestry from the Centre d'Etude du Polymorphism Humain (CEPH) collection, Han Chinese in Beijing, China, Chinese in Metropolitan Denver, Colorado, Gujarati Indians in Houston, Texas, Japanese in Tokyo, Japan, Luhya in Webuye, Kenya, Mexican ancestry in Los Angeles, California, Maasai in Kinyawa, Kenya, Toscans in Italy, and Yoruba in Ibadan, Nigeria (198).

3.3.4 Association analysis

I performed the single variant association tests using Factored Spectrally Transformed Linear Mixed Models (FaSTLMM) algorithm for GWAS (446).

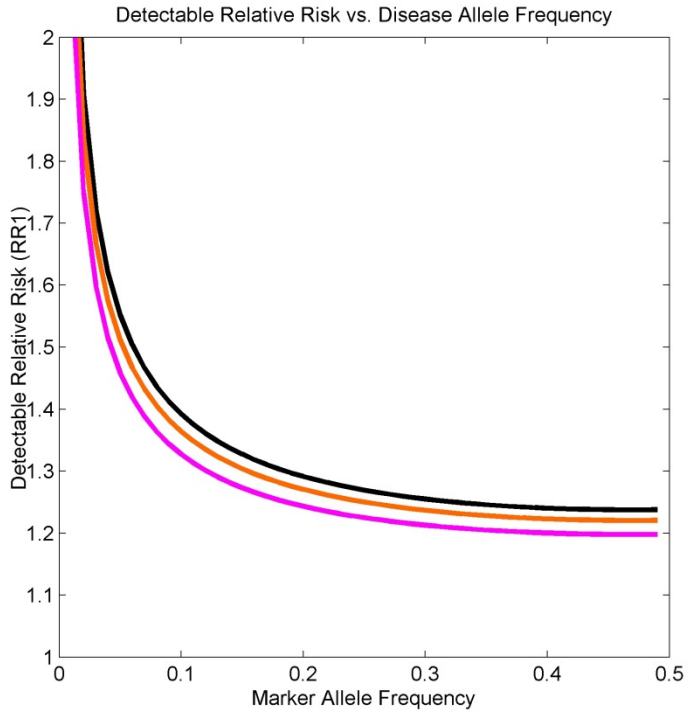
3.3.5 Power and sample size calculations

I performed power and sample size calculations using the Power for Genetic Association Analysis Calculator (464).

3.3.5.1 GWAS1 power and sample size calculations

I used the following assumptions for GWAS1 calculations: a co-dominant penetrance model (where two alleles have an equal effect on the phenotype) and a prevalence of DRE in the epilepsy population of 0.3. As evident from Figure 3.2, inclusion of 2,489 drug-resistant cases and 1,626 controls with drug-responsive epilepsy (control-to-case ratio 0.65) allows for 80% power to detect variants with MAF ≥ 0.05 carrying a RR of approximately 1.5, and variants with MAF ≥ 0.2 carrying a RR of approximately 1.2 to 1.3. Please see Section 2.4.2 for details on the sample sizes of both cohorts used for these calculations.

Figure 3.2: GWAS1 – minimum detectable RR at P = $5*10^{-8}$ for different power levels



All calculations assume a co-dominant model, prevalence of DRE in the epilepsy population 0.3, number of cases 2,489, number of controls 1,626, and a genome-wide statistical significance threshold $5*10^{-8}$.

90% power

80% power

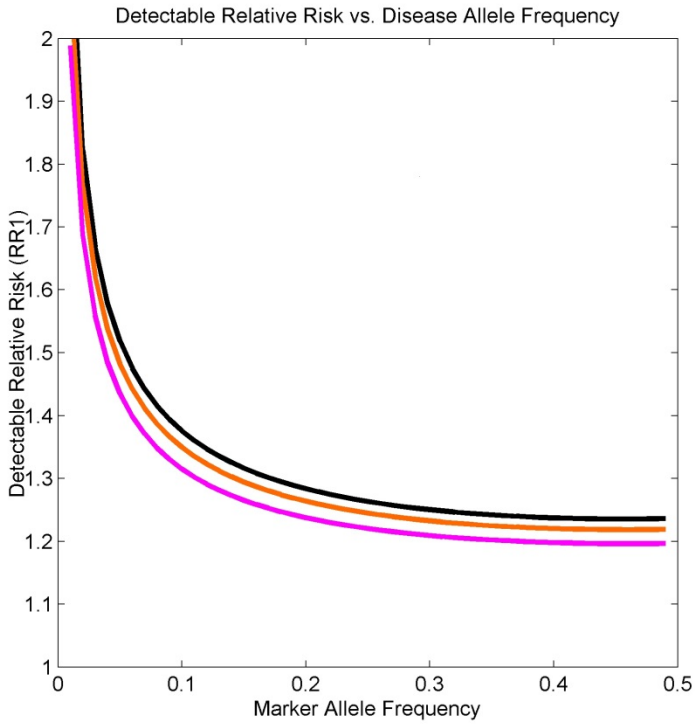
60% power

3.3.5.2 GWAS2 power and sample size calculations

I used the following assumptions for GWAS2 calculations: a co-dominant penetrance model, and a prevalence of DRE in the general population 0.0015. The prevalence of DRE was estimated assuming an epilepsy prevalence of 0.005 in the general population

(32) and a 0.3 rate of drug resistance (20, 28, 43). As evident from Figure 3.3, inclusion of 2,489 drug-resistant cases and 10,695 healthy controls (control to case ratio 4.3) would allow for 80% power to detect variants with MAF ≥ 0.05 carrying a RR of approximately 1.5, and variants with MAF ≥ 0.2 carrying a RR of approximately 1.2 to 1.3 (Figure 3.3). Please see Section 2.4.2 for details on the sample sizes of both cohorts used for these calculations.

Figure 3.3: GWAS2 – minimum detectable RR at P = $5*10^{-8}$ for different power levels



All calculations assume a co-dominant model, prevalence of DRE in the general population 0.0015, number of cases 2,489, number of controls 10,695, and a genome-wide statistical significance threshold $5*10^{-8}$.

90% power

80% power

60% power

For GWAS1 and GWAS2 power and sample size curves see Appendix 6.

3.3.6 Visualisation of results

I used R (463) to generate Manhattan and QQ plots, and LocusZoom (465) to generate plots of top associated loci.

3.3.7 Interpretation of results

To obtain insight into which genes may account for the genome-wide associations, I first identified the SNPs in high LD ($r2 > 0.8$) with the lead associated SNP at each locus using the HaploReg v4.1 tool which utilises LD information from the 1000 Genomes Project (466, 467). I used the ENSEMBL Variant Effect Predictor (468, 469) to check for potential impact of all identified SNPs on either protein structure or regulation of gene expression.

I prioritised candidate genes on the basis of the following:

- Proximity (location within 250kb (kilobase) from the lead SNP at a locus)
- Preferential expression in the brain, based on the gene expression data available in the Genotype-Tissue Expression (GTEx) database (470, 471)

- Presence of a missense variant in the genome-wide significant locus, as annotated by ENSEMBL (468, 469)
- Observation of a central nervous system phenotype in knockout mouse models, based on the information available in the Mouse Genome Informatics (MGI) database (472, 473)
- Presence of a significant cis-eQTL (Bonferroni corrected $P < 8*10^{-10}$) in the genome-wide significant locus, based on the information in publically available databases: GTEx (470, 471), Brain xQTL (474, 475), and Braineac (476, 477)

To place the results of the present work in the context of the relevant published literature, I examined the output of GWAS2 for top SNPs identified in both ILAE GWAS analyses as variants influencing epilepsy risk (261, 262). In addition, I extracted the results for variants in selected genes that had previously been associated with resistance to multiple AEDs from the output of GWAS1 and GWAS2. I selected the genes for which I have been able to identify multiple, rather than single reports (Table 3.3). As discussed in Section 1.8.2, the published literature for some of these candidate genes may be conflicting, with positive and negative assicociations identified by different research groups.

Table 3.3: Genes associated with drug resistance in epilepsy

Gene	Protein	Polymorphisms associated with drug response / resistance
CYP2C9	Cytochrome P450 2C9 enzyme	CYP2C9*2 (rs1799853), CYP2C9*3 (rs1057910) (274, 278, 279, 281)
CYP2C19	Cytochrome P450 2C9 enzyme	CYP2C19*2 (rs4244285) (382, 478)
ABCB1	P-gp	rs1045642 (297, 298)
SCN1A	Sodium voltage-gated channel type I alpha subunit; $Na_v1.1$	rs2298771 (479, 480)

3.4 Results

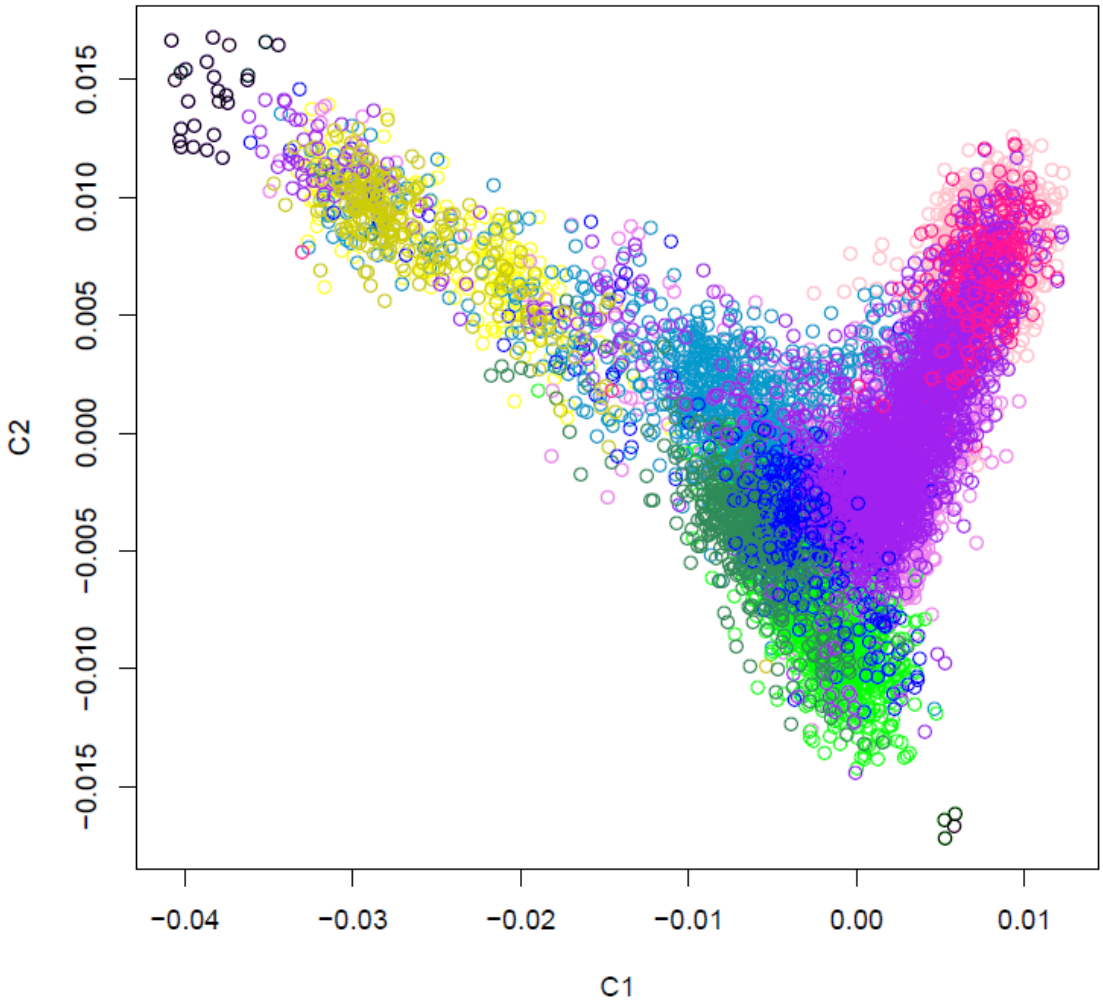
3.4.1 Genotyping and imputation

All 16 data sets were genotyped and imputed at deCODE using the 1000 Genomes July 2011 reference. This generated approximately 40,000,000 SNPs per case overall before QC. The X chromosome was not analysed as SNP data were not available for all cohorts.

3.4.2 Cohorts

I studied 14,829 samples in total, including: 2,489 cases with drug-resistant epilepsy and 1,645 cases with drug-responsive epilepsy from 7 cohorts (Belgium, Netherlands, Germany, Ireland, UK - London, UK - Liverpool, Italy), in addition to 10,695 healthy individuals from 6 cohorts (Belgium, Germany, Ireland, UK - NBS, UK - 1958 Birth Cohort, Italy). MDS analysis suggested that all samples clustered in the European ethnic origin, as expected (Figures 3.4 and 3.5). On the basis of visual inspection of the MDS plot (Figure 3.4), I excluded 29 population outliers from downstream analyses. In addition, I excluded 85 samples due to relatedness and 21 samples due to outlying heterozygosity values, leaving in total 14,694 samples after QC: 2,423 drug-resistant, 1,626 drug-responsive, and 10,645 healthy controls. Additional MDS plots showing the distribution of cases and controls for both GWAS analyses are included in Appendix 7.

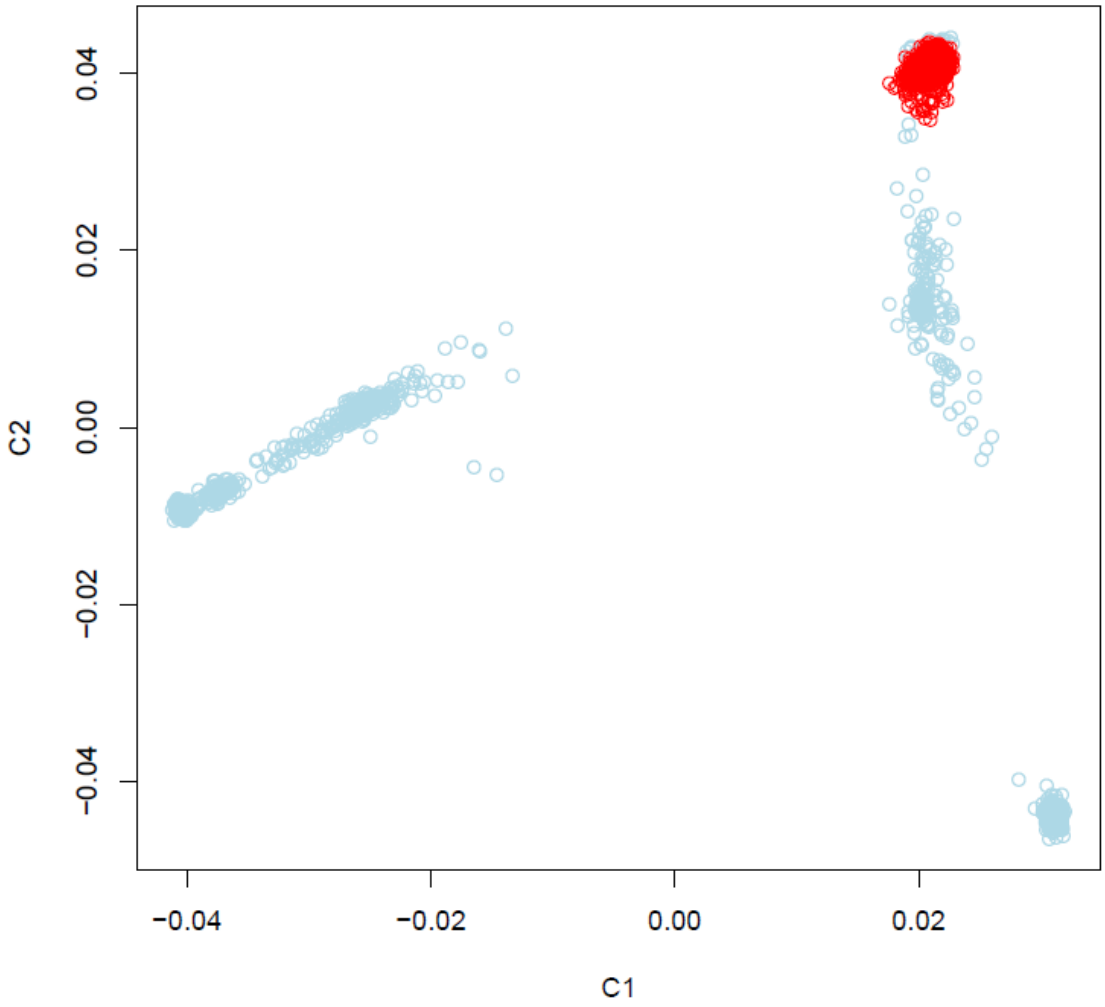
Figure 3.4: MDS analysis of cases and controls considered for drug resistance GWAS



- Belgian epilepsy cases
- Dutch epilepsy cases
- Belgian healthy controls
- German epilepsy cases
- German healthy controls
- Irish epilepsy cases
- Irish healthy controls
- UK epilepsy cases
- UK healthy controls
- Italian epilepsy cases
- Italian healthy controls
- Excluded samples

C1 = principal coordinate 1; C2 = principal coordinate 2. Epilepsy cases include both drug-resistant and drug-responsive cases.

Figure 3.5: MDS analysis of EpiPGX samples included in drug resistance GWAS and the HapMap population



O HapMap samples

O EpiPGX samples

C1 = principal coordinate 1; C2 = principal coordinate 2.

3.4.3 GWAS1 (comparison of drug-resistant individuals with drug responders)

After application of QC criteria, I included 2,423 cases with DRE and 1,626 cases with drug-responsive epilepsy in the association analysis. After application of marker QC,

5,952,081 SNPs remained in the association analysis. I performed association testing using FaSTLMM. I noted an inflation factor of 1.008, suggesting adequate control for possible cryptic stratification (see Figure 3.7 for QQ plot).

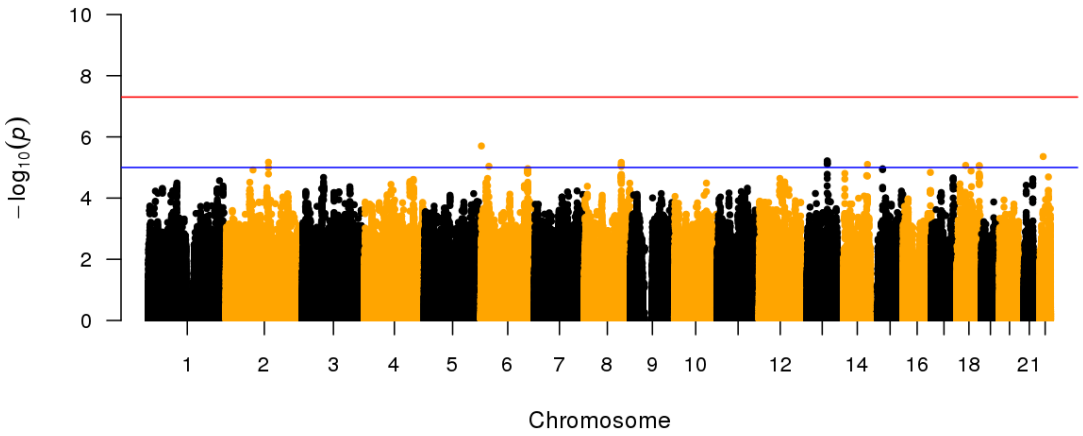
No SNPs reached genome-wide significance in this analysis. The top associated SNP was rs2816283 ($P = 1.98*10^{-06}$) located in an intergenic region at 6p25.3 (see Table 3.3).

Table 3.4: GWAS1 – top associated loci (none of these loci reached genome-wide significance)

SNP	Cytogenetic band	Base pair position	Allele 1, allele 2	MAF (minor allele)	Candidate gene	Annotation	P-value (FaSTLMM)	OR (95% CI)
rs2816283	6p25.3	1529872	T, C	0.2499 (T)	Multiple	Intergenic	$1.98*10^{-06}$	0.94 (0.92-9.97)
rs79316678	22q12.1	27786568	A, G	0.0296 (G)	Multiple	Intergenic	$4.38*10^{-06}$	1.16 (0.84-1.60)
rs4611336	13q31.1	83201920	T, C	0.3104 (C)	*SLITRK1*	Intergenic	$6.03*10^{-06}$	1.05 (0.94-1.18)
rs9531372	13q31.1	83197385	A, T	0.3894 (T)	*SLITRK1*	Intergenic	$6.40*10^{-06}$	1.05 (0.94-1.18)
rs1683557	2q21.2	134816184	G, A	0.0165 (A)	Multiple	Intergenic	$6.71*10^{-06}$	1.10 (0.89-1.35)

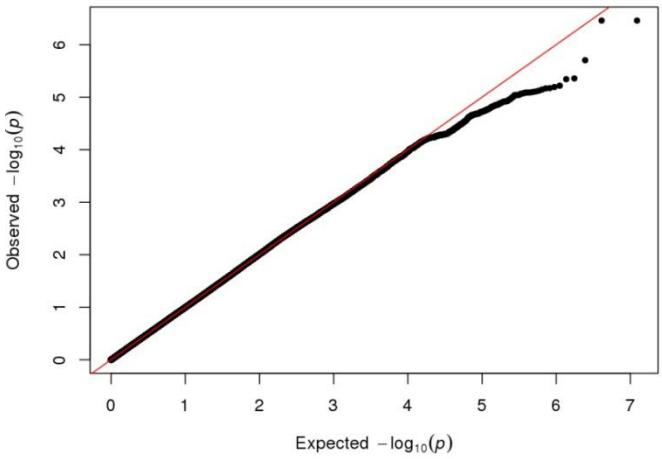
Base pair position refers to human genome build 37 (GRCh37, hg19). MAF is from all populations from the 1000 Genomes Project. Candidate gene refers to the most plausible candidate gene attributable to the signal, based on proximity (450). OR (odds ratio) corresponds to allele 2. Annotation refers to type of SNP; nc transcript variant refers to a transcript variant of a non-coding gene; upstream refers to upstream of a gene.

Figure 3.6: GWAS1 – Manhattan plot



X-axis shows genomic coordinates, Y-axis shows the negative algorithm of the association P-value for each SNP. Each SNP is represented as a dot. The red line shows the threshold of genome-wide significance.

Figure 3.7: GWAS1 – QQ plot



X-axis shows the negative algorithm of the expected association P-value for each SNP under the null hypothesis. Y-axis shows the negative algorithm of the observed P-value for each SNP. P-value refers to single association tests performed with FaSTLMM.

3.4.4 GWAS2 (comparison of drug-resistant individuals with healthy controls)

After application of the QC criteria, I included 2,423 cases with DRE and 10,645 healthy controls in the association analysis. After application of marker QC, 5,919,078 SNPs remained in the analysis. I performed association testing using FaSTLMM. I noted an inflation factor of 1.016, suggesting adequate control for possible cryptic stratification (see Figure 3.9 for QQ plot).

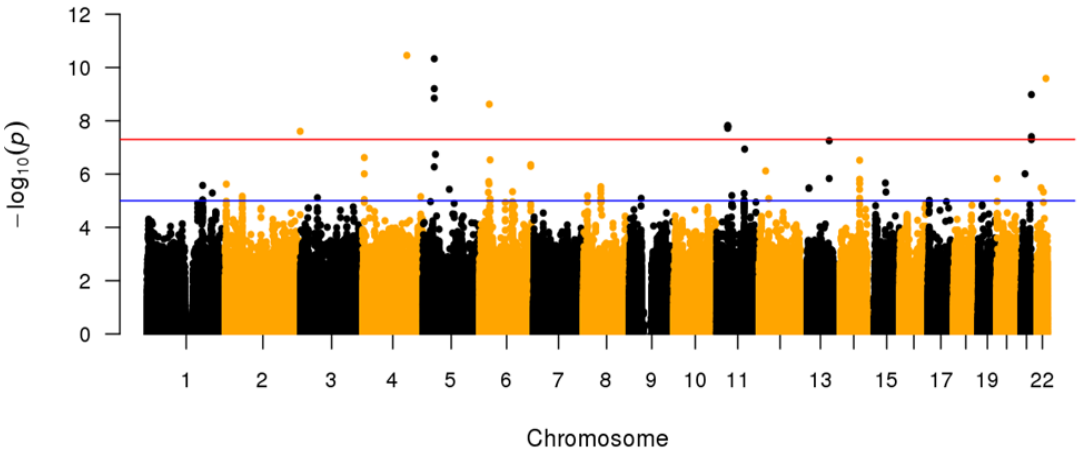
I identified 8 loci with genome-wide significance (Table 3.4 and Figure 3.8).

Table 3.5: GWAS2 – genome-wide associated loci at $P < 5*10^{-8}$

SNP	Cytogenetic band	Base pair position	Allele 1, allele 2	MAF (minor allele)	Candidate gene	Annotation	P-value (FaSTLMM)	OR (95% CI)
rs75700350	4q31.1	140189612	C, T	0.0529 (T)	*MGARP*	Intronic	$3.52*10^{-11}$	1.07 (1.05-1.09)
rs4629621	5p13.2	36426049	T, C	0.1512 (T)	*SLC1A3*	Intergenic	$4,69*10^{-11}$	1.06 (0.97-1.15)
rs5765116	22q13.31	45408407	G, A	0.4589 (G)	Multiple	Intergenic	$2.58*10^{-10}$	1.03 (0.98-1.09)
rs8127410	21q22.3	47507039	A, G	0.3894 (G)	Multiple	Intergenic	$1.04*10^{-09}$	0.97 (0.91-1.02)
rs150512569	6p21.33	30495053	G, A	0.1032 (A)	Multiple	Intergenic	$2.38*10^{-09}$	1.04 (0.98-1.11)
rs16927514	11p13	35505321	T, C	0.0280 (C)	*PAMR1*	Intronic	$1.52*10^{-08}$	0.94 (0.84-1.05)
rs73999651	2q37.3	238054898	G, T	0.0873 (T)	Multiple	Intergenic	$2.48*10^{-08}$	0.93 (0.82-1.06)
rs4516077	13q31.2	89193635	A, G	0.4922 (G)	*LINC00433*	nc transcript variant	$5.59*10^{-08}$	0.97 (0.93-1.02)

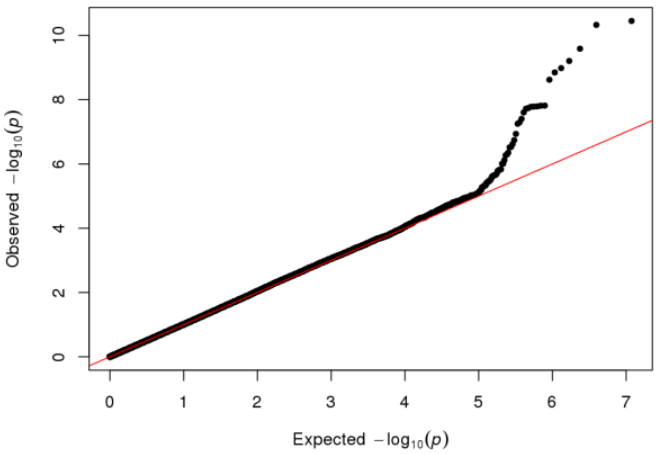
Base pair position refers to human genome build 37 (GRCh37, hg19). MAF is from all populations from the 1000 Genomes Project. Candidate gene refers to the most plausible candidate gene attributable to the signal, based on proximity (450). OR (odds ratio) corresponds to allele 2. Annotation refers to type of SNP; nc transcript variant refers to a transcript variant of a non-coding gene; upstream refers to upstream of a gene.

Figure 3.8: GWAS2 – Manhattan plot



X-axis shows genomic coordinates, Y-axis shows the negative algorithm of the association P-value for each SNP. Each SNP is represented as a dot. The red line shows the threshold of genome-wide significance.

Figure 3.9: GWAS2 – QQ plot

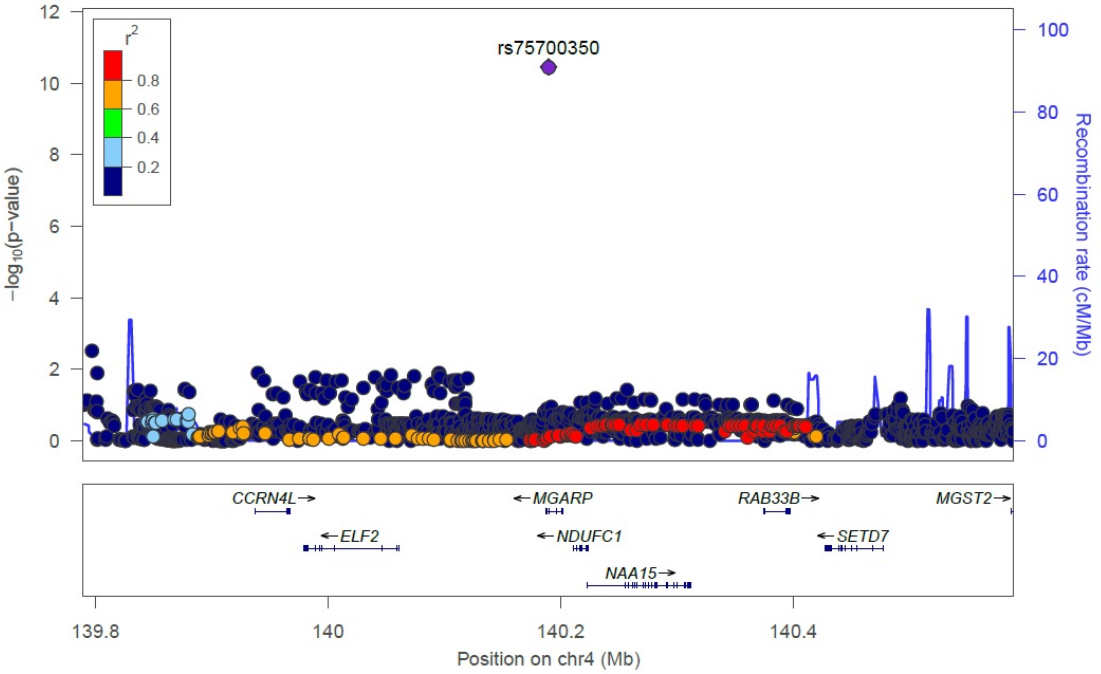


X-axis shows the negative algorithm of the expected association P-value for each SNP under the null hypothesis. Y-axis shows the negative algorithm of the observed P-value for each SNP. P-value refers to single association tests performed with FaSTLMM.

3.4.4.1 Signal at locus 4q31.1

The top genome-wide association signal (rs75700350) was located at 4q31.1 (Figure 3.10). This signal was centred on the mitochondria localised glutamic acid rich protein (*MGARP*) gene. rs75700350 was the only SNP with genome-wide significance in this region. Usually, associated regions in GWAS are expected to contain a set of SNPs with varying degrees of association due to local LD patterns (465). Although in 1000 Genomes, there are 34 SNPs in high LD with rs75700350 (467) and 33 of them were included in the association analysis, none of them reached (or was close to reaching) genome-wide significance indicating that this hit could be an artefact. rs75700350 was a Type0 SNP (i.e. imputed and not directly genotyped in any of the data sets). The imputation quality metrics (*info*, *average_maximum_posterior_call*, *r2_type0*, *concord_type0*) were satisfactory for all genotype data sets, as well as across the cases and the controls. That said, GWAS2 included publically available controls genotyped on different platforms than the DRE cases. In such settings, imputation can introduce bias, especially if the MAF in the study data differs substantially from the reference (481). In this case, there was a significant difference between the MAF in the reference (0.0529; 1000 Genomes) and the study data. In some of the data sets including cases as well as controls, the MAF was as low as 0.007.

Figure 3.10: Genomic context of the 4q31.1 signal (LocusZoom plot)



Chr = chromosome, cM = centimorgan, Mb = megabase, cM/Mb = recombination rate (shown as the blue line running through the plot, with peaks indicating the recombination hotspots), P-value refers to single variant association tests performed with FaSTLMM, r^2 = measure of LD. LD data were taken from the 1000 Genomes project (European population), human genome build 37 (GRCh37, hg19), March 2012.

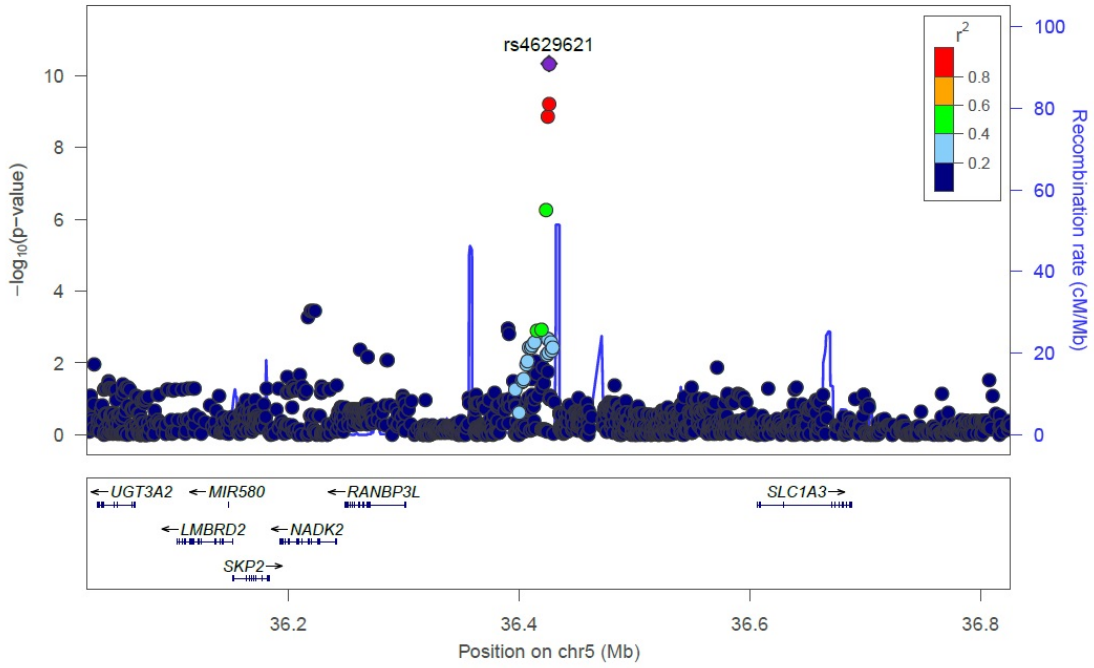
3.4.4.2 Signal at locus 5p13.2

The SNP with the second lowest P-value (rs4629621) was located in an intergenic region at 5p13.2 (Figure 3.11). Two other variants in high LD ($r^2 > 0.8$) with rs4629621 and genome-wide significance (rs4479866 and rs6861831) were found in this region, as well as several variants with nominal significance and $r^2 > 0.2$.

There are four genes within 250kb of the lead SNP (rs4629621): S-phase kinase-associated protein 2 (*SKP2*), RAN binding protein 3 like (*RANBP3L*), nicotinamide

adenine dinucleotide kinase 2 (*NADK2*), and solute carrier family 1 member 3 (*SLC1A3*). *SKP2* encodes one of the subunits of the ubiquitin-protein ligase complex and has a role in phosphorylation-dependent ubiquitination. It is a protooncogene, implicated in the pathogenesis of multiple types of cancer (482, 483). *RANBP3L* is associated with bone development and mesenchymal stem cell differentiation (484). The *NADK2* gene encodes a mitochondrial kinase that catalyses the phosphorylation of NAD to yield NAD phosphate (485). *SLC1A3* encodes excitatory amino acid transporter 1 (EAAT1), a glial glutamate transporter (486). None of the SNPs in the locus has been identified as a significant cis-eQTL for any of the four genes in the aforementioned eQTL databases (471, 475, 477). Two out of the four genes of interest (*RANBP3L* and *SLC1A3*) are preferentially expressed in the brain (471). While *RANBP3L* has not been associated with any relevant phenotype in animal models, *SLC1A3* knockout mice have been shown to have abnormal kindling responses and seizures (473). Mutations in *SLC1A3* have been linked to epilepsy (487), making *SLC1A3* the most interesting of the four genes.

Figure 3.11: Genomic context of the 5p13.2 signal (LocusZoom plot)



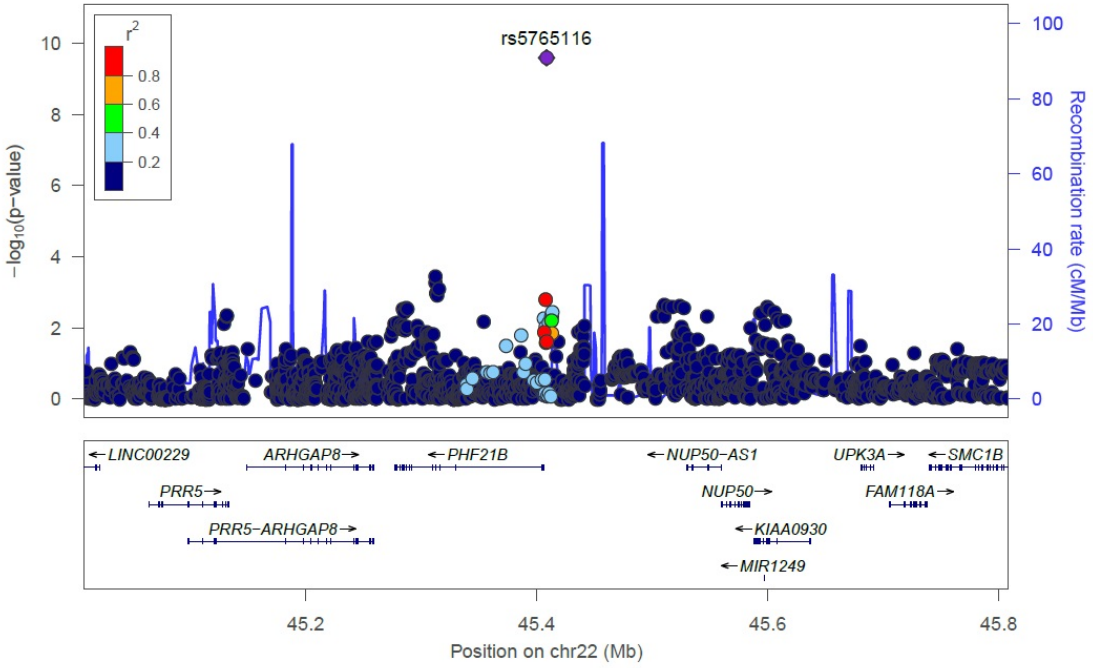
Chr = chromosome, cM = centimorgan, Mb = megabase, cM/Mb = recombination rate (shown as the blue line running through the plot, with peaks indicating the recombination hotspots), P-value refers to single variant association tests performed with FaSTLMM, r^2 = measure of LD. LD data were taken from the 1000 Genomes project (European population), human genome build 37 (GRCh37, hg19), March 2012.

3.4.4.3 Signal at locus 22q13.31

The third top associated SNP (rs5765116) was located in an intergenic region at 22q13.31 (Figure 3.12). No other variants in this LD block reached genome-wide significance. In 1000 Genomes, there are 16 SNPs in high LD with rs5765116 (467). While seven of them did not pass the QC and were not included in the association analysis, the remaining nine were included and they all reached nominal significance. However, upon close inspection of the LocusZoom plot, the pattern of association for these SNPs appears atypical, suggesting the possibility that this signal was an artefact.

rs5765116 was a Type0 SNP (i.e. imputed and not directly genotyped in any of the data sets). The imputation quality metrics (*info*, *average_maximum_posterior_call*, *r2_type0*, *concord_type0*) were satisfactory for all genotype data sets, as well as across the cases and the controls.

Figure 3.12: Genomic context of the 22q13.31 signal (LocusZoom plot)

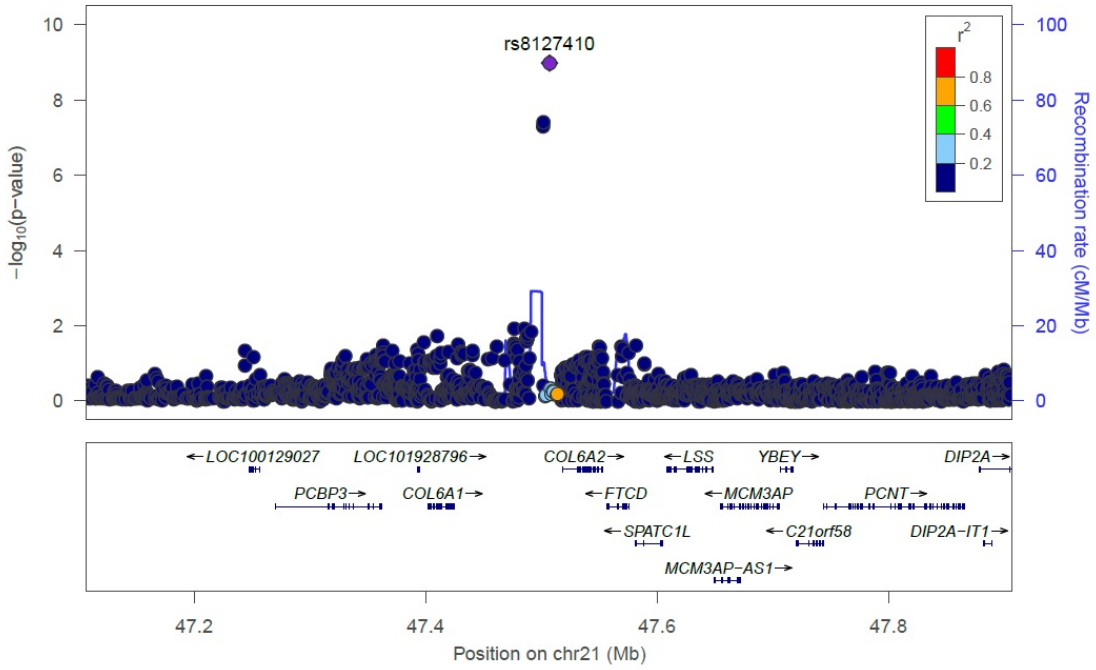


Chr = chromosome, cM = centimorgan, Mb = megabase, cM/Mb = recombination rate (shown as the blue line running through the plot, with peaks indicating the recombination hotspots), P-value refers to single variant association tests performed with FaSTLMM, r^2 = measure of LD. LD data were taken from the 1000 Genomes project (European population), human genome build 37 (GRCh37, hg19), March 2012.

3.4.4.4 Signal at locus 21q22.3

The fourth top associated SNP (rs8127410) was located in an intergenic region at 21q22.3 (Figure 3.13). In 1000 Genomes, there are six SNPs in high LD with rs8127410 (467); however, none of them passed the QC. Two other variants with genome-wide significance (rs2839104 and rs2839103) were found in this region. These two variants are not in LD with rs8127410, but they are in strong LD with each-other. Despite the missing information on SNPs in high LD with the associated variants, it is most likely that these hits were artefacts. Both associated variants were Type0 SNPs (i.e. imputed and not directly genotyped in any of the data sets). The imputation quality metrics (*info*, *average_maximum_posterior_call*, *r2_type0*, *concord_type0*) were satisfactory for all genotype data sets, as well as across the cases and the controls.

Figure 3.13: Genomic context of the 21q22.3 signal (LocusZoom plot)

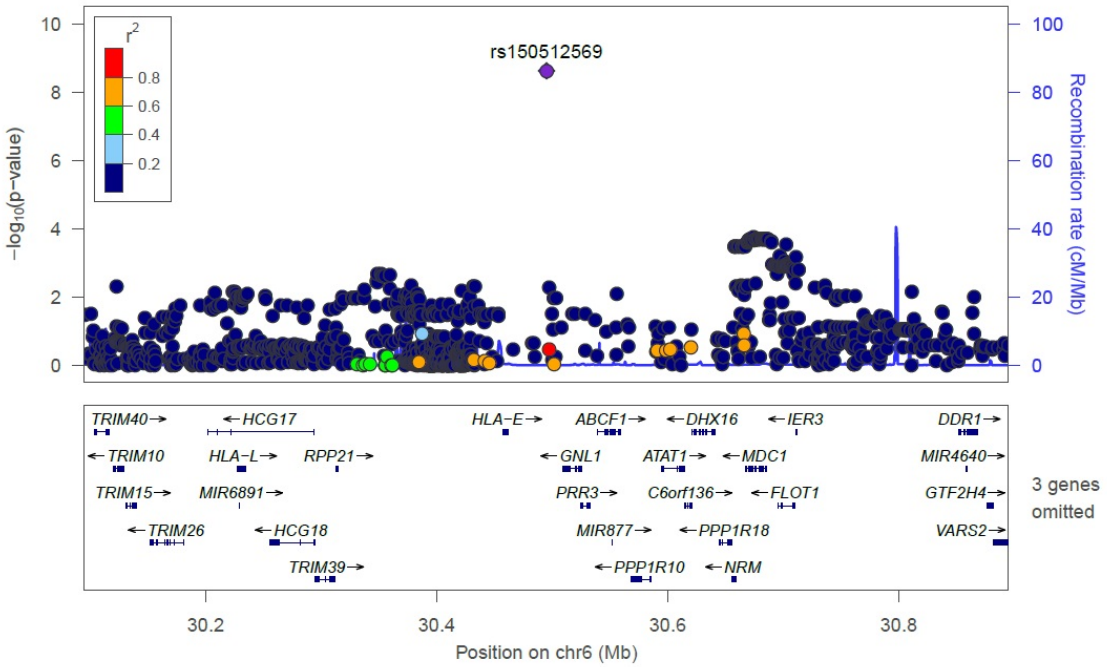


Chr = chromosome, cM = centimorgan, Mb = megabase, cM/Mb = recombination rate (shown as the blue line running through the plot, with peaks indicating the recombination hotspots), P-value refers to single variant association tests performed with FaSTLMM, r^2 = measure of LD. LD data were taken from the 1000 Genomes project (European population), human genome build 37 (GRCh37, hg19), March 2012.

3.4.4.5 Signal at locus 6p21.33

The next signal (rs150512569) was located in an intergenic region at 6p21.33 (Figure 3.14). In 1000 genomes, there are five SNPs in high LD with rs150512569 (467). These SNPs did not pass the QC and were thus not included in the association analysis. Nevertheless, it is very likely that that this SNP was an artefact as it is not supported by other SNPs with genome-wide or nominal significance.

Figure 3.14: Genomic context of the 6p21.33 signal (LocusZoom plot)



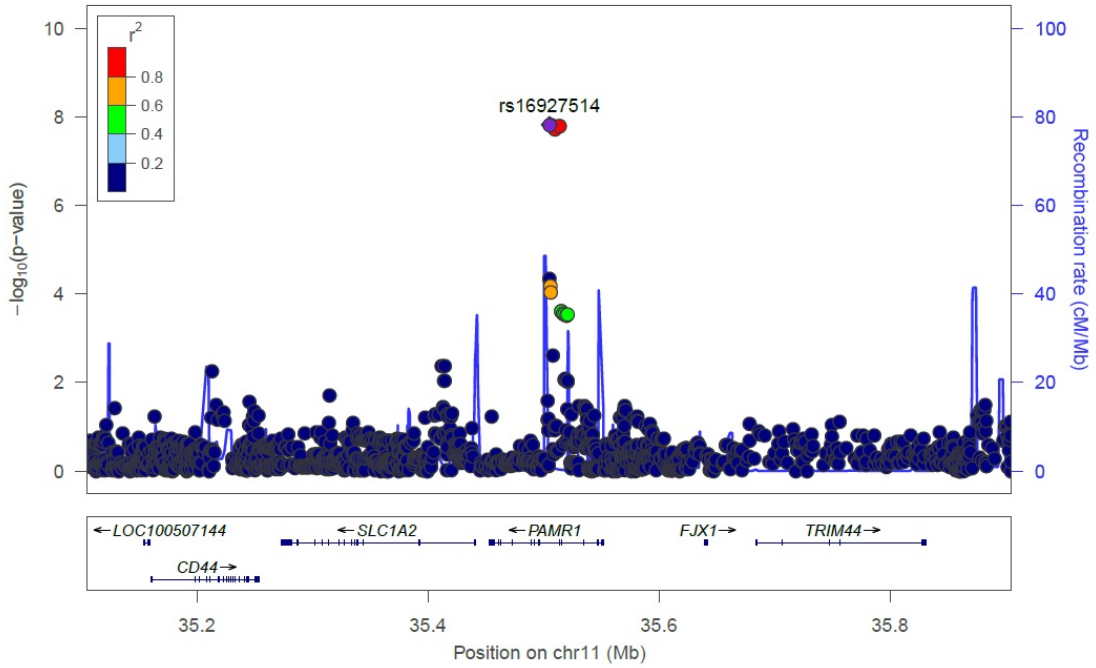
Chr = chromosome, cM = centimorgan, Mb = megabase, cM/Mb = recombination rate (shown as the blue line running through the plot, with peaks indicating the recombination hotspots), P-value refers to single variant association tests performed with FaSTLMM, r^2 = measure of LD. LD data were taken from the 1000 Genomes project (European population), human genome build 37 (GRCh37, hg19), March 2012.

3.4.4.6 Signal at locus 11p13

The next signal (rs16927514) was located at 11p13 (Figure 3.15). This signal was centred on the *PAMR1* (peptidase domain containing associated with muscle degeneration) gene (intronic variant) which has so far not been linked to epilepsy. In 1000 Genomes, there are seven SNPs in high LD with rs16927514 (467). Six of them (rs117711020, rs117145777, rs17726212, rs117054435, rs117524439, rs115838855) passed the QC and reached genome-wide significance. All these SNPs are annotated as intronic variants of *PAMR1* in ENSEMBL. In addition to *PAMR1*, there are three other

genes within 250kb from the lead SNP at this locus: solute carrier family 1 member 2 (*SLC1A2*), four-jointed box kinase 1 (*FJX1*), and tripartite motif containing 44 (*TRIM44*). *FJX1* and *SLC1A2* are predominantly expressed in the brain (471), and *SLC1A2* knockout mouse models exhibit seizures, increased susceptibility to pharmacologically induced seizures, and abnormalities of the hippocampus (473). None of the SNPs in the locus is a significant cis-eQTL for any of the four genes (471, 475, 477). Considering its function and expression pattern, *SLC1A2* could be a plausible candidate gene, even though all the significantly associated SNPs are intronic variants of *PAMR1*.

Figure 3.15: Genomic context of the 11p13 signal (LocusZoom plot)

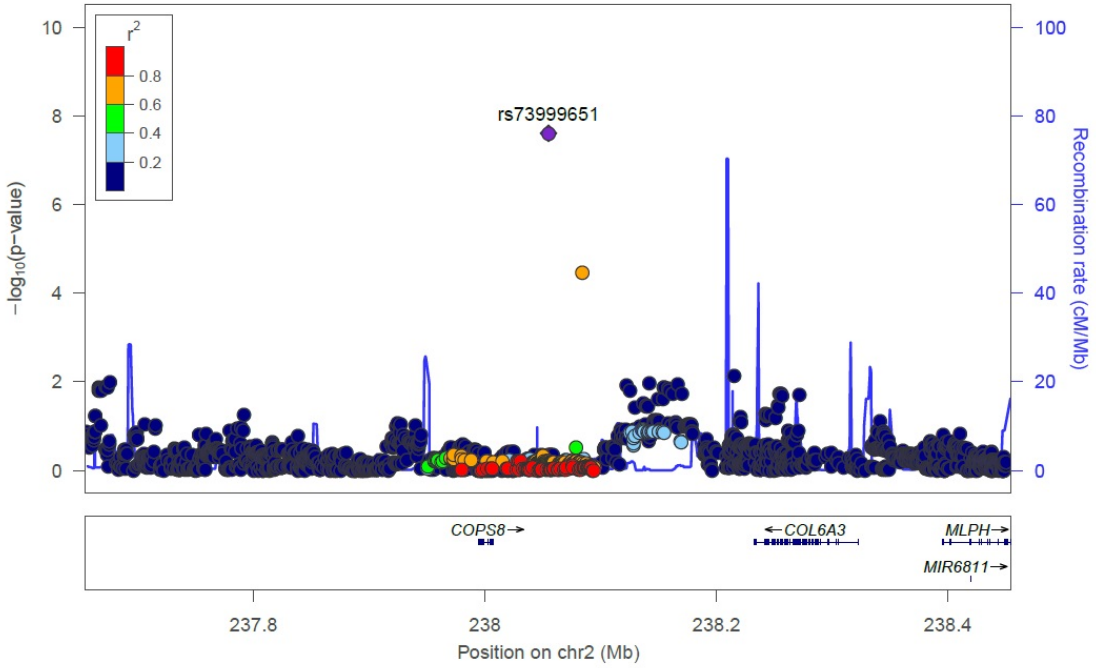


Chr = chromosome, cM = centimorgan, Mb = megabase, cM/Mb = recombination rate (shown as the blue line running through the plot, with peaks indicating the recombination hotspots), P-value refers to single variant association tests performed with FaSTLMM, r^2 = measure of LD. LD data were taken from the 1000 Genomes project (European population), human genome build 37 (GRCh37, hg19), March 2012.

3.4.4.7 Signal at locus 2q37.3

The next signal (rs73999651) was located in an intergenic region at 2q37.3 (Figure 3.16). In 1000 Genomes, there are 65 SNPs in high LD ($r^2 > 0.8$) with rs75700350 (467), 58 of which passed the QC and were included in the association analysis. None of them reached genome-wide or nominal significance. Only one variant in moderate LD ($r^2 > 0.6$) with rs73999651 had nominal significance. Considering that the lead SNP at this locus was not supported by other SNPs, it is possible that this hit was an artefact, potentially arising in the course of the imputation. rs73999651 was a Type0 SNP (i.e. imputed and not directly genotyped in any of the data sets). The imputation quality metrics (*info*, *average_maximum_posterior_call*, *r2_type0*, *concord_type0*) were satisfactory for all imputation batches, as well as across the cases and the controls. On detailed inspection of the characteristics of this SNP, I noted a significant difference between the MAF in the reference (0.0873; 1000 Genomes) and the UK controls (0.018 to 0.021).

Figure 3.16: Genomic context of the 2q37.3 signal (LocusZoom plot)

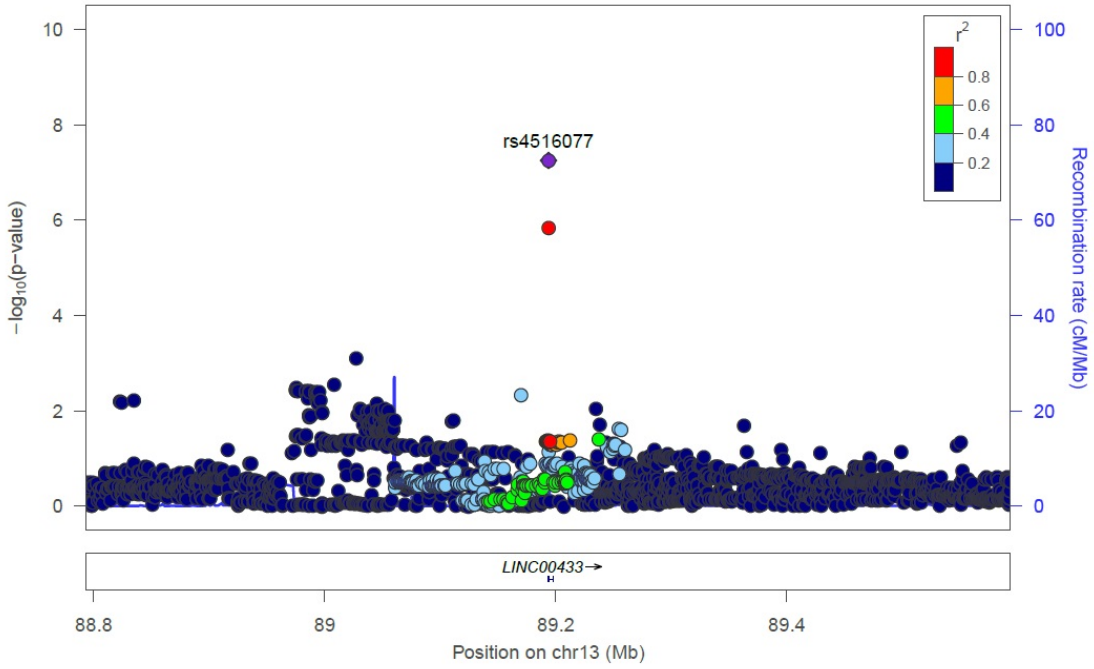


Chr = chromosome, cM = centimorgan, Mb = megabase, cM/Mb = recombination rate (shown as the blue line running through the plot, with peaks indicating the recombination hotspots), P-value refers to single variant association tests performed with FaSTLMM, r^2 = measure of LD. LD data were taken from the 1000 Genomes project (European population), human genome build 37 (GRCh37, hg19), March 2012.

3.4.4.8 Signal at locus 13q31.2

The final signal (rs4516077) was located at 13q31.2 (Figure 3.17), centred on the long intergenic non-protein coding RNA 433 (*LINC00433*) gene (transcript variant of a non-coding RNA gene). In 1000 Genomes, there is one SNP in high LD with the lead SNP (467). This SNP (rs7491918) also reached genome-wide significance. Apart from *LINC00433*, there are no other genes within 250kb of the lead SNP.

Figure 3.17: Genomic context of the 13q31.2 signal (LocusZoom plot)



Chr = chromosome, cM = centimorgan, Mb = megabase, cM/Mb = recombination rate (shown as the blue line running through the plot, with peaks indicating the recombination hotspots), P-value refers to single variant association tests performed with FaSTLMM, r^2 = measure of LD. LD data were taken from the 1000 Genomes project (European population), human genome build 37 (GRCh37, hg19), March 2012.

3.5 Discussion

3.5.1 Comparison of drug-resistant individuals with drug responders

In GWAS1, I compared 2,423 cases with DRE and 1,626 controls with drug-responsive epilepsy. To my knowledge, this is the largest GWAS of DRE to this date. The association analysis did not show a differential distribution of SNPs in individuals with DRE compared with individuals who respond to AEDs. This study was powered to

detect variants with MAF 0.05 carrying a RR for DRE of at least 1.5, or variants with MAF of at least 0.2 carrying a modest RR. The results suggest drug resistance may not be caused by common variants with modest or large effect sizes, although an influence of rare variants or variants with small effect sizes cannot be excluded.

It is possible that drug resistance is caused by a large number of variants with small effect sizes, in line with the so called infinitesimal model. In contrast with the common disease-common variant hypothesis, the infinitesimal model predicts that many variants with small effect sizes (RR less than 1.1) underlie the variable susceptibility for common traits (488). This could possibly mean hundreds or thousands of contributing loci, with affected individuals carrying an excess of risk variants compared with the general population (488, 489). Such pattern has been observed in sudden unexpected death in epilepsy (SUDEP), another severe phenotype associated with DRE. Leu *et al.* showed that the polygenic burden underlying SUDEP consists of contributions from more than a thousand genes (490).

Another possible scenario that cannot be excluded is that drug resistance is caused by a small number of rare variants (MAF < 0.01) which would not have been identified by the present analysis. This would be consistent with the rare variant hypothesis, according to which most of the phenotypic variability in complex traits is due to rare variants with relatively large effect sizes (491). Since an individual variant explains most of the risk only in a fraction of cases, the effects of rare variants may not account for a sufficient amount of the variability in the population to be detectable by standard GWAS (489).

Last but not least, it is possible that variants underlying drug resistance are specific to different types of epilepsy and not the same across all epilepsy types.

3.5.1.1 Limitations

The main limitation of this study is the relatively small sample size. Even though this is the largest GWAS of this type to date, the sample size is still modest, reflecting the difficulty obtaining large numbers of cases with well characterised pharmacologic outcomes. The sample size was insufficient to allow the detection of variants carrying a small risk for DRE. Furthermore, the small sample size did not allow for analysis of variants associated with DRE in epilepsy subtypes. Lastly, there is a lack of independent replication sample.

The analysis used genotype data generated separately on several platforms, which is normal for GWAS conducted in a consortium setting. To mitigate this, the data sets were imputed to the same reference using a unified protocol. It is important to note that the X chromosome was not included in the analysis as the data were not available for all cohorts. Any potentially significant variants on chromosome X could thus not be detected.

The limitations associated with case and control selection are discussed in detail in Section 2.5.3. Due to the retrospective data collection, the classification of cases relied on the quality of medical records. Since the patients were recruited at tertiary centres, there was over-representation of difficult-to-treat cases in the overall sample (2,489

cases with drug-resistant epilepsy and 1,626 cases with drug-responsive epilepsy were available for inclusion in GWAS1). In a general epilepsy population, one would expect only approximately 30% resistant cases (20, 28, 43). In addition, it is possible that the drug-responsive controls were less typical of the general responder population. Individuals recruited in EpiPGX might have required a longer time to achieve remission, or suffered relapses, warranting continuous follow-up at a tertiary centre. Individuals achieving a sustained remission with their first AED trial would be more likely followed locally and would consequently be under-represented in our cohort. The information on the time to remission was not available for the cohort of drug responders in this analysis. Consequently, I have not been able to make comparisons with the published community-based epilepsy cohort outcomes, or speculate whether the drug-responsive cohort included in GWAS1 could be genetically different from a general drug-responsive population and whether the analysis could have been affected. The average number of adequate and appropriate AED trials in individuals with drug-responsive epilepsy was approximately two, which is similar to what has been observed in another study (81).

The issue of the fluctuating course of epilepsy has been discussed in detail in Section 2.5.3.5. The classification of cases with a fluctuating course of epilepsy poses challenges with any definition of epilepsy outcomes because the classification is only valid at one point, and may change at another point in time. It is probable that some of the individuals in our drug-resistant cohort would have achieved remission and that some of the drug-responsive individuals would have relapsed down the line, thus falling into the opposite phenotypic category. As a consequence, both GWAS1 cohorts

might have included a small proportion of cases with a fluctuating epilepsy course that may have a different genomic signature than both the drug-resistant and the drug-responsive populations. The extent to which this issue could have affected the power and the results of GWAS1 is difficult to estimate. The proportion of individuals with a fluctuating course is estimated at approximately 10 to 15% (28, 401), so there could theoretically be a few percent of them in each GWAS1 cohort. Fluctuations between periods of uncontrolled seizures and remissions are more likely in individuals experiencing infrequent seizures (382). Since the EpiPGX definition of DRE required a minimum of four seizures in the 12 months prior to the last follow-up, it likely excluded a fair proportion of such cases. However, this was not the case with the definition of drug response.

3.5.1.2 Concluding remarks

In conclusion, GWAS1 does not provide a definitive answer regarding the genetic basis of drug resistance in epilepsy. It is possible that drug resistance is caused by rare variants or variants with very small effect sizes that could not be detected by this analysis.

3.5.2 Comparison of drug-resistant individuals with healthy controls

In GWAS2, I compared 2,423 cases with DRE and 10,645 healthy controls. Association analysis revealed eight loci with genome-wide significance: 4q31.1, 5p13.2, 22q13.31,

21q22.3, 6p21.33, 11p13, 2q37.3, and 13q31.2. I examined the genomic context, LD structure, and potential functional significance of these loci using information from 1000 Genomes, ENSEMBL Variant Effect Predictor (468, 469), GTEx (470, 471), Brain xQTL (474, 475), Braineac (476, 477), and MGI databases (472, 473). Exploring these resources allowed me to speculate which genes could be plausible candidates at each locus. More advanced methods like the exploration of chromatin interaction data, protein-protein interactions, and transcriptome-wide association analysis (262) were beyond the scope of this work.

All genome-wide significant SNPs were located in non-coding regions of the genome, which is consistent with GWAS findings in other complex diseases. It is estimated that approximately 80% of all GWAS signals are located in introns and intergenic regions, often within promoter or enhancer sequences. Functional variants driving these signals are likely to affect the regulation of gene expression (492, 493).

The genome-wide significant SNPs from GWAS2 can be grouped as follows:

1. Intronic: 11p13
2. Intergenic: 5p13.2
3. Non-coding transcript variant: 13q31.2
4. Associated variants not supported by other SNPs at the same locus, likely representing artefacts: 4q31.1, 6p21.33, 2q37.7, 21q22.3, 22q13.31

3.5.2.1 Signal at locus 11p13

The signal at locus 11p13 was centred on the second intron of the *PAMR1* (peptidase domain containing associated with muscle regeneration 1) gene. *PAMR1* has so far not been linked to epilepsy or neuronal excitation (494). Instead, it has been proposed that *PAMR1* could be a tumour suppressor as it has been found to be frequently inactivated by promoter hypermethylation in breast cancer tissues (495). Even though the lead SNP and all the SNPs in high LD with it are *PAMR1* intronic variants, this does not necessarily mean that *PAMR1* is the biological candidate. There are known examples of intronic variants affecting the expression of adjacent genes (496) and there is a theoretical possibility that another gene within the locus (such as *SLC1A2* or *FJX1*) could be affected. None of the SNPs in the locus is an eQTL for any of the aforementioned genes, based on the data available in public databases (471, 475, 477); however, the available information is limited and may be incomplete.

On the basis of its function and expression pattern, the most interesting gene within this locus is *SLC1A2*, encoding the excitatory amino acid transporter 2 (EAAT2). Located in the membranes of astroglia, EAAT2 is responsible for removing glutamate from the extracellular space at the synapse, allowing timely glutamatergic transmission and maintaining the extracellular glutamate concentration at low levels to protect neurons from its toxic effects. Heterozygous *Slc1a2* loss of function mutations in mice result in impaired glutamate uptake and excitotoxicity (497, 498). De novo mutations in *SLC1A2* cause EE with onset in the first week of life, multiple seizure types, and profound learning disability (102, 392). Lower regional content of EAAT2 has been shown in

dysplastic tissue of individuals with FCD (499). Zhang *et al*. identified *SLC1A2* as a pivotal gene in MTLE using a systems biology approach focussing on gene co-expression networks. They found low expression of *SLC1A2* in individuals with MTLE. In addition, the authors studied the rat pilocarpne model of epilepsy and showed low levels of *Slc1a2* in the hippocampal tissue after SE. Rats transfected with pheochromocytoma cells overexpressing human *SLC1A2* exhibited reduced seizure activity and less pronounced neuronal loss, astrocytosis, and inflammation (500). Combined, these results indicate that genetic variation impacting *SLC1A2* expression could influence disease progression across different epilepsy types by affecting the glutamatergic pathway. Higher expression of EAAT2 could mean more efficient clearance of glutamate from the synapse and less excitotoxicity following different insults (e.g. SE, brain trauma). Existence of variants influencing the regulation of *SLC1A2* expression could be a potential non-specific mechanism underlying drug resistance in common epilepsies.

3.5.2.2 Signal at locus 5p13.2

The signal at locus 5.13 was located in a non-coding region, within 250kb of four genes: *SKP2*, *RANBP3L*, *NADK2*, and *SLC1A3*. Publically available eQTL data do not indicate any involvement of the variants in this locus in the regulation of these genes in the CNS (471, 475, 477). However, the available data may be incomplete and potential influences cannot be excluded.

Based on its function and high expression in the brain, *SLC1A3*, encoding the glial glutamate transporter EAAT1, is the most interesting gene in the region. Since the reuptake of glutamate via astroglial glutamate transporters is the main mechanism for terminating excitatory neurotransmission, EAAT1 is critical for the regulation of extracellular glutamate and excitotoxicity (501), in the same way as the aforementioned EAAT2. In addition to regulating the extent and duration of glutamate-mediated signal, EAAT1 also has anion channel activity preventing further glutamate release (502-504).

Variation in *SLC1A3* (rs4869682 genotype) has been shown to correlate with the risk of developing epilepsy as a consequence of traumatic brain injury (505). Mutations in *SLC1A3* have been linked to seizures in the context of episodic ataxia and alternating hemiplegia (487, 506). Jen *et al.* described a case with a heterozygous mutation of *SLC1A3* and episodic ataxia, focal seizures, migraine, and alternating hemiplegia. The mutation in this individual resulted in a substitution of arginine for proline at a highly conserved site in the EAAT1 protein. The glutamate uptake assay showed markedly reduced glutamate uptake capacity of mutant EAAT1. Insufficient clearance of glutamate from the synapse can lead to to hyperexcitability, manifesting as baseline epileptiform discharges and a lower seizure threshold, as well as other paroxysmal neurologic disturbances, which was clearly demonstrated in this case. The overlapping clinical features resulting from mutations in proteins important in glutamate release and reuptake strengthen a role for glutamate transmission in paroxysmal neurological disorders (487).

CNVs at the 5p13.2 locus result in complex syndromes that can feature seizures as one of the symptoms, often in combination with autistic spectrum disorder (ASD) and attention deficit hyperactivity disorder (ADHD). Chromosome 5p13 duplication syndrome (OMIM #613174) has been described in rare individuals with developmental delay, behavioural problems, and facial dysmorphism. Duplications of 5p13 described up to date vary in size from 0.25 to 13.6 Mb and span a variable number of genes. Clinical presentations are heterogenous and may include epilepsy, ASD, and ADHD (507-510). The main dosage-sensitive gene driving the phenotype is most likely *NIPBL* (nipped-B-like; regulator of the sister chromatid cohesion complex). *SLC1A3,* commony included in 5p13 microduplications, likely contributes to the clinical features. Van Amen-Hellebrekers *et al.* described a series of four children with very short duplications of 5p13.2 encompassing only the *SLC1A3* gene. The most consistent clinical findings were developmental delay, learning disability, facial dysmorphism, behavioural problems resembling ADHD, and autistic features. Interestingly, one of the cases also had EE with onset at four years of age. They concluded that *SLC1A3* might be a risk gene for ASD and ADHD (511). Locus 5p13-q11 has also been linked to ADHD in several other studies. The region contains several plausible candidate genes for ADHD, based on their role in the CNS function and development, *SLC1A3* being one of them (512, 513). Triplications of 5p12-14 region have been associated with epilepsy and ASD (514, 515).

As detailed above, the region on chromosome 5 harbouring the GWAS signal has already been implicated in epilepsy, ASD and ADHD. The most interesting plausible candidate gene within this locus is *SLC1A3*. Similar to *SLC1A2*, variants affecting *SLC1A3*

could be involved in drug resistance across different epilepsy types by affecting the clearance of glutamate, the main excitatory neurotransmitter in the CNS. In addition, connections with ASD and ADHD are particularly interesting because these are common comorbidities in PWE. Co-occurrence of epilepsy and ASD in children is estimated at approximately 20 to 25% (516). The prevalence of ADHD in the epilepsy population ranges from 23 to 40% (517). ASD and ADHD are more likely in children with DRE than with epilepsy that is easy to treat (518, 519).

The signal at locus 5.13 could potentially reflect a causal variant underlying DRE, epilepsy, or merely the higher prevalence of behavioural comorbidities in DRE. Unfortunately, the information on these comorbidities was not systematically collected in EpiPGX. A confirmation of this association signal in a second GWAS is important as the next step. In addition, expression data could help determine which genes are directly affected.

3.5.2.3 Signal at locus 13q31.2

The signal at locus 13q31.2 was centred on the *LINC00433* gene (transcript variant of a long non-coding RNA gene). Long non-coding RNAs (lncRNAs) are a type of transcribed RNA molecules that are more than 200 nucleotides in length and do not encode proteins. They are believed to have a wide array of functions, including gene inhibition and gene activation through various mechanisms. They may regulate genes in close proximity or at a distance from the transcribed lncRNA (520, 521). An example of lncRNA with a known function is the *XIST* (X-inactive specific transcript) lncRNA which

plays an important role in X chromosome inactivation (522). To date, very few lncRNAs have been characterised in detail. No information is available on the function of *LINC00433*. Publically available eQTL data for both associated SNPs at this locus do not indicate a significant effect on the expression of any genes (477). Nevertheless, it is possible that variants affecting *LINC00433* could result in changes in gene regulation underlying epilepsy or DRE. Further expression studies could help determine the function of *LINC00433*.

3.5.2.4 Associated loci representing artefacts

Associated loci that likely represent artefacts include: 4q31.1, 6p21.33, 2q37.3, 21q22.3, and 22q13.3. Based on the examination of GWAS output for SNPs in high LD with the associated SNPs at these loci, there is less confidence that these could be true associations. The cohorts used in this GWAS were genotyped in several separate batches on different platforms and there is a possibility of imputation artefacts. Despite the use of stringent marker QC, issues can arise if SNPs in the studied cohorts differ substantially from the reference. Further work may be required to eliminate these issues, including repeating the analysis in an independent cohort.

3.5.2.5 Limitations

As mentioned in the previous section, the main limitation of this analysis was the use of genotype data generated separately on various platforms which may have resulted

in artefacts despite imputing all data sets using the same reference. As with GWAS1, the sample size, in particular of the DRE cohort, was modest. As shown by GWAS in other areas, increasing the sample size usually reveals additional associated loci (220) and allows the detection of variants carrying small risks. Furthermore, increasing the sample size would also allow for analysis of variants associated with DRE in epilepsy subtypes. It is important to note that the X chromosome was not included in the analysis as the data were not available for all cohorts. Any potentially significant variants on chromosome X could thus not be detected. Lastly, there is a lack of independent replication sample.

I used several public databases to prioritise candidate genes at each locus based on the proximity to the lead SNP, expression patterns, eQTL data, functional consequences of variants within the locus, and knockout mouse phenotype data. Together, these resources have provided only limited insight into the potential relevance of the associated signals rather than allowing definitive conclusions. There are additional methods that could have provided valuable information, for example taking into account chromatin interactions data, protein-protein interactions, and transcriptome-wide association analysis; however, these were beyond the scope of this work. When examining the expression and mouse phenotype data, I focussed predominantly on the central nervous system. Theoretically, genes involved in DRE could be associated with other organ systems, such as the liver, or the immune system.

Assembling large cohorts of epilepsy cases with well characterised long-term outcomes represents a challenge even for large international consortia. The EpiPGX definition of DRE required the availability of sufficient information on individual AED treatment

outcomes and seizure frequency in the medical records. Due to the retrospective data collection, complete information was not always available, resulting in a lower number of cases classified as resistant than what would be expected with less stringent definitions. The DRE cohort available for inclusion in GWAS2 was thus modest. This is discussed in detail in Section 2.5.3, together with other reasons for undefined outcomes. An important part of classifying the cases as drug-resistant was checking for causes of pseudoresistance, for example NES or poor adherence to medication. As this relied on the quality of medical records, it is possible that the exclusion of such cases was not complete. As with GWAS1, the classification of cases with a fluctuating course of epilepsy represents an additional challenge. It has been shown that individuals meeting criteria for DRE may achieve late remissions (401, 406), so it is possible that some individuals in the drug-resistant cohort would have achieved remission with further drug changes. Since GWAS2 included healthy controls this issue might have affected it to a lesser extent than GWAS1 which also included drug responders.

Additional efforts and collaboration with other epilepsy consortia will be required to generate an independent replication cohort.

3.5.2.6 Concluding remarks

GWAS2 revealed several interesting loci associated with the drug-resistant phenotype, most notably 5p13.2 and 11p13 which both harbour potentially biologically relevant genes. This includes the genes for both glial glutamate transporters (*SLC1A2* and *SLC1A3*), responsible for the clearance of glutamate from the synaptic cleft and

protection against glutamate excitotoxicity. In addition, one of these two loci has been linked to ASD and ADHD. These are important comorbidities in epilepsy, more frequently encountered in DRE compared with drug responders. These loci could be risk factors for drug resistance or epilepsy susceptibility (especially difficult-to-treat epilepsy). In addition, the locus on chromosome 5 could reflect common pathophysiological pathways underlying DRE and comorbidities such as ASD and ADHD. All variants identified in GWAS2 were located in non-coding regions. Some of them could potentially affect gene regulation. Expression and functional work would be needed to establish the target genes and the effects of these variants.

3.5.3 Comparison and interpretation of GWAS1 and GWAS2 results

I performed GWAS1 and GWAS2 as two complementary analyses to explore the contribution of common genomic variation to DRE. I used the same DRE cases in both GWAS1 and GWAS2. While GWAS1 did not yield any genome-wide significant results, GWAS2 revealed several interesting associated loci. To get a more complete understanding of the results, I examined the output of GWAS1 for SNPs with genome-wide significance in GWAS2. Similarly, I examined the GWAS2 results for SNPs that had genome-wide significance in the previously published studies of genetic determinants of common epilepsies (261, 262).

3.5.3.1 Comparison of GWAS1 and GWAS2 results

No SNPs reached genome-wide significance in GWAS1 (comparison of drug-resistant cases with drug responders) and eight SNPs reached genome-wide significance in GWAS2 (comparison of drug-resistant cases with drug responders). Four out of these eight SNPs (rs4629621, rs8127410, rs16927514, and rs4516077) showed nominal significance in GWAS1 (see Table 3.5). None of these SNPs or SNPs in high LD with them were close to reaching genome-wide significance in GWAS1.

Table 3.6: P-values from single association tests of genome-wide associated loci from GWAS2 in GWAS1

SNP	Cytogenetic band	Base pair position	Allele 1, allele 2	MAF (minor allele)	P-value (FaSTLMM) GWAS2	OR (95% CI)	P-value (FaSTLMM) GWAS1	OR (95% CI)
rs75700350	4q31.1	140189612	C, T	0.0529 (T)	$3.52*10^{-11}$	1.07 (1.05-1.09)	$8.87*10^{-01}$	1.05 (0.99-1.11)
rs4629621	5p13.2	36426049	T, C	0.1512 (T)	$4{,}69*10^{-11}$	1.06 (0.97-1.15)	$4.73*10^{-03}$	1.05 (0.88-1.25)
rs5765116	22q13.31	45408407	G, A	0.4589 (G)	$2.58*10^{-10}$	1.03 (0.98-1.09)	$5.36*10^{-01}$	1.01 (0.90-1.19)
rs8127410	21q22.3	47507039	A, G	0.3894 (G)	$1.04*10^{-09}$	0.97 (0.91-1.02)	$3.31*10^{-02}$	0.97 (0.86-1.10)
rs150512569	6p21.33	30495053	G, A	0.1032 (A)	$2.38*10^{-09}$	1.04 (0.98-1.11)	$1.39*10^{-01}$	1.02 (0.88-1.19)
rs16927514	11p13	35505321	T, C	0.0280 (C)	$1.52*10^{-08}$	0.94 (0.84-1.05)	$2.32*10^{-02}$	0.95 (0.75-1.20)
rs73999651	2q37.3	238054898	G, T	0.0873 (T)	$2.48*10^{-08}$	0.93 (0.82-1.06)	$6.79*10^{-01}$	0.99 (0.78-1.26)
rs4516077	13q31.2	89193635	A, G	0.4922 (G)	$5.59*10^{-08}$	0.97 (0.93-1.02)	$1.29*10^{-02}$	0.97 (0.87-1.09)

Base pair position refers to human genome build 37 (GRCh37, hg19). MAF is from all populations from the 1000 Genomes Project. Candidate gene refers to the most plausible candidate gene attributable to the signal, based on proximity (450). OR (odds ratio) corresponds to allele 2. Annotation refers to type of SNP; nc transcript variant refers to a transcript variant of a non-coding gene; upstream refers to upstream of a gene.

3.5.3.2 Comparison of ILAE GWAS and GWAS2 results

The ILAE Consortium on Complex Epilepsies was established to explore the genetic susceptibility factors underlying common epilepsies. In 2014, the ILAE Consortium published a meta-analysis combining GWAS data from 12 cohorts of PWE and publically available controls. It included 8,696 cases with epilepsy and 26,157 healthy controls. Analyses were performed separately for three phenotypic groups: all epilepsies, focal epilepsies, and GGE. Analysis of all epilepsies identified two significantly associated loci. The first one was 2q24.3, containing the *SCN1A* gene encoding $Na_v1.1$. The second hit was at 4p15.1, harbouring *PCDH7*. Two additional loci approached genome-wide significance, implicating the golgi integral membrane protein 4 (*GOLIM4*) and GABA type A receptor alpha2 subunit (*GABRA2*) genes (261). Extended ILAE meta-analysis including all epilepsy cases confirmed two previously reported associations at 2p16.1, harbouring *FANCL*, and 2q24.3, harbouring *SCN1A*, and led to the identification of one novel genome-wide significant locus at 16q12.1, harbouring *HEATR3* and *BRD7* (262).

The original and extended ILAE GWAS (all epilepsies analysis) and GWAS2 all compared individuals with epilepsy and healthy controls, the main differences being the selection of epilepsy cases and cohort size. While the ILAE studies included drug-resistant and drug-responsive cases, GWAS2 included only the former. The original and extended ILAE GWAS included 10,064 and 15,212 PWE, respectively, compared with 2,489 individuals with DRE included in GWAS2. It is important to note that four EpiPGX

centres also contributed cases to the ILAE meta-analyses. Consequently, there was partial overlap between the two studies, with 932 out of 2,423 drug-resistant cases from the present analysis also included in the ILAE meta-analyses.

I examined the top ILAE and extended ILAE GWAS SNPs from the analysis of all epilepsies in the GWAS2 output. The comparison of results is shown in Tables 3.6 and 3.7. None of the top SNPs from the ILAE analyses reached genome-wide significance in GWAS2. The lead SNPs at 2q24.3 (rs6732655 and rs6432877) and 16q12.1 (rs4638568) reached nominal significance. The ILAE analyses and GWAS2 all included epilepsy cases and healthy controls; however, the structure of the epilepsy populations in these studies differed significantly. While GWAS2 included only cases with DRE, the proportion of cases with DRE and drug responders in the ILAE sample is unknown as the drug response was not assessed (261, 262). Due to these differences, GWAS2 may not have been positioned to replicate the ILAE results. In addition, the top associated SNPs in the ILAE analyses have low effect sizes, with OR ranging from 0.89 to 1.16 (261). Featuring significantly lower numbers of cases and controls, GWAS2 was not powered to detect SNPs with such low effect sizes. Having said that, the genome-wide associated SNPs in GWAS2 had similarly low OR. GWAS2 could have yielded different results because the signal was driven by difficult-to-treat epilepsy or comorbidities more commonly associated with DRE. These results could support the intrinsic hypothesis of drug resistance.

Table 3.7: Comparison of the genome-wide associated variants identified in the ILAE GWAS study of genetic determinants of common epilepsies (261) and GWAS2

SNP	Cytogenetic band	Base pair position	Candidate gene	Allele 1, allele 2	MAF (minor allele)	ILAE GWAS P-value (LMM)	ILAE GWAS OR	GWAS2 P-value (FaSTLMM)
rs6732655	2q24.3	166895066	*SCN1A*	T, A	0.22 (A)	$8.71*10^{-10}$	0.89	$1.22*10^{-02}$
rs28498976	4p15.1	31151357	*PCDH7*	A, G	0.46 (A)	$5.44*10^{-9}$	0.90	$6.67*10^{-02}$
rs111577701	3q26.2	167861408	*GOLIM4*	T, C	0.09 (T)	$4.42*10^{-07}$	1.16	$7.76*10^{-03}$
rs535066	4p12	46240287	*GABRA2*	T, G	0.40 (G)	$1.71*10^{-07}$	1.10	$1.56*10^{-01}$

Base pair position refers to human genome build 37 (GRCh37, hg19); *GOLIM4* = golgi integral membrane protein 4; *GABRA2* = gamma-aminobutyric acid type A receptor alpha2 subunit. MAF is from all populations from the 1000 Genomes Project. Candidate gene refers to the most plausible candidate gene attributable to the signal. LMM refers to linear mixed-model meta-analysis used in the ILAE GWAS study. OR (odds ratio) corresponds to allele 2. Annotation refers to type of SNP.

Table 3.8: Comparison of the genome-wide associated variants identified in the extended ILAE GWAS study of genetic determinants of common epilepsies (262) and GWAS2

SNP	Cytogenetic band	Base pair position	Candidate gene	Allele 1, allele 2	MAF (minor allele)	ILAE GWAS P-value (LMM)	ILAE GWAS OR	GWAS2 P-value (FaSTLMM)
rs6432877	2q24.3	166998767	*SCN3A, SCN2A, TTC21B, SCN1A*	T, A	0.26 (G)	$1.7*10^{-13}$	NA	$2.82*10^{-03}$
rs4671319	2p16.1	57950346	*FANCL, BCL11A*	A, G	0.44 (G)	$5.44*10^{-9}$	NA	$9.84*10^{-02}$
rs4638568	16q12.1	50045839	*HEATR3, BRD7*	T, C	0.09 (T)	$4.42*10^{-07}$	NA	$8.91*10^{-03}$

Base pair position refers to human genome build 37 (GRCh37, hg19); *SCN1A* = sodium voltage-gated channel type I alpha subunit; *SCN2A* = sodium voltage-gated channel type II alpha subunit; *SCN3A* = sodium voltage-gated channel type III alpha subunit; *TTC21B* = tetratricopeptide repeat domain 21B; *FANCL* = Fanconi anaemia complementation group L; *BCL11A* = BAF chromatin remodeling complex subunit 11A; *HEATR3* = HEAT repeat containing protein 3; *BRD7* = Bromodomain containing protein 7. MAF is from all populations from the 1000 Genomes Project. Candidate gene refers to the most plausible candidate gene attributable to the signal. LMM refers to linear mixed-model meta-analysis used in the ILAE GWAS study. OR (odds ratio) corresponds to allele 2. Annotation refers to type of SNP. NA = not available.

3.5.3.3 Exploration of GWAS1 and GWAS2 results for SNPs in genes previously associated with drug resistance in epilepsy

I examined the results of both GWAS analyses for SNPs in genes previously associated with drug resistance to multiple AEDs, including *CYP2C9*, *CYP2C19*, *ABCB1*, and *SCN1A* (for results, see Appendix 7). None of the SNPs in *CYP2C9*, *CYP2C19*, and *ABCB1* reached genome-wide or nominal significance in any of the analyses. The lack of significance in GWAS2 analysis that compared individuals with DRE and healthy individuals was expected since none of the three genes are considered epilepsy genes. The lack of any sub-threshold signals of interest in GWAS1 comparing individuals with DRE and drug responders might be due to the limited sample size. Alternatively, it is possible that no signal was detected because variation in these genes may only affect the response to a subset of AEDs. In the case of *ABCB1* polymorphisms, published research has produced contradictory conclusions regarding their relationship with DRE. The association of *ABCB1* polymorphisms with AED resistance remains uncertain (14). The results of the present work provide an additional insight into the complex question of *ABCB1* involvement in DRE.

Several SNPs within the *SCN1A* gene reached nominal significance in GWAS2, but not in GWAS1, including rs6732655 previously identified as an epilepsy risk factor in the original ILAE analysis. Although these values are not close to genome-wide significance, they are in line with the published literature linking variation in *SCN1A* with an increased risk of developing common epilepsies (261, 262).

3.6 Conclusions

The main findings of the present GWAS work are:

1. GWAS1 did not show a differential distribution of SNPs in individuals with DRE compared with drug responders. Hypothesis 1 could not be supported. GWAS1 was powered to identify variants with at least moderate effect sizes; the existence of common variants with very small effect sizes or rare variants underlying drug resistance is possible. Factors underlying drug resistance could also be specific to underlying aetiology.

2. GWAS2 showed a differential distribution of SNPs in individuals with DRE compared with healthy individuals. Considering the negative results of GWAS1, Hypothesis 2 can be neither confirmed nor rejected. It is possible that variants identified in GWAS2 are epilepsy susceptibility variants.

3.6.1 Further work

The present GWAS work can be viewed as the pilot to the more extensive and methodologically improved analyses that are going to follow.

The current Consortium efforts focus on generating suitable replication cohorts for both GWAS1 and GWAS2 through collaborations with other epilepsy consortia. Plans

for further GWAS work predict meta-analyses including both the original and the new cohorts. As the Consortium GWAS1 did not identify any variants of genome-wide significance, a meta-analysis approach could help increase the likelihood of relevant findings. GWAS2 identified several interesting loci that require further exploration. Performing a replication analysis will help clarify the significance of the identified loci and hopefully establish whether the less reliable findings were indeed artefacts. Further marker QC measures may be needed to address the issue of possible artefacts. Any confirmed variants will require further work to evaluate their functional significance, for example eQTL analyses.

Generating additional cohorts will also enable the exploration of common genomic variability in epilepsy sub-types, for example in focal epilepsies. In addition, there is ongoing NGS work in DRE exploring the contribution of rare variants.

The expertise within EpiPGX and other international epilepsy consortia could be leveraged to redefine the phenotypes for use in genome-wide studies. As discussed in Section 2.5.3.5, the DRE definitions used in research do not address the issue of fluctuating epilepsy course in some PWE, resulting in classification issues and reduced power. One potential solution currently explored within EpiPGX is to compare 'extreme non-responders' with 'super-responders'. The former could be defined as individuals failing a minimum of five to six AEDs without ever achieving a 12-month remission. This is based on the studies showing that the likelihood of remission after this number of AED trials is extremely low (49, 78, 79). 'Super-responders' could be defined as individuals achieving sustained response with the first adequate and appropriately used AED. A consensus would be needed on the minimum required follow-up

period to apply such a definition. Another potential option that could be explored is using quantitative rather than dichotomous phenotypes (for example proportion of time spent in remission, number of failed AEDs, etc.). Regardless of the definitions, prospective data collection would improve the data quality and the outcomes. Large international consortia have the potential to generate sufficiently large and adequately phenotyped prospective cohorts.

The genotype data generated for the purposes of the present work have been shared with other consortia in the field of epilepsy and beyond to contribute to various research efforts. A proportion of the GWAS data has been included in the Brainstorm Consortium meta-analysis that explored the shared genetics of 25 common brain disorders in 265,218 affected individuals and 784,643 controls, as well as their relationship to 17 phenotypes from 1,191,588 individuals. The analysis quantified the overlapping genetic risk factors between these disorders (523). Furthermore, the data will contribute to other cross-consortia efforts combining multi-layered data sets (for example the work by the ENIGMA Consortium featuring genomic and imaging data (524)).

CHAPTER 4: HERITABILITY ANALYSIS

4.1 Introduction

4.1.1 The principles of SNP heritability analysis

The purpose of SNP heritability analysis is to estimate how much of the phenotypic variance of a trait can be explained by all the measured SNPs in a data set. This is fundamentally different to GWAS discovery where millions of SNPs are tested individually (216).

In a SNP heritability analysis, genetic similarity across a large number of SNPs is correlated with phenotypic similarity in a cohort of unrelated individuals, each compared pairwisewith every other individual in the sample (216). The amount of genomic variation shared by conventionally unrelated individuals is small and highly variable. However, these small genotype-phenotype association signals are accumulated using a large amount of information available in a matrix of thousands of individuals. If randomly selected individuals with similar phenotypic features are relatively similar genetically, this indicates that the assessed SNPs (capturing the contribution from causal variants in LD with these SNPs) influence the trait (209).

The statistical test behind the SNP heritability analysis is linear regression using a mixed model. Mixed model is implemented for the partitioning of the phenotypic variance into the variance explained by one or more fixed effects such as individual SNPs or other covariates and the residual variance due to an unobserved random effect, usually interpreted as the polygenic contribution to the disease risk (525, 526). The partitioning is achieved by comparing prediction error against relatedness (526). Inter-individual phenotype differences are used as the primary variable, and genomic relatedness as the dependent variable, modelled as a combination of fixed effects (such as genome-wide associated SNPs identified in GWAS analyses, age, and sex) and a random effect using restricted maximum likelihood method; REML. This type of analysis is referred to as genomic relatedness matrix REML (GREML) (209).

The two software packages implementing GREML are Genome-wide Complex Trait Analysis (GCTA) and LD-Adjusted Kinships (LDAK) (525). Both of them use a two-step procedure to perform SNP heritability estimation:

1. First, a kinship matrix, also called genetic relationships matrix (GRM) is estimated from genome-wide SNP data. Each element in the matrix represents the genetic similarity of two individuals. The main difference between the two packages is that LDAK calculates a modified kinship matrix in which SNPs are weighted according to local LD to account for potential biases due to LD. These biases may arise since the contributions of causal variants to h2 may be overestimated in chromosomal regions with strong LD and underestimated in those with low LD (525).

2. The second step is fitting the kinship matrix in a mixed linear model to estimate how much of the phenotypic variance is attributable to genomic variation via the REML method. This is achieved by comparing prediction error against relatedness (527).

4.1.2 QC procedures in SNP heritability analysis

As is the case with GWAS, the results of SNP heritability analyses depend heavily on the overall data quality. The QC procedures in SNP heritability analyses are similar to those in GWAS and aim at excluding suboptimal quality markers and samples.

4.1.2.1 Marker QC procedures

Marker QC procedures include removing SNPs with excessive missing genotype data, SNPs with a significant deviation from HWE, SNPs with significantly different missingness rates in cases and controls, and SNPs with a very low MAF (422). Details are provided in Section 3.3.3.

4.1.2.2 Sample QC procedures

Sample QC procedures include removing samples with potential identity problems, samples with excessive missing genotype data, heterozygosity, population

stratification, or relatedness issues. Procedures are similar to those in GWAS (see Section 3.1.1.3.2 for details); however, controlling for relatedness may be more stringent in SNP heritability analyses.

When estimating SNP heritability, it is very important to exclude close relatives. If close relatives are part of the analysis, the phenotypic correlations between them can drive the estimate of genetic variance, potentially causing bias, for example due to common environmental effects. Both GCTA and LDAK have a function to iteratively exclude one sample out of each pair of individuals with genomic relatedness that is higher than a pre-specified cutoff value, e.g. IBD > 0.05. The function ensures that exclusion is done in a way to keep the maximum possible number of samples in the analysis (526).

4.1.2.3 Covariates

SNP heritability analyses are sensitive to potential biases due to population structure. Confounding can occur because individuals who are genetically similar are generally also geographically close and thus live in a similar environment. Consequently, it is difficult to ascertain whether phenotypic similarity is caused by shared genetic factors or by shared environment (528). Thus, it is common practice to include principal components as covariates in the mixed linear model to capture variance due to population structure (526).

4.2 Aim

The purpose of the present SNP heritability analysis was to estimate how much of the phenotypic variance of drug resistance in epilepsy can be explained by all the measured SNPs in the data set. As outlined in Section 1.10, the SNP heritability analysis tests the following hypothesis:

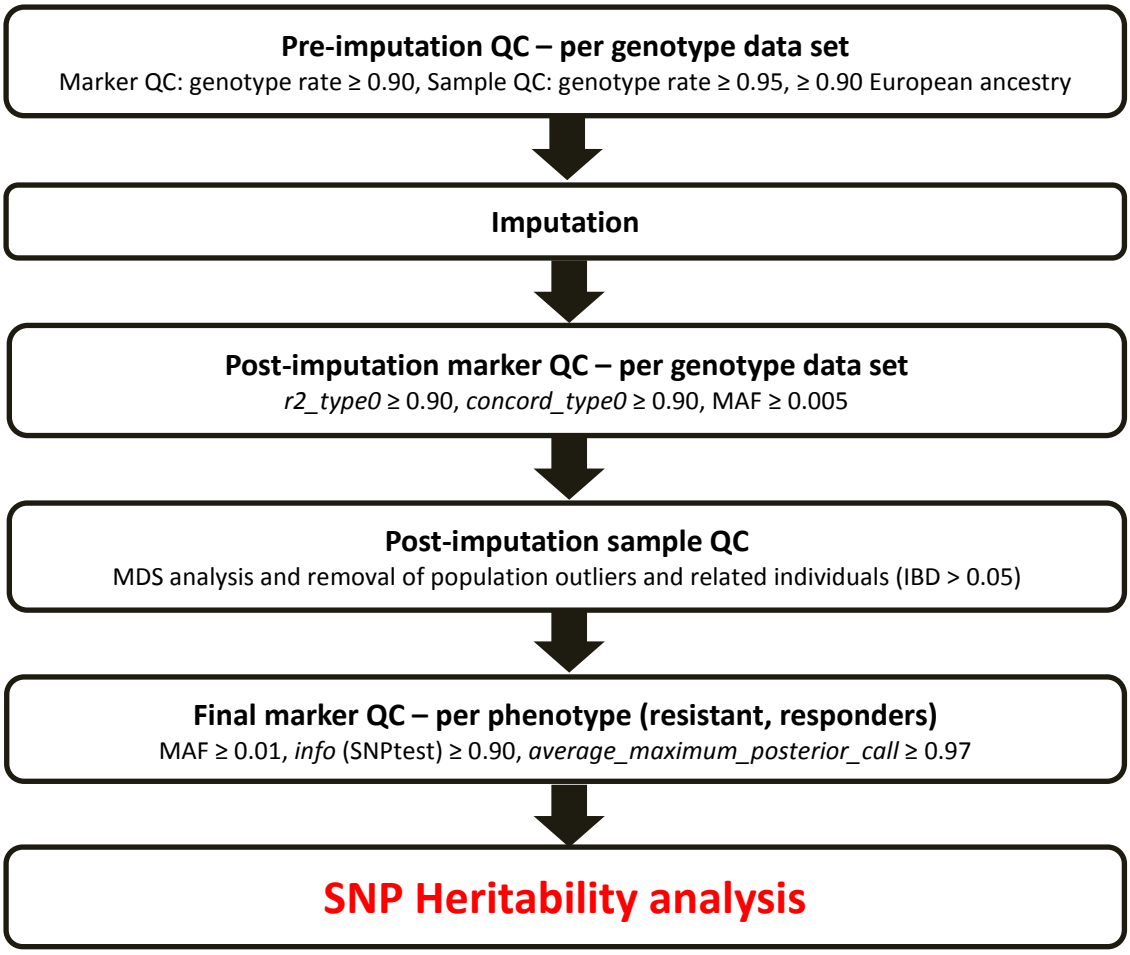
Hypothesis 3:

Drug resistance in epilepsy has a polygenic inheritance component.

4.3 Methods

I performed the heritability analysis using the same genotype data as in GWAS1 (comparing individuals with DRE and drug responders). I followed the protocol outlined in Figure 4.1. I used LDAK for the analysis (525).

Figure 4.1: Heritability analysis protocol



QC = quality control, MAF = minorallele frequency, HWE = Hardy-Weinberg equilibrium, MDS = multidimensional scaling. IMPUTE metrics include *r2_type0* and *concord_type0*. SNPTEST metrics include *average_maximum_posterior_call* and *info.*

4.3.1 Peripheral blood collection, DNA extraction, SNP genotyping, and imputation

The methods for peripheral blood collection, DNA extraction, SNP genotyping, and imputation are described in Sections 3.3.1 and 3.3.2.

4.3.2 Sample and marker QC procedures

I performed marker and sample QC as for GWAS1 (See Section 3.3.3), with the exception of controlling for relatedness where I applied a more stringent threshold than in GWAS1. I created a subset of markers independent of each other with respect to LD using a window size of 150 markers, shifting by 10 markers at a time and removing one half of every SNP pair with genotypic r2 > 0.4 (PLINK command: *--indep-pairwise 150 10 0.4*). Using this subset of markers, I calculated IBS/IBD (PLINK command: *--genome*). I then pruned one out of each pair of individuals with > 0.05 relatedness within each genotype data set and across the entire sample.

4.3.3 Calculating population axes

Using the genotype data for the subset of independent markers created with PLINK as described above (after removing SNPs in LD r2 > 0.4), I constructed a kinship matrix and performed PCA (LDAK commands *--calc-kins-direct, --pca*). This way I obtained the principal components (referred to as the population axes).

4.3.4 Calculating SNP weights and computing kinships

I calculated LD-based SNP weights for all autosomal markers that passed QC (LDAK command *--calc-weights*). I then computed LD-weighted kinships (LDAK command *--calc-kins-direct*) and created a GRM (LDAK command *--add-grm*).

4.3.5 Estimating SNP heritability

I performed h2 SNP estimation (LDAK command *--reml*) using the following parameters:

- Prevalence (LDAK command *--prevalence*) 0.3 (proportion of people with DRE in the epilepsy population)
- Covariates (LDAK command *--covar*): gender, epilepsy type (focal, generalised, unclassified), the first ten components (principal axes) of PCA

It is common to include SNPs that are associated with the trait in question in GWAS as fixed-effect covariates in heritability analysis (529). As no SNPs reached genome-wide significance in GWAS1 (comparing individuals with DRE and drug responders), no individual SNPs were included as covariates.

4.4 Results

4.4.1 Genotyping and imputation

Genotyping and imputation generated approximately 40,000,000 SNPs per case overall before QC (for details see Section 3.4.1).

4.4.2 SNP heritability analysis

I included 2,489 cases with DRE and 1,626 cases with drug-responsive epilepsy. After application of sample QC criteria, 2,254 cases with DRE and 1,501 cases with drug-responsive epilepsy remained in the analysis. After application of marker QC, 5,919,078 SNPs were taken forward in the h^2_{SNP} analysis. The h^2_{SNP} for DRE was estimated at 0.22 (SD 0.17). This estimate represents the proportion of inter-individual variance in resistance to AEDs in PWE accounted for by common SNPs in LD with causal genomic variants that underlie drug resistance.

4.5 Discussion

4.5.1 Interpreting SNP heritability results in absence of pedigree studies

As discussed in Section 1.8.2, the heritability (i.e. extent of genetic contribution) of drug resistance in epilepsy has not been quantified yet. The GWAS approach used in the present work has provided limited insights into the contribution of common SNPs to DRE. When comparing individuals with DRE and drug responders in GWAS1, I did not identify any genome-wide associated common SNPs. In the present SNP heritability analysis, I explored the joint contribution of all tested SNPs to the drug resistance phenotype.

The resulting h^2_{SNP} estimate of 22% provides the most convincing evidence for the existence of a polygenic heritable component of DRE to date. The present estimate is free from confounders arising from shared environment. In absence of conclusive pedigree studies, h^2 of drug resistance in epilepsy is unknown. Consequently, I am unable to compare the h^2_{SNP} estimate obtained in this work with existing h^2 estimates.

The relationship between h^2_{SNP} and h^2 for several traits, including height, has been explored in large GWAS data sets using GCTA and simulation analyses (216, 220, 530). h^2_{SNP} estimates are expected to be lower than h^2. The gap is observed because it is difficult to capture the additive effects of rare SNPs from GWAS data and because of statistical "noise" arising from small sample sizes. In addition, h^2 can be overestimated

due to the confounding resulting from shared family environment effects (530). For height, h^2_{SNP} estimates from more recent studies with bigger sample sizes and using dense genotyping arrays approach h^2 (220). Although the data from the present heritability work do not allow me to estimate the size of the gap between h^2_{SNP} and h^2, it is fair to assume that there is a gap between the h^2_{SNP} and h^2 of drug resistance in epilepsy. Nevertheless, this result indicates that the contribution of common SNPs captured on the genotyping platforms is limited. Other genetic contributions may arise from rare SNPs, CNVs, non-additive genetic effects, and environmental effects.

4.5.2 Limitations

As with GWAS1, the main limitations of this SNP heritability analysis are the relatively small sample size, under-representation of drug respoders in the sample, and potential issues with classification of individuals with a fluctuating course of epilepsy. This is discussed in more detail in Section 3.5.1.1.

The resulting h^2_{SNP} estimate is limited by the sample size, as well as the capacity of genotyping platforms and imputation methods. Using dense genotyping arrays or NGS data including rare variants and CNVs in larger samples sizes would likely result in higher h^2_{SNP} estimates and potentially predict individuals at risk of DRE. The addition of gene-gene and gene-environment interaction effects may increase the h^2_{SNP} estimate and allow better prediction of risk (530).

4.6 Conclusions

To my knowledge, the present work provides the first estimate of the SNP heritability of DRE to date. The estimate of 22% is modest, but sufficient to conclude that DRE has a polygenic inheritance component, confirming Hypothesis 3. This estimate reflects only SNPs tested in this work, but it does not account for the potential contributions of rare variants or CNVs.

Further work includes refining the SNP heritability estimate by using larger cohorts and dense genotyping arrays or NGS. Performing SNP heritability analyses for epilepsy sub-types may provide additional relevant insights.

CHAPTER 5: CONCLUSIONS

Over 25 AEDs with different mechanisms of action are currently available for the treatment of epilepsy, yet in individuals with DRE, resistance occurs to many or all of them. In some cases, AEDs may not work because they do not target the underlying pathophysiology (e.g. carbamazepine in Dravet syndrome), in line with the intrinsic hypothesis of drug resistance. The existence of such examples, however, does not exclude the possibility of a cross-syndrome component of DRE. Drug resistance is observed clinically across the entire spectrum of AEDs and epilepsies, regardless of the underlying aetiology, implying that it could be driven by a non-specific mechanism.

Various research approaches have been used to determine the underlying causes of drug resistance over the past decades. Clinical predictors of DRE have been known for decades; however, the genetic factors underlying DRE have proven difficult to identify. There are no comprehensive pedigree studies of DRE heritability. Numerous candidate gene studies have been undertaken, but no convincing genetic cause has been identified yet.

In the present work, I explored the genetic contribution to DRE and attempted to identify common genomic variants associated with drug resistance in epilepsy of any aetiology using a large data set of human SNP genotype data. In GWAS1, I searched for genome-wide significant common SNPs underlying drug resistance by comparing individuals with DRE and drug-responders. This study did not identify any common

SNPs associated with DRE; however, this analysis is not definitive and the existence of common variants with very small effect sizes or rare variants underlying drug resistance is possible. In GWAS2, I compared individuals with DRE and healthy controls. This study identified several interesting signals which could potentially be driven by epilepsy (in particular difficult-to-treat epilepsy) susceptibility variants or variants underlying comorbidities.

I performed a SNP heritability analysis to explore the joint contribution of all tested SNPs to drug resistance, showing that DRE has a polygenic heritable component (22%). This value is modest, indicating other influences, such as that of the environment. Considering that the GWAS comparing individuals with DRE and individuals with drug-responsive epilepsy did not identify any genome-wide significant variants, but the heritability analysis using the same data showed a combined effect of all SNPs, it is possible that drug resistance is caused by a large number of variants with small effect sizes, in line with the infinitesimal model. Showing that DRE has a heritable component is important as it provides a rationale for further genome-wide studies. Uncovering the genetic architecture of drug resistance in epilepsy has the potential to guide the development of novel therapies and personalised medicine.

BIBLIOGRAPHY

1. Baker GA, Camfield C, Camfield P, Cramer JA, Elger CE, Johnson AL, et al. Commission on Outcome Measurement in Epilepsy, 1994-1997: final report. Epilepsia. 1998;39(2):213-31.

2. Bonnett L, Smith CT, Smith D, Williamson P, Chadwick D, Marson AG. Prognostic factors for time to treatment failure and time to 12 months of remission for patients with focal epilepsy: post-hoc, subgroup analyses of data from the SANAD trial. Lancet Neurol. 2012;11(4):331-40.

3. Noble AJ, Marson AG. Which outcomes should we measure in adult epilepsy trials? The views of people with epilepsy and informal carers. Epilepsy Behav. 2016;59:105-10.

4. Gaitatzis A, Sander JW. The mortality of epilepsy revisited. Epileptic Disord. 2004;6:3-13.

5. Trinka E. Epilepsy: comorbidity in the elderly. Acta Neurol Scand. 2003;108(s180):33-6.

6. Kwan P, Brodie MJ. Neuropsychological effects of epilepsy and antiepileptic drugs. Lancet. 2001;357(9251):216-22.

7. McCagh J, Fisk JE, Baker GA. Epilepsy, psychosocial and cognitive functioning. Epilepsy Res. 2009;86(1):1-14.

8. Spatt J, Bauer G, Baumgartner C, Feucht M, Graf M, Mamoli B, et al. Predictors for Negative Attitudes toward Subjects with Epilepsy: A Representative Survey in the General Public in Austria. Epilepsia. 2005;46(5):736-42.

9. Keezer MR, Bell GS, Neligan A, Novy J, Sander JW. Cause of death and predictors of mortality in a community-based cohort of people with epilepsy. Neurology. 2016;86(8):704-12.

10. Lhatoo SD, Johnson AL, Goodridge DM, MacDonald BK, Sander JW, Shorvon SD. Mortality in epilepsy in the first 11 to 14 years after diagnosis: Multivariate analysis of a long-term, prospective, population-based cohort. Ann Neurol. 2001;49(3):336-44.

11. Mohanraj R, Norrie J, Stephen LJ, Kelly K, Hitiris N, Brodie MJ. Mortality in adults with newly diagnosed and chronic epilepsy: a retrospective comparative study. Lancet Neurol. 2006;5(6):481-7.

12. Sheth SG, Krauss G, Krumholz A, Li G. Mortality in epilepsy: driving fatalities vs other causes of death in patients with epilepsy. Neurology. 2004;63(6):1002-7.

13. Trinka E, Bauer G, Oberaigner W, Ndayisaba JP, Seppi K, Granbichler CA. Cause-specific mortality among patients with epilepsy: Results from a 30-year cohort study. Epilepsia. 2013;54(3):495-501.

14. Balestrini S, Sisodiya SM. Pharmacogenomics in epilepsy. Neurosci Lett. 2017;667:27-39.

15. Franco V, Perucca E. The pharmacogenomics of epilepsy. Expert Rev Neurother. 2015;15(10):1161-70.

16. Aneesh TP, Sonal Sekhar M, Jose A, Chandran L, Zachariah SM. Pharmacogenomics: The Right Drug to the Right Person. J Clin Med Res. 2009;1(4):191-4.

17. Kwan P, Brodie MJ. Early identification of refractory epilepsy. N Engl J Med. 2000;342(5):314-9.

18. Wieser HG, Ortega M, Friedman A, Yonekawa Y. Long-term seizure outcomes following amygdalohippocampectomy. J Neurosurg. 2003;98(4):751-63.

19. Roden DM, Wilke RA, Kroemer HK, Stein CM. Pharmacogenomics: the genetics of variable drug responses. Circulation. 2011;123(15):1661-70.

20. Arts WF, Geerts AT, Brouwer OF, Boudewyn Peters AC, Stroink H, van Donselaar CA. The early prognosis of epilepsy in childhood: the prediction of a poor outcome. The Dutch study of epilepsy in childhood. Epilepsia. 1999;40(6):726-34.

21. Berg AT, Shinnar S, Levy SR, Testa FM, Smith-Rapaport S, Beckerman B. Early development of intractable epilepsy in children: a prospective study. Neurology. 2001;56(11):1445-52.

22. Camfield P, Camfield C. Nova Scotia pediatric epilepsy study. In: Jallon PB, A.; Dulac, O.; Hauser, W. A., editor. Prognosis of Epilepsies. Montrogue, France: John Libbey Eurotext; 2003. p. 113-26.

23. Dlugos D, Sammel MD, Strom BL, Farrar JT. Response to first drug trial predicts outcome in childhood temporal lobe epilepsy. Neurology. 2001;57(12):2259-64.

24. Berg AT, Levy SR, Novotny EJ, Shinnar S. Predictors of Intractable Epilepsy in Childhood: A Case-Control Study. Epilepsia. 1996;37(1):24-30.

25. Kwan P, Hao XT. Update and overview of the International League Against Epilepsy consensus definition of drug-resistant epilepsy. US Neurology. 2010;6(2):122-4.

26. Kwan P, Arzimanoglou A, Berg AT, Brodie MJ, Allen Hauser W, Mathern G, et al. Definition of drug resistant epilepsy: consensus proposal by the ad hoc Task Force of the ILAE Commission on Therapeutic Strategies. Epilepsia. 2010;51(6):1069-77.

27. Hanley JA, Lippman-Hand A. If nothing goes wrong, is everything all right? Interpreting zero numerators. JAMA. 1983;249(13):1743-5.

28. Brodie MJ, Barry SJ, Bamagous GA, Norrie JD, Kwan P. Patterns of treatment response in newly diagnosed epilepsy. Neurology. 2012;78(20):1548-54.

29. Geerts A, Arts WF, Stroink H, Peeters E, Brouwer O, Peters B, et al. Course and outcome of childhood epilepsy: a 15-year follow-up of the Dutch Study of Epilepsy in Childhood. Epilepsia. 2010;51(7):1189-97.

30. Bell GS, Neligan A, Sander JW. An unknown quantity—The worldwide prevalence of epilepsy. Epilepsia. 2014;55(7):958-62.

31. Fiest KM, Sauro KM, Wiebe S, Patten SB, Kwon C-S, Dykeman J, et al. Prevalence and incidence of epilepsy: A systematic review and meta-analysis of international studies. Neurology. 2017;88(3):296-303.

32. Ngugi AK, Bottomley C, Kleinschmidt I, Sander JW, Newton CR. Estimation of the burden of active and life-time epilepsy: A meta-analytic approach. Epilepsia. 2010;51(5):883-90.

33. Commission on Epidemiology and Prognosis ILAE. Guidelines for Epidemiologic Studies on Epilepsy. Epilepsia. 1993;34(4):592-6.

34. Thurman DJ, Beghi E, Begley CE, Berg AT, Buchhalter JR, Ding D, et al. Standards for epidemiologic studies and surveillance of epilepsy. Epilepsia. 2011;52(s7):2-26.

35. Ngugi AK, Kariuki SM, Bottomley C, Kleinschmidt I, Sander JW, Newton CR. Incidence of epilepsy: a systematic review and meta-analysis. Neurology. 2011;77(10):1005-12.

36. Camfield C, Camfield P, Gordon K, Smith B, Dooley J. Outcome of childhood epilepsy: A population-based study with a simple predictive scoring system for those treated with medication. J Pediatr. 1993;122(6):861-8.

37. Sillanpää M, Schmidt D. Natural history of treated childhood-onset epilepsy: prospective, long-term population-based study. Brain. 2006;129(Pt 3):617-24.

38. Shorvon SD, Goodridge DMG. Longitudinal cohort studies of the prognosis of epilepsy: contribution of the National General Practice Study of Epilepsy and other studies. Brain. 2013;136(11):3497-510.

39. Neligan A, Sander JW. The prognosis of epilepsy. In: Rugg-Gunn FJ, Smalls JE, editors. Epilepsy 2015, From Channels to Commissioning, A Practical Guide to Epilepsy. East Sussex: Meritus Communications; 2015. p. 373-8.

40. Chen Z, Brodie MJ, Liew D, Kwan P. Treatment Outcomes in Patients With Newly Diagnosed Epilepsy Treated With Established and New Antiepileptic Drugs: A 30-Year Longitudinal Cohort Study. JAMA Neurol. 2018;75(3):279-86.

41. Cockerell OC, Johnson AL, Sander JW, Shorvon SD. Prognosis of epilepsy: a review and further analysis of the first nine years of the British National General Practice Study of Epilepsy, a prospective population-based study. Epilepsia. 1997;38(1):31-46.

42. Hanaoka Y, Yoshinaga H, Kobayashi K. A ten-year follow-up cohort study of childhood epilepsy: Changes in epilepsy diagnosis with age. Brain Dev. 2017;39(4):312-20.

43. Mohanraj R, Brodie MJ. Diagnosing refractory epilepsy: response to sequential treatment schedules. Eur J Neurol. 2006;13(3):277-82.

44. Annegers JF, Hauser WA, Elveback LR. Remission of seizures and relapse in patients with epilepsy. Epilepsia. 1979;20(6):729-37.

45. Bell GS, Neligan A, Giavasi C, Keezer MR, Novy J, Peacock JL, et al. Outcome of seizures in the general population after 25 years: a prospective follow-up, observational cohort study. J Neurol Neurosurg Psychiatry. 2016;87(8):843-50.

46. Wakamoto H, Nagao H, Hayashi M, Morimoto T. Long-term medical, educational, and social prognoses of childhood-onset epilepsy: a population-based study in a rural district of Japan. Brain Dev. 2000;22(4):246-55.

47. Giussani G, Canelli V, Bianchi E, Franchi C, Nobili A, Erba G, et al. A population-based study of active and drug-resistant epilepsies in Northern Italy. Epilepsy Behav. 2016;55:30-7.

48. Lindsten H, Stenlund H, Forsgren L. Remission of seizures in a population-based adult cohort with a newly diagnosed unprovoked epileptic seizure. Epilepsia. 2001;42(8):1025-30.

49. Zhang Y, Yu N, Su L, Di Q. A prospective cohort study of prognosis for newly diagnosed epilepsy in east China. BMC Neurol. 2013;13:116.

50. Picot MC, Baldy-Moulinier M, Daures JP, Dujols P, Crespel A. The prevalence of epilepsy and pharmacoresistant epilepsy in adults: a population-based study in a Western European country. Epilepsia. 2008;49(7):1230-8.

51. Berg AT, Zelko FA, Levy SR, Testa FM. Age at onset of epilepsy, pharmacoresistance, and cognitive outcomes: a prospective cohort study. Neurology. 2012;79(13):1384-91.

52. Berg AT, Kelly MM. Defining intractability: comparisons among published definitions. Epilepsia. 2006;47(2):431-6.

53. Kong ST, Ho CS, Ho PC, Lim SH. Prevalence of drug resistant epilepsy in adults with epilepsy attending a neurology clinic of a tertiary referral hospital in Singapore. Epilepsy Res. 2014;108(7):1253-62.

54. Ramos-Lizana J, Rodriguez-Lucenilla MI, Aguilera-López P, Aguirre-Rodríguez J, Cassinello-García E. A study of drug-resistant childhood epilepsy testing the new ILAE criteria. Seizure. 2012;21(4):266-72.

55. Kobow K, Blumcke I. Epigenetics in epilepsy. Neurosci Lett. 2017;667:40-6.

56. Mohanraj R, Brodie MJ. Early predictors of outcome in newly diagnosed epilepsy. Seizure. 2013;22(5):333-44.

57. Semah F, Picot MC, Adam C, Broglin D, Arzimanoglou A, Bazin B, et al. Is the underlying cause of epilepsy a major prognostic factor for recurrence? Neurology. 1998;51(5):1256-62.

58. Sillanpää M, Jalava M, Kaleva O, Shinnar S. Long-term prognosis of seizures with onset in childhood. N Engl J Med. 1998;338(24):1715-22.

59. Sillanpää M, Schmidt D. Early seizure frequency and aetiology predict long-term medical outcome in childhood-onset epilepsy. Brain. 2009;132(Pt 4):989-98.

60. Sillanpää M, Anttinen A, Rinne JO, Joutsa J, Sonninen P, Erkinjuntti M, et al. Childhood-onset epilepsy five decades later. A prospective population-based cohort study. Epilepsia. 2015;56(11):1774-83.

61. Okuma T, Kumashiro H. Natural history and prognosis of epilepsy: report of a multi-institutional study in Japan. The group for the study of prognosis of epilepsy in Japan. Epilepsia. 1981;22(1):35-53.

62. Russo A, Posar A, Conti S, Parmeggiani A. Prognostic factors of drug-resistant epilepsy in childhood: An Italian study. Pediatr Int. 2015;57(6):1143-8.

63. Trinka E, Martin F, Luef G, Unterberger I, Bauer G. Chronic epilepsy with complex partial seizures is not always medically intractable--a long-term observational study. Acta Neurol Scand. 2001;103(4):219-25.

64. Stephen LJ, Kwan P, Brodie MJ. Does the Cause of Localisation-Related Epilepsy Influence the Response to Antiepileptic DrugTreatment? Epilepsia. 2001;42(3):357-62.

65. Mattson RH, Cramer JA, Collins JF. Prognosis for total control of complex partial and secondarily generalized tonic clonic seizures. Department of Veterans Affairs Epilepsy Cooperative Studies No. 118 and No. 264 Group. Neurology. 1996;47(1):68-76.

66. MacDonald BK, Johnson AL, Goodridge DM, Cockerell OC, Sander JW, Shorvon SD. Factors predicting prognosis of epilepsy after presentation with seizures. Ann Neurol. 2000;48(6):833-41.

67. Goodridge DM, Shorvon SD. Epileptic seizures in a population of 6000. II: Treatment and prognosis. Br Med J (Clin Res Ed). 1983;287(6393):645-7.

68. Beghi E, Tognoni G. Prognosis of Epilepsy in Newly Referred Patients: A Multicenter Prospective Study. Epilepsia. 1988;29(3):236-43.

69. Collaborative Group for the Study of Epilepsy. Prognosis of epilepsy in newly referred patients: a multicenter prospective study of the effects of monotherapy on the long-term course of epilepsy. Epilepsia. 1992;33(1):45-51.

70. Sillanpää M, Schmidt D. Seizure clustering during drug treatment affects seizure outcome and mortality of childhood-onset epilepsy. Brain. 2008;131(Pt 4):938-44.

71. Hitiris N, Mohanraj R, Norrie J, Sills GJ, Brodie MJ. Predictors of pharmacoresistant epilepsy. Epilepsy Res. 2007;75(2):192-6.

72. Feksi AT, Kaamugisha J, Sander JWAS, Shorvon SD, Gatiti S. Comprehensive primary health care antiepileptic drug treatment programme in rural and semi-urban Kenya. Lancet. 1991;337(8738):406-9.

73. Shorvon SD, Reynolds EH. Early prognosis of epilepsy. Br Med J (Clin Res Ed). 1982;285(6356):1699-701.

74. Kwan P. Defining drug-resistant epilepsy. Neurology Asia. 2011;16:67-9.

75. Gowers WR. Epilepsy and other chronic convulsive diseases: their causes, symptoms, & treatment. London: Churchill; 1881.

76. Sillanpää M. Remission of Seizures and Predictors of Intractability in Long-Term Follow-Up. Epilepsia. 1993;34(5):930-6.

77. Camfield C, Camfield P, Gordon K, Dooley J. Does the number of seizures before treatment influence ease of control or remission of childhood epilepsy? Not if the number is 10 or less. Neurology. 1996;46(1):41-4.

78. Schiller Y, Najjar Y. Quantifying the response to antiepileptic drugs. Effect of past treatment history. Neurology. 2008;70(1):54-65.

79. Callaghan B, Schlesinger M, Rodemer W, Pollard J, Hesdorffer D, Allen Hauser W, et al. Remission and relapse in a drug-resistant epilepsy population followed prospectively. Epilepsia. 2011;52(3):619-26.

80. Casetta I, Granieri E, Monetti VC, Gilli G, Tole MR, Paolino E, et al. Early predictors of intractability in childhood epilepsy: a community-based case–control study in Copparo, Italy. Acta Neurol Scand. 1999;99(6):329-33.

81. Ko TS, Holmes GL. EEG and clinical predictors of medically intractable childhood epilepsy. Clin Neurophysiol. 1999;110(7):1245-51.

82. Sillanpää M, Schmidt D. Is incident drug-resistance of childhood-onset epilepsy reversible? A long-term follow-up study. Brain. 2012;135(Pt 7):2256-62.

83. Hauser E, Freilinger M, Seidl R, Groh C. Prognosis of Childhood Epilepsy in Newly Referred Patients. J Child Neurol. 1996;11(3):201-4.

84. Kwan P, Brodie MJ. Drug Treatment of Epilepsy: When Does It Fail and How to Optimize Its Use? CNS Spectrums. 2004;9(2):110-9.

85. Rogawski MA, Johnson MR. Intrinsic severity as a determinant of antiepileptic drug refractoriness. Epilepsy Curr. 2008;8(5):127-30.

86. Sills GJ. Mechanisms of action of antiepileptic drugs. In: Rugg-Gunn FJ, Smalls JE, editors. Epilepsy 2015, From Channels to Commissioning, A Practical Guide to Epilepsy. East Sussex: Meritus Communications; 2015. p. 373-8.

87. Brodie MJ, Sills GJ. Combining antiepileptic drugs--rational polytherapy? Seizure. 2011;20(5):369-75.

88. Pohlmann-Eden B, Weaver DF. The puzzle(s) of pharmacoresistant epilepsy. Epilepsia. 2013;54(s2):1-4.

89. Faingold CL. Anticonvulsant Drugs Are Neuronal Network-Modifying Agents (NMAs). In: Stein J, Coen C, Rolls ET, Mann EO, Goodwin G, Morris J, et al., editors. Reference Module in Neuroscience and Biobehavioral Psychology: Elsevier; 2017.

90. White HS. Preclinical Development of Antiepileptic Drugs: Past, Present, and Future Directions. Epilepsia. 2003;44(s7):2-8.

91. Klitgaard H. Levetiracetam: The Preclinical Profile of a New Class of Antiepileptic Drugs? Epilepsia. 2001;42(s4):13-8.

92. Sisodiya SM, Marini C. Genetics of antiepileptic drug resistance. Curr Opin Neurol. 2009;22(2):150-6.

93. Potschka H. Animal and human data: where are our concepts for drug-resistant epilepsy going? Epilepsia. 2013;54 Suppl 2:29-32.

94. Remy S, Beck H. Molecular and cellular mechanisms of pharmacoresistance in epilepsy. Brain. 2006;129(1):18-35.

95. Löscher W. How to Explain Multidrug Resistance in Epilepsy? Epilepsy Curr. 2005;5(3):107-12.

96. Rogawski MA. The intrinsic severity hypothesis of pharmacoresistance to antiepileptic drugs. Epilepsia. 2013;54 Suppl 2:33-40.

97. Martínez-Juárez IE, Hernández-Vanegas LE, Rodríguez y Rodríguez N, León-Aldana JA, Delgado-Escueta AV. Genes Involved in Pharmacoresistant Epilepsy. In: Rocha L, Cavalheiro EA, editors. Pharmacoresistance in Epilepsy: From Genes and Molecules to Promising Therapies: Springer; 2013. p. 11-25.

98. Löscher W, Sills GJ. Drug resistance in epilepsy: Why is a simple explanation not enough? Epilepsia. 2007;48(12):2370-2.

99. Schmidt D, Löscher W. Drug Resistance in Epilepsy: Putative Neurobiologic and Clinical Mechanisms. Epilepsia. 2005;46(6):858-77.

100. Vreugdenhil M, Wadman WJ. Modulation of Sodium Currents in Rat CA1 Neurons by Carbamazepine and Valproate After Kindling Epileptogenesis. Epilepsia. 1999;40(11):1512-22.

101. Remy S, Gabriel S, Urban BW, Dietrich D, Lehmann TN, Elger CE, et al. A novel mechanism underlying drug resistance in chronic epilepsy. Ann Neurol. 2003;53(4):469-79.

102. Epi4K Consortium, Epilepsy Phenome/Genome Project, Allen AS, Berkovic SF, Cossette P, Delanty N, et al. De novo mutations in epileptic encephalopathies. Nature. 2013;501(7466):217-21.

103. Hirose S. Mutant GABAA receptor subunits in genetic (idiopathic) epilepsy. In: Steinlein OK, editor. Progress in Brain Research. 213: Elsevier; 2014. p. 55-85.

104. Hernandez CC, Kong W, Hu N, Zhang Y, Shen W, Jackson L, et al. Altered Channel Conductance States and Gating of GABAA Receptors by a Pore Mutation Linked to Dravet Syndrome. eNeuro. 2017;4(1).

105. Naylor DE. Treating acute seizures with benzodiazepines: does seizure duration matter? Epileptic Disord. 2014;16 Spec No 1:S69-83.

106. Mazarati AM, Baldwin RA, Sankar R, Wasterlain CG. Time-dependent decrease in the effectiveness of antiepileptic drugs during the course of self-sustaining status epilepticus. Brain Res. 1998;814(1):179-85.

107. Mayer SA, Claassen J, Lokin J, Mendelsohn F, Dennis LJ, Fitzsimmons B. Refractory status epilepticus: Frequency, risk factors, and impact on outcome. Arch Neurol. 2002;59(2):205-10.

108. Yuan F, Jia R, Gao Q, Yang F, Yang X, Jiang Y, et al. Early Predictors of Drug-Resistant Epilepsy Development after Convulsive Status Epilepticus. Eur Neurol. 2018;79(5-6):325-32.

109. Deeb TZ, Maguire J, Moss SJ. Possible alterations in GABAA receptor signaling that underlie benzodiazepine-resistant seizures. Epilepsia. 2012;53 Suppl 9:79-88.

110. Koepp MJ, Richardson MP, Labbé C, Brooks DJ, Cunningham VJ, Ashburner J, et al. 11C-flumazenil PET, volumetric MRI, and quantitative pathology in mesial temporal lobe epilepsy. Neurology. 1997;49(3):764-73.

111. Hand KS, Baird VH, Van Paesschen W, Koepp MJ, Revesz T, Thom M, et al. Central benzodiazepine receptor autoradiography in hippocampal sclerosis. Br J Pharmacol. 1997;122(2):358-64.

112. Ryvlin P, Bouvard S, Le Bars D, De Lamérie G, Grégoire MC, Kahane P, et al. Clinical utility of flumazenil-PET versus [18F]fluorodeoxyglucose-PET and MRI in refractory partial epilepsy. A prospective study in 100 patients. Brain. 1998;121(11):2067-81.

113. Bouvard S, Costes N, Bonnefoi F, Lavenne F, Mauguière F, Delforge J, et al. Seizure-related short-term plasticity of benzodiazepine receptors in partial epilepsy: a [11C]flumazenil-PET study. Brain. 2005;128(6):1330-43.

114. Kapur J, Macdonald RL. Rapid Seizure-Induced Reduction of Benzodiazepine and Zn(2+) Sensitivity of Hippocampal Dentate Granule Cell GABA(A) Receptors. J Neurosci. 1997;17(19):7532-40.

115. Brooks-Kayal AR, Shumate MD, Jin H, Rikhter TY, Coulter DA. Selective changes in single cell GABAA receptor subunit expression and function in temporal lobe epilepsy. Nat Med. 1998;4:1166.

116. Kwan P, Brodie MJ. Refractory epilepsy: a progressive, intractable but preventable condition? Seizure. 2002;11(2):77-84.

117. Dallas S, Miller DS, Bendayan R. Multidrug Resistance-Associated Proteins: Expression and Function in the Central Nervous System. Pharmacol Rev. 2006;58(2):140-61.

118. Zhang C, Kwan P, Zuo Z, Baum L. The transport of antiepileptic drugs by P-glycoprotein. Adv Drug Deliv Rev. 2012;64(10):930-42.

119. Tishler DM, Weinberg KI, Hinton DR, Barbaro N, Annett GM, Raffel C. MDR1 Gene Expression in Brain of Patients with Medically Intractable Epilepsy. Epilepsia. 1995;36(1):1-6.

120. Sisodiya SM, Lin WR, Harding BN, Squier MV, Thom M. Drug resistance in epilepsy: expression of drug resistance proteins in common causes of refractory epilepsy. Brain. 2002;125(Pt 1):22-31.

121. Gidal BE. P-glycoprotein Expression and Pharmacoresistant Epilepsy: Cause or Consequence? Epilepsy Curr. 2014;14(3):136-8.

122. Hoffmeyer S, Burk O, von Richter O, Arnold HP, Brockmoller J, Johne A, et al. Functional polymorphisms of the human multidrug-resistance gene: multiple sequence variations

and correlation of one allele with P-glycoprotein expression and activity in vivo. Proc Natl Acad Sci U S A. 2000;97(7):3473-8.

123. Siddiqui A, Kerb R, Weale ME, Brinkmann U, Smith A, Goldstein DB, et al. Association of Multidrug Resistance in Epilepsy with a Polymorphism in the Drug-Transporter Gene ABCB1. N Engl J Med. 2003;348(15):1442-8.

124. Hauser WA, Rich SS, Annegers JF, Anderson VE. Seizure recurrence after a 1st unprovoked seizure: an extended follow-up. Neurology. 1990;40(8):1163-70.

125. Camfield P, Camfield C, Smith S, Dooley J, Smith E. Long-term Outcome Is Unchanged by Antiepileptic Drug Treatment after a First Seizure: A 15-year Follow-up from a Randomized Trial in Childhood. Epilepsia. 2002;43(6):662-3.

126. Marson A, Jacoby A, Johnson A, Kim L, Gamble C, Chadwick D. Immediate versus deferred antiepileptic drug treatment for early epilepsy and single seizures: a randomised controlled trial. Lancet. 2005;365(9476):2007-13.

127. Barton ME, Klein BD, Wolf HH, Steve White H. Pharmacological characterization of the 6 Hz psychomotor seizure model of partial epilepsy. Epilepsy Res. 2001;47(3):217-27.

128. Metcalf CS, West PJ, Thomson K, Edwards S, Smith MD, White HS, et al. Development and Pharmacological Characterization of the Rat 6 Hz Model of Partial Seizures. Epilepsia. 2017;58(6):1073-84.

129. Tang F, Hartz AMS, Bauer B. Drug-Resistant Epilepsy: Multiple Hypotheses, Few Answers. Front Neurol. 2017;8:301.

130. Fang M, Xi ZQ, Wu Y, Wang XF. A new hypothesis of drug refractory epilepsy: neural network hypothesis. Med Hypotheses. 2011;76(6):871-6.

131. Kobow K, El-Osta A, Blumcke I. The methylation hypothesis of pharmacoresistance in epilepsy. Epilepsia. 2013;54 Suppl 2:41-7.

132. Tinel N, Lauritzen I, Chouabe C, Lazdunski M, Borsotto M. The KCNQ2 potassium channel: splice variants, functional and developmental expression. Brain localization and comparison with KCNQ3. FEBS Letters. 1998;438(3):171-6.

133. Brown DA, Passmore GM. Neural KCNQ (Kv7) channels. Br J Pharmacol. 2009;156(8):1185-95.

134. Charlier C, Singh NA, Ryan SG, Lewis TB, Reus BE, Leach RJ, et al. A pore mutation in a novel KQT-like potassium channel gene in an idiopathic epilepsy family. Nat Genet. 1998;18(1):53-5.

135. Singh NA, Charlier C, Stauffer D, DuPont BR, Leach RJ, Melis R, et al. A novel potassium channel gene, KCNQ2, is mutated in an inherited epilepsy of newborns. Nat Genet. 1998;18(1):25-9.

136. Yang W-P, Levesque PC, Little WA, Conder ML, Ramakrishnan P, Neubauer MG, et al. Functional Expression of Two KvLQT1-related Potassium Channels Responsible for an Inherited Idiopathic Epilepsy. Journal of Biological Chemistry. 1998;273(31):19419-23.

137. Martin HC, Kim GE, Pagnamenta AT, Murakami Y, Carvill GL, Meyer E, et al. Clinical whole-genome sequencing in severe early-onset epilepsy reveals new genes and improves molecular diagnosis. Hum Mol Genet. 2014;23(12):3200-11.

138. Saitsu H, Kato M, Koide A, Goto T, Fujita T, Nishiyama K, et al. Whole exome sequencing identifies KCNQ2 mutations in Ohtahara syndrome. Ann Neurol. 2012;72(2):298-300.

139. Mucha M, Ooi L, Linley JE, Mordaka P, Dalle C, Robertson B, et al. Transcriptional control of KCNQ channel genes and the regulation of neuronal excitability. J Neurosci. 2010;30(40):13235-45.

140. Palm K, Belluardo N, Metsis M, Timmusk T. Neuronal Expression of Zinc Finger Transcription Factor REST/NRSF/XBR Gene. J Neurosci. 1998;18(4):1280-96.

141. Calderone A, Jover T, Noh K-m, Tanaka H, Yokota H, Lin Y, et al. Ischemic Insults Derepress the Gene Silencer REST in Neurons Destined to Die. J Neurosci. 2003;23(6):2112-21.

142. Vezzani A, Bartfai T, Bianchi M, Rossetti C, French J. Therapeutic potential of new antiinflammatory drugs. Epilepsia. 2011;52 Suppl 8:67-9.

143. Vezzani A, Aronica E, Mazarati A, Pittman QJ. Epilepsy and brain inflammation. Exp Neurol. 2013;244:11-21.

144. Vezzani A, Lang B, Aronica E. Immunity and Inflammation in Epilepsy. Cold Spring Harb Perspect Med. 2016;6(2):a022699.

145. Gershen LD, Zanotti-Fregonara P, Dustin IH, Liow JS, Hirvonen J, Kreisl WC, et al. Neuroinflammation in Temporal Lobe Epilepsy Measured Using Positron Emission Tomographic Imaging of Translocator Protein. JAMA Neurol. 2015;72(8):882-8.

146. Vezzani A, French J, Bartfai T, Baram TZ. The role of inflammation in epilepsy. Nat Rev Neurol. 2011;7(1):31-40.

147. Viviani B, Gardoni F, Marinovich M. Cytokines and Neuronal Ion Channels in Health and Disease. International Review of Neurobiology. 82: Elsevier; 2007. p. 247-63.

148. Vezzani A, Maroso M, Balosso S, Sanchez M-A, Bartfai T. IL-1 receptor/Toll-like receptor signaling in infection, inflammation, stress and neurodegeneration couples hyperexcitability and seizures. Brain, Behavior, and Immunity. 2011;25(7):1281-9.

149. Takeuchi H, Jin S, Wang J, Zhang G, Kawanokuchi J, Kuno R, et al. Tumor Necrosis Factor-α Induces Neurotoxicity via Glutamate Release from Hemichannels of Activated Microglia in an Autocrine Manner. J Biol Chem. 2006;281(30):21362-8.

150. Stellwagen D, Beattie EC, Seo JY, Malenka RC. Differential Regulation of AMPA Receptor and GABA Receptor Trafficking by Tumor Necrosis Factor-α. J Neurosci. 2005;25(12):3219-28.

151. Meisler MH, O'Brien JE, Sharkey LM. Sodium channel gene family: epilepsy mutations, gene interactions and modifier effects. J Physiol. 2010;588(Pt 11):1841-8.

152. Gorter JA, Potschka H. Drug resistance. In: Noebels JL, Avoli M, Rogawski MA, Olsen RW, Delgado-Escueta AV, editors. Jasper's basic mechanisms of the epilepsies 4th ed. Bethesda: National Center for Biotechnology Information; 2012.

153. Potschka H, Brodie MJ. Pharmacoresistance. In: Stefan H, Theodore WH, editors. Handbook of Clinical Neurology. 108: Elsevier; 2012. p. 741-57.

154. Winkler H. Verbreitung und Ursache der Parthenogenesis im Pflanzen und Tierreiche. Jena: Fischer; 1920.

155. Noguera-Solano R, Ruiz-Gutierrez R, Rodriguez-Caso JM. Genome: twisting stories with DNA. Endeavour. 2013;37(4):213-9.

156. Tate SK, Sisodiya SM. Multidrug resistance in epilepsy: a pharmacogenomic update. Expert Opin Pharmacother. 2007;8(10):1441-9.

157. Cardon LR, Palmer LJ. Population stratification and spurious allelic association. Lancet. 2003;361(9357):598-604.

158. Walker LE, Mirza N, Yip VLM, Marson AG, Pirmohamed M. Personalized medicine approaches in epilepsy. J Intern Med. 2015;277(2):218-34.

159. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. Nature. 2004;431:931.

160. Ezkurdia I, Juan D, Rodriguez JM, Frankish A, Diekhans M, Harrow J, et al. Multiple evidence strands suggest that there may be as few as 19,000 human protein-coding genes. Hum Mol Genet. 2014;23(22):5866-78.

161. Genome Reference Consortium Human Build 38 patch release 12 (GRCh38.p12) [Internet]. 2017 [cited 1.6.2018]. Available from: https://www.ncbi.nlm.nih.gov/assembly/GCA_000001405.27.

162. Genome Reference Consortium. Human Genome Overview [Internet]. 2017 [cited 1.6.2018]. Available from: https://www.ncbi.nlm.nih.gov/grc/human.

163. 1000 Genomes Project Consortium, 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.

164. Zheng-Bradley X, Flicek P. Applications of the 1000 Genomes Project resources. Brief Funct Genomics. 2017;16(3):163-70.

165. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, et al. Global variation in copy number in the human genome. Nature. 2006;444(7118):444-54.

166. Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, Patil N, et al. Characterization of single-nucleotide polymorphisms in coding regions of human genes. Nat Genet. 1999;22(3):231-8.

167. MacDonald JR, Ziman R, Yuen RK, Feuk L, Scherer SW. The Database of Genomic Variants: a curated collection of structural variation in the human genome. Nucleic Acids Res. 2014;42(Database issue):D986-92.

168. Brookes AJ. The essence of SNPs. Gene. 1999;234(2):177-86.

169. Karki R, Pandya D, Elston RC, Ferlini C. Defining "mutation" and "polymorphism" in the era of personal genomics. BMC Med Genomics. 2015;8:37.

170. Bush WS, Moore JH. Chapter 11: Genome-wide association studies. PLoS Comput Biol. 2012;8(12):e1002822.

171. Auer PL, Johnsen JM, Johnson AD, Logsdon BA, Lange LA, Nalls MA, et al. Imputation of Exome Sequence Variants into Population- Based Samples and Blood-Cell-Trait-Associated Loci in African Americans: NHLBI GO Exome Sequencing Project. Am J Hum Genet. 2012;91(5):794-808.

172. Griffith OL, Montgomery SB, Bernier B, Chu B, Kasaian K, Aerts S, et al. ORegAnno: an open-access community-driven resource for regulatory annotation. Nucleic Acids Res. 2008;36(Database issue):D107-D13.

173. Myles S, Davison D, Barrett J, Stoneking M, Timpson N. Worldwide population differentiation at disease-associated SNPs. BMC Med Genomics. 2008;1(1):22.

174. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature. 2009;461:747.

175. Low SK, Takahashi A, Mushiroda T, Kubo M. Genome-Wide Association Study: A Useful Tool to Identify Common Genetic Variants Associated with Drug Toxicity and Efficacy in Cancer Pharmacogenomics. Clin Cancer Res. 2014;20(10):2541-52.

176. Risch NJ. Searching for genetic determinants in the new millennium. Nature. 2000;405(6788):847-56.

177. Takeuchi F, McGinnis R, Bourgeois S, Barnes C, Eriksson N, Soranzo N, et al. A Genome-Wide Association Study Confirms VKORC1, CYP2C9, and CYP4F2 as Principal Genetic Determinants of Warfarin Dose. PLoS Genet. 2009;5(3):e1000433.

178. 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.

179. Moulard B, Chaigne D, Mouthon D, Buresi C, Guipponi M, Malafosse A. Identification of a New Locus for Generalized Epilepsy with Febrile Seizures Plus (GEFS+) on Chromosome 2q24-q33. Am J Hum Genet. 1999;65(5):1396-400.

180. Escayg A, MacDonald BT, Meisler MH, Baulac S, Huberfeld G, An-Gourfinkel I, et al. Mutations of SCN1A, encoding a neuronal sodium channel, in two families with GEFS+2. Nat Genet. 2000;24(4):343-5.

181. Claes L, Del-Favero J, Ceulemans B, Lagae L, Van Broeckhoven C, De Jonghe P. De novo mutations in the sodium-channel gene SCN1A cause severe myoclonic epilepsy of infancy. Am J Hum Genet. 2001;68(6):1327-32.

182. Kwon JM, Goate AM. The candidate gene approach. Alcohol Res Health. 2000;24(3):164-8.

183. St George-Hyslop PH, Tanzi RE, Polinsky RJ, Haines JL, Nee L, Watkins PC, et al. The genetic defect causing familial Alzheimer's disease maps on chromosome 21. Science. 1987;235(4791):885-90.

184. Goate A, Chartier-Harlin MC, Mullan M, Brown J, Crawford F, Fidani L, et al. Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. Nature. 1991;349(6311):704-6.

185. Bailey-Wilson JE, Wilson AF. Linkage analysis in the next-generation sequencing era. Hum Hered. 2011;72(4):228-36.

186. Bowden DW, An SS, Palmer ND, Brown WM, Norris JM, Haffner SM, et al. Molecular basis of a linkage peak: exome sequencing and family-based analysis identify a rare genetic variant in the ADIPOQ gene in the IRAS Family Study. Hum Mol Genet. 2010;19(20):4112-20.

187. Kasperavičiūtė D, Sisodiya SM. Epilepsy pharmacogenetics. Pharmacogenomics. 2009;10(5):817-36.

188. Chung WH, Hung SI, Hong HS, Hsih MS, Yang LC, Ho HC, et al. Medical genetics: a marker for Stevens-Johnson syndrome. Nature. 2004;428(6982):486.

189. McCormack M, Alfirevic A, Bourgeois S, Farrell JJ, Kasperavičiūtė D, Carrington M, et al. HLA-A*3101 and Carbamazepine-Induced Hypersensitivity Reactions in Europeans. The New England journal of medicine. 2011;364(12):1134-43.

190. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. Nat Rev Genet. 2005;6(2):95-108.

191. Pearson TA, Manolio TA. How to interpret a genome-wide association study. JAMA. 2008;299(11):1335-44.

192. Lander ES. The New Genomics: Global Views of Biology. Science. 1996;274(5287):536-9.

193. Chakravarti A. Population genetics—making sense out of sequence. Nat Genet. 1999;21(1):56-60.

194. Risch N, Merikangas K. The future of genetic studies of complex human diseases. Science. 1996;273(5281):1516-7.

195. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007;447(7145):661-78.

196. International HapMap Consortium, Altshuler D, Donnelly P. A haplotype map of the human genome. Nature. 2005;437:1299.

197. International HapMap Consortium, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, et al. A second generation human haplotype map of over 3.1 million SNPs. Nature. 2007;449(7164):851-61.

198. International HapMap 3 Consortium, Altshuler DM, Gibbs RA, Peltonen L, Altshuler DM, Gibbs RA, et al. Integrating common and rare genetic variation in diverse human populations. Nature. 2010;467(7311):52-8.

199. Barrett JC, Cardon LR. Evaluating coverage of genome-wide association studies. Nat Genet. 2006;38(6):659-62.

200. Mullen SA, Crompton DE, Carney PW, Helbig I, Berkovic SF. A neurologist's guide to genome-wide association studies. Neurology. 2009;72(6):558-65.

201. Amos CI. Successful design and conduct of genome-wide association studies. Hum Mol Genet. 2007;16(R2):R220-R5.

202. Haines JL, Hauser MA, Schmidt S, Scott WK, Olson LM, Gallins P, et al. Complement Factor H Variant Increases the Risk of Age-Related Macular Degeneration. Science. 2005;308(5720):419-21.

203. Li MJ, Liu Z, Wang P, Wong MP, Nelson MR, Kocher JP, et al. GWASdb v2: an update database for human genetic variants identified by genome-wide association studies. Nucleic Acids Res. 2016;44(D1):D869-76.

204. Stranger BE, Stahl EA, Raj T. Progress and promise of genome-wide association studies for human complex trait genetics. Genetics. 2011;187(2):367-83.

205. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. Am J Hum Genet. 2017;101(1):5-22.

206. Alonso N, Lucas G, Hysi P. Big data challenges in bone research: genome-wide association studies and next-generation sequencing. Bonekey Rep. 2015;4:635.

207. Lee S, Abecasis GR, Boehnke M, Lin X. Rare-variant association analysis: study designs and statistical tests. Am J Hum Genet. 2014;95(1):5-23.

208. Witte JS, Visscher PM, Wray NR. The contribution of genetic variants to disease depends on the ruler. Nat Rev Genet. 2014;15(11):765-76.

209. Yang J, Zeng J, Goddard ME, Wray NR, Visscher PM. Concepts, estimation and interpretation of SNP-based heritability. Nat Genet. 2017;49(9):1304-10.

210. Lee SH, Yang J, Goddard ME, Visscher PM, Wray NR. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. Bioinformatics. 2012;28(19):2540-2.

211. Wray NR, Visscher PM. Estimating trait heritability. Nature Education [Internet]. 2008; 1(1):[29 p.].

212. Visscher PM, Medland SE, Ferreira MAR, Morley KI, Zhu G, Cornes BK, et al. Assumption-Free Estimation of Heritability from Genome-Wide Identity-by-Descent Sharing between Full Siblings. PLoS Genet. 2006;2(3):e41.

213. Maher B. Personal genomes: The case of the missing heritability. Nature. 2008;456(7218):18-21.

214. Gibson G. Hints of hidden heritability in GWAS. Nat Genet. 2010;42(7):558-60.

215. Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, et al. Missing heritability and strategies for finding the underlying causes of complex disease. Nat Rev Genet. 2010;11(6):446-50.

216. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, et al. Common SNPs explain a large proportion of the heritability for human height. Nat Genet. 2010;42(7):565-9.

217. Yang J, Bakshi A, Zhu Z, Hemani G, Vinkhuyzen AAE, Lee SH, et al. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. Nat Genet. 2015;47(10):1114-20.

218. Macgregor S, Cornes BK, Martin NG, Visscher PM. Bias, precision and heritability of self-reported and clinically measured height in Australian twins. Hum Genet. 2006;120(4):571-80.

219. Silventoinen K, Sammalisto S, Perola M, Boomsma DI, Cornes BK, Davis C, et al. Heritability of adult body height: a comparative study of twin cohorts in eight countries. Twin Res. 2003;6(5):399-408.

220. Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. Nat Genet. 2014;46(11):1173-86.

221. Peljto AL, Barker-Cummings C, Vasoli VM, Leibson CL, Hauser WA, Buchhalter JR, et al. Familial risk of epilepsy: a population-based study. Brain. 2014;137(Pt 3):795-805.

222. Lennox WG, Lennox MA. Epilepsy and related disorders. Boston: Little Brown; 1960.

223. Kjeldsen MJ, Corey LA, Christensen K, Friis ML. Epileptic seizures and syndromes in twins: the importance of genetic factors. Epilepsy Res. 2003;55(1-2):137-46.

224. Kjeldsen MJ, Corey LA, Solaas MH, Friis ML, Harris JR, Kyvik KO, et al. Genetic Factors in Seizures: A Population-Based Study of 47,626 US, Norwegian and Danish Twin Pairs. Twin Res Hum Genet. 2005;8(02):138-47.

225. Berkovic SF, Howell RA, Hay DA, Hopper JL. Epilepsies in twins: Genetics of the major epilepsy syndromes. Ann Neurol. 1998;43(4):435-45.

226. Vadlamudi L, Milne RL, Lawrence K, Heron SE, Eckhaus J, Keay D, et al. Genetics of epilepsy: The testimony of twins in the molecular era. Neurology. 2014;83(12):1042-8.

227. Corey LA, Pellock JM, Kjeldsen MJ, Nakken KO. Importance of genetic factors in the occurrence of epilepsy syndrome type: a twin study. Epilepsy Res. 2011;97(1-2):103-11.

228. Sillanpää M, Koskenvuo M, Romanov K, Kaprio J. Genetic factors in epileptic seizures: evidence from a large twin population. Acta Neurol Scand. 1991;84(6):523-6.

229. Miller LL, Pellock JM, DeLorenzo RJ, Meyer JM, Corey LA. Univariate genetic analyses of epilepsy and seizures in a population-based twin study: The Virginia twin registry. Genet Epidemiol. 1998;15(1):33-49.

230. Kjeldsen MJ, Kyvik KO, Christensen K, Friis ML. Genetic and environmental factors in epilepsy: a population-based study of 11 900 Danish twin pairs. Epilepsy Res. 2001;44(2):167-78.

231. Pal DK, Pong AW, Chung WK. Genetic evaluation and counseling for epilepsy. Nat Rev Neurol. 2010;6(8):445-53.

232. Wang J, Lin ZJ, Liu L, Xu HQ, Shi YW, Yi YH, et al. Epilepsy-associated genes. Seizure. 2017;44:11-20.

233. Møller RS, Larsen LHG, 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.

234. Scheffer IE, Berkovic SF. Generalized epilepsy with febrile seizures plus. A genetic disorder with heterogeneous clinical phenotypes. Brain. 1997;120 (Pt 3):479-90.

235. Thomas RH, Berkovic SF. The hidden genetics of epilepsy—a clinically important new paradigm. Nat Rev Neurol. 2014;10(5):283-92.

236. Chipaux M, Szurhaj W, Vercueil L, Milh M, Villeneuve N, Cances C, et al. Epilepsy diagnostic and treatment needs identified with a collaborative database involving tertiary centers in France. Epilepsia. 2016;57(5):757-69.

237. Mullen SA, Carvill GL, Bellows S, Bayly MA, Trucks H, Lal D, et al. Copy number variants are frequent in genetic generalized epilepsy with intellectual disability. Neurology. 2013;81(24):2148-.

238. Dunn P, Albury CL, Maksemous N, Benton MC, Sutherland HG, Smith RA, et al. Next Generation Sequencing Methods for Diagnosis of Epilepsy Syndromes. Front Genet. 2018;9:20.

239. Steinlein OK, Mulley JC, Propping P, Wallace RH, Phillips HA, Sutherland GR, et al. A missense mutation in the neuronal nicotinic acetylcholine receptor α4 subunit is associated with autosomal dominant nocturnal frontal lobe epilepsy. Nat Genet. 1995;11(2):201-3.

240. Mulley JC, Scheffer IE, Petrou S, Dibbens LM, Berkovic SF, Harkin LA. SCN1A mutations and epilepsy. Hum Mutat. 2005;25(6):535-42.

241. Veeramah KR, O'Brien JE, Meisler MH, Cheng X, Dib-Hajj SD, Waxman SG, et al. De novo pathogenic SCN8A mutation identified by whole-genome sequencing of a family quartet affected by infantile epileptic encephalopathy and SUDEP. Am J Hum Genet. 2012;90(3):502-10.

242. Gardella E, Becker F, Møller RS, Schubert J, Lemke JR, Larsen LH, et al. Benign infantile seizures and paroxysmal dyskinesia caused by an SCN8A mutation. Ann Neurol. 2016;79(3):428-36.

243. Nava C, Dalle C, Rastetter A, Striano P, de Kovel CG, Nabbout R, et al. De novo mutations in HCN1 cause early infantile epileptic encephalopathy. Nat Genet. 2014;46(6):640-5.

244. Carvill GL, Weckhuysen S, McMahon JM, Hartmann C, Møller RS, Hjalgrim H, et al. GABRA1 and STXBP1: Novel genetic causes of Dravet syndrome. Neurology. 2014;82(14):1245-53.

245. Carvill GL, Heavin SB, Yendle SC, McMahon JM, O'Roak BJ, Cook J, et al. Targeted resequencing in epileptic encephalopathies identifies de novo mutations in CHD2 and SYNGAP1. Nat Genet. 2013;45(7):825-30.

246. Galizia EC, Myers CT, Leu C, de Kovel CGF, Afrikanova T, Cordero-Maldonado ML, et al. CHD2 variants are a risk factor for photosensitivity in epilepsy. Brain. 2015;138(5):1198-208.

247. Thomas RH, Zhang LM, Carvill GL, Archer JS, Heavin SB, Mandelstam SA, et al. CHD2 myoclonic encephalopathy is frequently associated with self-induced seizures. Neurology. 2015;84(9):951-8.

248. Mignot C, von Stülpnagel C, Nava C, Ville D, Sanlaville D, Lesca G, et al. Genetic and neurodevelopmental spectrum of SYNGAP1-associated intellectual disability and epilepsy. J Med Genet. 2016;53(8):511-22.

249. Atkins L, Miller WL, Salam M. A ring-20 chromosome. J Med Genet. 1972;9(3):377-80.

250. Stewart LR, Hall AL, Kang SH, Shaw CA, Beaudet AL. High frequency of known copy number abnormalities and maternal duplication 15q11-q13 in patients with combined schizophrenia and epilepsy. BMC Med Genet. 2011;12(1):154.

251. Helbig I, Mefford HC, Sharp AJ, Guipponi M, Fichera M, Franke A, et al. 15q13.3 microdeletions increase risk of idiopathic generalized epilepsy. Nat Genet. 2009;41(2):160-2.

252. Mefford HC. CNVs in Epilepsy. Curr Genet Med Rep. 2014;2:162-7.

253. Addis L, Rosch RE, Valentin A, Makoff A, Robinson R, Everett KV, et al. Analysis of rare copy number variation in absence epilepsies. Neurol Genet. 2016;2(2):e56.

254. de Kovel CGF, Trucks H, Helbig I, Mefford HC, Baker C, Leu C, et al. Recurrent microdeletions at 15q11.2 and 16p13.11 predispose to idiopathic generalized epilepsies. Brain. 2010;133(1):23-32.

255. Monlong J, Girard SL, Meloche C, Cadieux-Dion M, Andrade DM, Lafreniere RG, et al. Global characterization of copy number variants in epilepsy patients from whole genome sequencing. PLOS Genetics. 2018;14(4):e1007285.

256. Epilepsy Phenome/Genome Project, Epi4K Consortium. Copy number variant analysis from exome data in 349 patients with epileptic encephalopathy. Ann Neurol. 2015;78(2):323-8.

257. Mefford HC, Yendle SC, Hsu C, Cook J, Geraghty E, McMahon JM, et al. Rare copy number variants are an important cause of epileptic encephalopathies. Ann Neurol. 2011;70(6):974-85.

258. EPICURE Consortium, EMINet Consortium, Steffens M, Leu C, Ruppert AK, Zara F, et al. Genome-wide association analysis of genetic generalized epilepsies implicates

susceptibility loci at 1q43, 2p16.1, 2q22.3 and 17q21.32. Hum Mol Genet. 2012;21(24):5359-72.

259. Kasperavičiūtė D, Catarino CB, Matarin M, Leu C, Novy J, Tostevin A, et al. Epilepsy, hippocampal sclerosis and febrile seizures linked by common genetic variation around SCN1A. Brain. 2013;136(10):3140-50.

260. Koeleman BPC. What do genetic studies tell us about the heritable basis of common epilepsy? Polygenic or complex epilepsy? Neurosci Lett. 2018;667:10-6.

261. International League Against Epilepsy Consortium on Complex Epilepsies. Genetic determinants of common epilepsies: a meta-analysis of genome-wide association studies. Lancet Neurol. 2014;13(9):893-903.

262. International League Against Epilepsy Consortium on Complex Epilepsies. Genome-wide mega-analysis identifies 16 loci and highlights diverse biological mechanisms in the common epilepsies. Nat Commun. 2018;9(1):5269.

263. Neale BM, Sklar P. Genetic analysis of schizophrenia and bipolar disorder reveals polygenicity but also suggests new directions for molecular interrogation. Curr Opin Neurobiol. 2015;30:131-8.

264. Jansen PR, Watanabe K, Stringer S, Skene N, Bryois J, Hammerschlag AR, et al. Genome-wide Analysis of Insomnia (N=1,331,010) Identifies Novel Loci and Functional Pathways. bioRxiv. 2018.

265. Vadlamudi L, Andermann E, Lombroso CT, Schachter SC, Milne RL, Hopper JL, et al. Epilepsy in twins: insights from unique historical data of William Lennox. Neurology. 2004;62(7):1127-33.

266. Johnson MR, Milne RL, Torn-Broers Y, Hopper JL, Scheffer IE, Berkovic SF. A Twin Study of Genetic Influences on Epilepsy Outcome. Twin Res. 2003;6(2):140-6.

267. Löscher W, Rundfeldt C, Hönack D. Pharmacological characterization of phenytoin-resistant amygdala-kindled rats, a new model of drug-resistant partial epilepsy. Epilepsy Res. 1993;15(3):207-19.

268. Löscher W, Cramer S, Ebert U. Selection of Phenytoin Responders and Nonresponders in Male and Female Amygdala-Kindled Sprague–Dawley Rats. Epilepsia. 1998;39(11):1138-47.

269. Cramer S, Ebert U, Löscher W. Characterization of phenytoin-resistant kindled rats, a new model of drug-resistant partial epilepsy: comparison of inbred strains. Epilepsia. 1998;39(10):1046-53.

270. Urso R, Blardi P, Giorgi G. A short introduction to pharmacokinetics. Eur Rev Med Pharmacol Sci. 2002;6(2):33-44.

271. Emich-Widera E, Likus W, Kazek B, Niemiec P, Balcerzyk A, Sieroń AL, et al. CYP3A5∗3 and C3435T MDR1 Polymorphisms in Prognostication of Drug-Resistant Epilepsy in Children and Adolescents. BioMed Research International. 2013;2013:526837.

272. Mamiya K, Ieiri I, Shimamoto J, Yukawa E, Imai J, Ninomiya H, et al. The Effects of Genetic Polymorphisms of CYP2C9 and CYP2C 19 on Phenytoin Metabolism in Japanese Adult Patients with Epilepsy: Studies in Stereoselective Hydroxylation and Population Pharmacokinetics. Epilepsia. 1998;39(12):1317-23.

273. van der Weide J, Steijns LSW, van Weelden MJM, de Haan K. The effect of genetic polymorphism of cytochrome P450 CYP2C9 on phenytoin dose requirement. Pharmacogenetics. 2001;11(4):287-91.

274. Tate SK, Depondt C, Sisodiya SM, Cavalleri GL, Schorge S, Soranzo N, et al. Genetic predictors of the maximum doses patients receive during clinical use of the anti-epileptic drugs carbamazepine and phenytoin. Proc Natl Acad Sci U S A. 2005;102(15):5507-12.

275. Mosher CM, Tai G, Rettie AE. CYP2C9 Amino Acid Residues Influencing Phenytoin Turnover and Metabolite Regio- and Stereochemistry. J Pharmacol Exp Ther. 2009;329(3):938-44.

276. Rosemary J, Surendiran A, Rajan S, Shashindran CH, Adithan C. Influence of the CYP2C9 AND CYP2C19 polymorphisms on phenytoin hydroxylation in healthy individuals from south India. Indian J Med Res. 2006;123(5):665-70

277. Ninomiya H, Mamiya K, Matsuo S, Ieiri I, Higuchi S, Tashiro N. Genetic polymorphism of the CYP2C subfamily and excessive serum phenytoin concentration with central nervous system intoxication. Ther Drug Monit. 2000;22(2):230-2.

278. 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.

279. Depondt C, Godard P, Espel RS, Da Cruz AL, Lienard P, Pandolfo M. A candidate gene study of antiepileptic drug tolerability and efficacy identifies an association of CYP2C9 variants with phenytoin toxicity. Eur J Neurol. 2011;18(9):1159-64.

280. Daci A, Beretta G, Vllasaliu D, Shala A, Govori V, Norata GD, et al. Polymorphic Variants of SCN1A and EPHX1 Influence Plasma Carbamazepine Concentration, Metabolism and Pharmacoresistance in a Population of Kosovar Albanian Epileptic Patients. PLoS One. 2015;10(11):e0142408.

281. Lakhan R, Kumari R, Singh K, Kalita J, Misra UK, Mittal B. Possible role of CYP2C9 & CYP2C19 single nucleotide polymorphisms in drug refractory epilepsy. Indian J Med Res. 2011;134(3):295-301.

282. Seven M, Batar B, Unal S, Yesil G, Yuksel A, Guven M. The drug-transporter gene MDR1 C3435T and G2677T/A polymorphisms and the risk of multidrug-resistant epilepsy in Turkish children. Mol Biol Rep. 2014;41(1):331-6.

283. Wolf SJ, Bachtiar M, Wang J, Sim TS, Chong SS, Lee CGL. An update on ABCB1 pharmacogenetics: insights from a 3D model into the location and evolutionary conservation of residues corresponding to SNPs associated with drug pharmacokinetics. Pharmacogenomics J. 2011;11:315.

284. Ieiri I, Takane H, Hirota T, Otsubo K, Higuchi S. Genetic polymorphisms of drug transporters: pharmacokinetic and pharmacodynamic consequences in pharmacotherapy. Expert Opin Drug Metab Toxicol. 2006;2(5):651-74.

285. Ensembl release 93; Gene: ABCB1 [Internet]. 2018. Available from: https://www.ensembl.org/Homo_sapiens/Gene/Variation_Gene/Table?db=core;g=ENSG00000085563;r=7:87503633-87713323.

286. Sayyah M, Kamgarpour F, Maleki M, Karimipoor M, Gharagozli K, Shamshiri AR. Association analysis of intractable epilepsy with C3435T and G2677T/A ABCB1 gene polymorphisms in Iranian patients. Epileptic Disord. 2011;13(2):155-65.

287. Sánchez MB, Herranz JL, Leno C, Arteaga R, Oterino A, Valdizán EM, et al. Genetic factors associated with drug-resistance of epilepsy: relevance of stratification by patient age and aetiology of epilepsy. Seizure. 2010;19(2):93-101.

288. Hung CC, Jen Tai J, Kao PJ, Lin MS, Liou HH. Association of polymorphisms in NR1I2 and ABCB1 genes with epilepsy treatment responses. Pharmacogenomics. 2007;8(9):1151-8.

289. Soranzo N, Cavalleri GL, Weale ME, Wood NW, Depondt C, Marguerie R, et al. Identifying Candidate Causal Variants Responsible for Altered Activity of the ABCB1 Multidrug Resistance Gene. Genome Res. 2004;14(7):1333-44.

290. Kwan P, Baum L, Wong V, Ng PW, Lui CHT, Sin NC, et al. Association between ABCB1 C3435T polymorphism and drug-resistant epilepsy in Han Chinese. Epilepsy Behav. 2007;11(1):112-7.

291. Shaheen U, Prasad DK, Sharma V, Suryaprabha T, Ahuja YR, Jyothy A, et al. Significance of MDR1 gene polymorphism C3435T in predicting drug response in epilepsy. Epilepsy Res. 2014;108(2):251-6.

292. Tan NCK, Heron SE, Scheffer IE, Pelekanos JT, McMahon JM, Vears DF, et al. Failure to confirm association of a polymorphism in ABCB1 with multidrug-resistant epilepsy. Neurology. 2004;63(6):1090-2.

293. Sills GJ, Mohanraj R, Butler E, McCrindle S, Collier L, Wilson EA, et al. Lack of Association between the C3435T Polymorphism in the Human Multidrug Resistance (MDR1) Gene and Response to Antiepileptic Drug Treatment. Epilepsia. 2005;46(5):643-7.

294. Manna I, Gambardella A, Labate A, Mumoli L, Ferlazzo E, Pucci F, et al. Polymorphism of the multidrug resistance 1 gene MDR1/ABCB1 C3435T and response to antiepileptic drug treatment in temporal lobe epilepsy. Seizure. 2015;24:124-6.

295. Bournissen FG, Moretti ME, Juurlink DN, Koren G, Walker M, Finkelstein Y. Polymorphism of the MDR1/ABCB1 C3435T drug-transporter and resistance to anticonvulsant drugs: A meta-analysis. Epilepsia. 2009;50(4):898-903.

296. Haerian BS, Lim KS, Tan HJ, Mohamed EH, Tan CT, Raymond AA, et al. Association between ABCB1 polymorphism and response to sodium valproate treatment in Malaysian epilepsy patients. Epileptic Disord. 2011;13(1):65-75.

297. Li M, Tan J, Yang X, Su L, Xie J, Liang B, et al. The ABCB1-C3435T polymorphism likely acts as a risk factor for resistance to antiepileptic drugs. Epilepsy Res. 2014;108(6):1052-67.

298. Li SX, Liu YY, Wang QB. ABCB1 gene C3435T polymorphism and drug resistance in epilepsy: evidence based on 8,604 subjects. Med Sci Monit. 2015;21:861-8.

299. Chouchi M, Kaabachi W, Klaa H, Tizaoui K, Turki IB, Hila L. Relationship between ABCB1 3435TT genotype and antiepileptic drugs resistance in Epilepsy: updated systematic review and meta-analysis. BMC Neurol. 2017;17(1):32.

300. Beck H. Plasticity of antiepileptic drug targets. Epilepsia. 2007;48 Suppl 1:14-8.

301. Dravet C. Dravet syndrome history. Dev Med Child Neurol. 2011;53 Suppl 2:1-6.

302. Yu FH, Mantegazza M, Westenbroek RE, Robbins CA, Kalume F, Burton KA, et al. Reduced sodium current in GABAergic interneurons in a mouse model of severe myoclonic epilepsy in infancy. Nat Neurosci. 2006;9:1142.

303. Ogiwara I, Miyamoto H, Morita N, Atapour N, Mazaki E, Inoue I, et al. Nav1.1 localizes to axons of parvalbumin-positive inhibitory interneurons: a circuit basis for epileptic seizures in mice carrying an Scn1a gene mutation. J Neurosci. 2007;27(22):5903-14.

304. Arzimanoglou A. Dravet syndrome: From electroclinical characteristics to molecular biology. Epilepsia. 2009;50(s8):3-9.

305. Liu Y, Lopez-Santiago LF, Yuan Y, Jones JM, Zhang H, O'Malley HA, et al. Dravet syndrome patient-derived neurons suggest a novel epilepsy mechanism. Ann Neurol. 2013;74(1):128-39.

306. Isom LL. Opposing Phenotypes in Dravet Syndrome Patient-Derived Induced Pluripotent Stem Cell Neurons: Can Everyone Be Right? Epilepsy Curr. 2017;17(4):244-7.

307. Wirrell EC, Laux L, Donner E, Jette N, Knupp K, Meskis MA, et al. Optimizing the Diagnosis and Management of Dravet Syndrome: Recommendations From a North American Consensus Panel. Pediatr Neurol. 2017;68:18-34.e3.

308. Billakota S, Devinsky O, Marsh E. Cannabinoid therapy in epilepsy. Curr Opin Neurol. 2019;32(2):220-6.

309. Zhang Y, Kecskes A, Copmans D, Langlois M, Crawford AD, Ceulemans B, et al. Pharmacological characterization of an antisense knockdown zebrafish model of Dravet syndrome: inhibition of epileptic seizures by the serotonin agonist fenfluramine. PLoS One. 2015;10(5):e0125898.

310. Schoonjans AS, Lagae L, Ceulemans B. Low-dose fenfluramine in the treatment of neurologic disorders: experience in Dravet syndrome. Ther Adv Neurol Disord. 2015;8(6):328-38.

311. Schoonjans A, Paelinck BP, Marchau F, Gunning B, Gammaitoni A, Galer BS, et al. Low-dose fenfluramine significantly reduces seizure frequency in Dravet syndrome: a prospective study of a new cohort of patients. Eur J Neurol. 2017;24(2):309-14.

312. BRIEF-Zogenix Announces New Data From Pivotal Phase 3 Clinical Trial Of ZX008 [Internet]. Reuters. 2017 [cited 4.8.2018]. Available from: https://www.reuters.com/article/brief-zogenix-announces-new-data-from-pi/brief-zogenix-announces-new-data-from-pivotal-phase-3-clinical-trial-of-zx008-idUSASB0BWCE.

313. Margari L, Legrottaglie AR, Vincenti A, Coppola G, Operto FF, Buttiglione M, et al. Association between SCN1A gene polymorphisms and drug resistant epilepsy in pediatric patients. Seizure. 2018;55:30-5.

314. Heinzen EL, Yoon W, Tate SK, Sen A, Wood NW, Sisodiya SM, et al. Nova2 interacts with a cis-acting polymorphism to influence the proportions of drug-responsive splice variants of SCN1A. The American Journal of Human Genetics. 2007;80(5):876-83.

315. Tate SK, Singh R, Hung CC, Tai JJ, Depondt C, Cavalleri GL, et al. A common polymorphism in the SCN1A gene associates with phenytoin serum levels at maintenance dose. Pharmacogenet Genomics. 2006;16(10):721-6.

316. Zimprich F, Stogmann E, Bonelli S, Baumgartner C, Mueller JC, Meitinger T, et al. A functional polymorphism in the SCN1A gene is not associated with carbamazepine dosages in Austrian patients with epilepsy. Epilepsia. 2008;49(6):1108-9.

317. Abe T, Seo T, Ishitsu T, Nakagawa T, Hori M, Nakagawa K. Association between SCN1A polymorphism and carbamazepine-resistant epilepsy. Br J Clin Pharmacol. 2008;66(2):304-7.

318. Angelopoulou C, Veletza S, Heliopoulos I, Vadikolias K, Tripsianis G, Stathi C, et al. Association of SCN1A gene polymorphism with antiepileptic drug responsiveness in the population of Thrace, Greece. Arch Med Sci. 2017;13(1):138-47.

319. Zhou L, Cao Y, Long H, Long L, Xu L, Liu Z, et al. ABCB1, ABCC2, SCN1A, SCN2A, GABRA1 gene polymorphisms and drug resistant epilepsy in the Chinese Han population. Pharmazie. 2015;70(6):416-20.

320. Manna I, Gambardella A, Bianchi A, Striano P, Tozzi R, Aguglia U, et al. A functional polymorphism in the SCN1A gene does not influence antiepileptic drug responsiveness in Italian patients with focal epilepsy. Epilepsia. 2011;52(5):e40-4.

321. Kumari R, Lakhan R, Kumar S, Garg RK, Misra UK, Kalita J, et al. SCN1AIVS5-91G>A polymorphism is associated with susceptibility to epilepsy but not with drug responsiveness. Biochimie. 2013;95(6):1350-3.

322. Yip TSC, O'Doherty C, Tan NCK, Dibbens LM, Suppiah V. SCN1A variations and response to multiple antiepileptic drugs. Pharmacogenomics J. 2014;14(4):385-9.

323. Seidner G, Alvarez MG, Yeh JI, O'Driscoll KR, Klepper J, Stump TS, et al. GLUT-1 deficiency syndrome caused by haploinsufficiency of the blood-brain barrier hexose carrier. Nat Genet. 1998;18(2):188-91.

324. Larsen J, Johannesen KM, Ek J, Tang S, Marini C, Blichfeldt S, et al. The role of SLC2A1 mutations in myoclonic astatic epilepsy and absence epilepsy, and the estimated frequency of GLUT1 deficiency syndrome. Epilepsia. 2015;56(12):e203-e8.

325. Lebon S, Suarez P, Alija S, Korff CM, Fluss J, Mercati D, et al. When should clinicians search for GLUT1 deficiency syndrome in childhood generalized epilepsies? Eur J Paediatr Neurol. 2015;19(2):170-5.

326. Rotstein M, Engelstad K, Yang H, Wang D, Levy B, Chung WK, et al. Glut1 deficiency: inheritance pattern determined by haploinsufficiency. Ann Neurol. 2010;68(6):955-8.

327. Yang H, Wang D, Engelstad K, Bagay L, Wei Y, Rotstein M, et al. Glut1 deficiency syndrome and erythrocyte glucose uptake assay. Ann Neurol. 2011;70(6):996-1005.

328. Leen WG, Klepper J, Verbeek MM, Leferink M, Hofste T, van Engelen BG, et al. Glucose transporter-1 deficiency syndrome: the expanding clinical and genetic spectrum of a treatable disorder. Brain. 2010;133(3):655-70.

329. De Vivo DC, Leary L, Wang D. Glucose transporter 1 deficiency syndrome and other glycolytic defects. J Child Neurol. 2002;17 Suppl 3:3S15-23.

330. French JA, Lawson JA, Yapici Z, Ikeda H, Polster T, Nabbout R, et al. Adjunctive everolimus therapy for treatment-resistant focal-onset seizures associated with tuberous sclerosis (EXIST-3): a phase 3, randomised, double-blind, placebo-controlled study. Lancet. 2016;388(10056):2153-63.

331. Pierson TM, Yuan H, Marsh ED, Fuentes-Fajardo K, Adams DR, Markello T, et al. GRIN2A mutation and early-onset epileptic encephalopathy: personalized therapy with memantine. Ann Clin Transl Neurol. 2014;1(3):190-8.

332. Poduri A. Meta-Analysis Revives Genome-Wide Association Studies in Epilepsy. Epilepsy Curr. 2015;15(3):122-3.

333. Speed D, Hoggart C, Petrovski S, Tachmazidou I, Coffey A, Jorgensen A, et al. A genome-wide association study and biological pathway analysis of epilepsy prognosis in a prospective cohort of newly treated epilepsy. Hum Mol Genet. 2014;23(1):247-58.

334. Epi4K Consortium. Epi4K: gene discovery in 4,000 genomes. Epilepsia. 2012;53(8):1457-67.

335. Epi4K Consortium. Phenotypic analysis of 303 multiplex families with common epilepsies. Brain. 2017;140(8):2144-56.

336. Delanty N, Cavallleri G. Genomics-Guided Precise Anti-Epileptic Drug Development. Neurochem Res. 2017;42(7):2084-8.

337. EpiPM Consortium. A roadmap for precision medicine in the epilepsies. Lancet Neurol. 2015;14(12):1219-28.

338. Reif PS, Tsai MH, Helbig I, Rosenow F, Klein KM. Precision medicine in genetic epilepsies: break of dawn? Expert Rev Neurother. 2017;17(4):381-92.

339. Striano P, Vari MS, Mazzocchetti C, Verrotti A, Zara F. Management of genetic epilepsies: From empirical treatment to precision medicine. Pharmacol Res. 2016;107:426-9.

340. McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet. 2008;9(5):356-69.

341. Tellez-Zenteno JF, Hernandez-Ronquillo L, Buckley S, Zahagun R, Rizvi S. A validation of the new definition of drug-resistant epilepsy by the International League Against Epilepsy. Epilepsia. 2014;55(6):829-34.

342. Hlebokazov F, Dakukina T, Ihnatsenko S, Kosmacheva S, Potapnev M, Shakhbazau A, et al. Treatment of refractory epilepsy patients with autologous mesenchymal stem cells reduces seizure frequency: An open label study. Adv Med Sci. 2017;62(2):273-9.

343. Cação G, Parra J, Mannan S, Sisodiya SM, Sander JW. Juvenile myoclonic epilepsy refractory to treatment in a tertiary referral center. Epilepsy Behav. 2018;82:81-6.

344. Vignoli A, Peron A, Turner K, Scornavacca GF, La Briola F, Chiesa V, et al. Long-term outcome of epilepsy with onset in the first three years of life: Findings from a large cohort of patients. Eur J Paediatr Neurol. 2016;20(4):566-72.

345. Voll A, Hernandez-Ronquillo L, Buckley S, Tellez-Zenteno JF. Predicting drug resistance in adult patients with generalized epilepsy: A case-control study. Epilepsy Behav. 2015;53:126-30.

346. Chung SS, Hogan RE, Blatt I, Lawson P B, Nguyen H, Clark AM, et al. Long-term safety and sustained efficacy of USL255 (topiramate extended-release capsules) in patients with refractory partial-onset seizures. Epilepsy Behav. 2016;59:13-20.

347. Collart Dutilleul P, Ryvlin P, Kahane P, Vercueil L, Semah F, Biraben A, et al. Exploratory Phase II Trial to Evaluate the Safety and the Antiepileptic Effect of Pitolisant (BF2.649) in Refractory Partial Seizures, Given as Adjunctive Treatment During 3 Months. Clin Neuropharmacol. 2016;39(4):188-93.

348. Krueger DA, Wilfong AA, Mays M, Talley CM, Agricola K, Tudor C, et al. Long-term treatment of epilepsy with everolimus in tuberous sclerosis. Neurology. 2016;87(23):2408-15.

349. Lim K-S, Lotay N, White R, Kwan P. Efficacy and safety of retigabine/ezogabine as adjunctive therapy in adult Asian patients with drug-resistant partial-onset seizures: A randomized, placebo-controlled Phase III study. Epilepsy Behav. 2016;61:224-30.

350. Moseley BD, Sperling MR, Asadi-Pooya AA, Diaz A, Elmouft S, Schiemann J, et al. Efficacy, safety, and tolerability of adjunctive brivaracetam for secondarily generalized tonic-clonic seizures: Pooled results from three Phase III studies. Epilepsy Res. 2016;127:179-85.

351. Overwater IE, Rietman AB, Bindels-de Heus K, Looman CW, Rizopoulos D, Sibindi TM, et al. Sirolimus for epilepsy in children with tuberous sclerosis complex: A randomized controlled trial. Neurology. 2016;87(10):1011-8.

352. Yu L, Liao WP, Yi YH, Qiu G. ABCB1 G2677T/A polymorphism is associated with the risk of drug-resistant epilepsy in Asians. Epilepsy Res. 2015;115:100-8.

353. Lv N, Qu J, Long H, Zhou L, Cao Y, Long L, et al. Association study between polymorphisms in the CACNA1A, CACNA1C, and CACNA1H genes and drug-resistant epilepsy in the Chinese Han population. Seizure. 2015;30:64-9.

354. Bertok S, Dolzan V, Goricar K, Podkrajsek KT, Battelino T, Rener-Primec Z. The association of SCN1A p.Thr1067Ala polymorphism with epilepsy risk and the response to antiepileptic drugs in Slovenian children and adolescents with epilepsy. Seizure. 2017;51:9-13.

355. Butilă AT, Zazgyva A, Sin AI, Szabo ER, Tilinca MC. GABRG2 C588T gene polymorphisms might be a predictive genetic marker of febrile seizures and generalized recurrent seizures: a case-control study in a Romanian pediatric population. Arch Med Sci. 2018;14(1):157-66.

356. Asadi-Pooya AA, Emami M, Ashjazadeh N, Nikseresht A, Shariat A, Petramfar P, et al. Reasons for uncontrolled seizures in adults; the impact of pseudointractability. Seizure. 2013;22(4):271-4.

357. Kutlu G, Erdal A, Gomceli YB, Inan LE. Pseudo-refractory epilepsy. Neurosciences (Riyadh). 2013;18(3):284-6.

358. Trinka E. Ideal characteristics of an antiepileptic drug: how do these impact treatment decisions for individual patients? Acta Neurol Scand. 2012;126(s194):10-8.

359. Brigo F, Igwe SC, Lattanzi S. Ethosuximide, sodium valproate or lamotrigine for absence seizures in children and adolescents. Cochrane Database Syst Rev. 2019;2:CD003032.

360. Glauser TA, Cnaan A, Shinnar S, Hirtz DG, Dlugos D, Masur D, et al. Ethosuximide, Valproic Acid, and Lamotrigine in Childhood Absence Epilepsy. The New England journal of medicine. 2010;362(9):790-9.

361. Marson AG, Al-Kharusi AM, Alwaidh M, Appleton R, Baker GA, Chadwick DW, et al. The SANAD study of effectiveness of valproate, lamotrigine, or topiramate for generalised and unclassifiable epilepsy: an unblinded randomised controlled trial. Lancet. 2007;369(9566):1016-26.

362. Veroniki AA, Cogo E, Rios P, Straus SE, Finkelstein Y, Kealey R, et al. Comparative safety of anti-epileptic drugs during pregnancy: a systematic review and network meta-analysis of congenital malformations and prenatal outcomes. BMC Med. 2017;15(1):95.

363. Veroniki AA, Rios P, Cogo E, Straus SE, Finkelstein Y, Kealey R, et al. Comparative safety of antiepileptic drugs for neurological development in children exposed during pregnancy and breast feeding: a systematic review and network meta-analysis. BMJ Open. 2017;7(7):e017248.

364. Bromley RL, Weston J, Marson AG. Maternal Use of Antiepileptic Agents During Pregnancy and Major Congenital Malformations in ChildrenMaternal Use of Antiepileptic Agents and Major Congenital Malformations in ChildrenMaternal Use of Antiepileptic Agents and Major Congenital Malformations in Children. JAMA 2017;318(17):1700-1.

365. Tomson T, Battino D, Bonizzoni E, Craig J, Lindhout D, Perucca E, et al. Comparative risk of major congenital malformations with eight different antiepileptic drugs: a prospective cohort study of the EURAP registry. Lancet Neurol. 2018;17(6):530-8.

366. National Institute for Health and Care Excellence Clinical guideline [CG185] Bipolar disorder: assessment and management [Internet]. NICE. 2014 [cited 15.8.2019]. Available from: https://www.nice.org.uk/guidance/cg185/chapter/1-Recommendations#managing-bipolar-disorder-in-adults-in-the-longer-term-in-secondary-care-2.

367. European Medicines Agency. New measures to avoid valproate exposure in pregnancy endorsed [Internet]. 2018 [cited 15.8.2019]. Available from: https://www.ema.europa.eu/en/documents/referral/valproate-article-31-referral-new-measures-avoid-valproate-exposure-pregnancy-endorsed_en-0.pdf.

368. Shakespeare J, Sisodiya SM. Guidance Document on Valproate Use in Women and Girls of Childbearing Years. Royal College of General Practitioners and Association of British Neurologists and Royal College of Physicians; 2019.

369. Proposal for Revised Classification of Epilepsies and Epileptic Syndromes. Epilepsia. 1989;30(4):389-99.

370. Okudan ZV, Özkara Ç. Reflex epilepsy: triggers and management strategies. Neuropsychiatr Dis Treat. 2018;14:327-37.

371. WHO Collaborating Centre for Drug Statistics Methodology. Guidelines for ATC classification and DDD assignment 2013. Oslo, Norway: WHO; 2012.

372. Altman DG. Practical Statistics for Medical Research. London: Chapman and Hall/CRC; 1991.

373. Hao XT, Wong IS, Kwan P. Interrater reliability of the international consensus definition of drug-resistant epilepsy: a pilot study. Epilepsy Behav. 2011;22(2):388-90.

374. Lao O, Lu TT, Nothnagel M, Junge O, Freitag-Wolf S, Caliebe A, et al. Correlation between Genetic and Geographic Structure in Europe. Curr Biol. 2008;18(16):1241-8.

375. Nelis M, Esko T, Mägi R, Zimprich F, Zimprich A, Toncheva D, et al. Genetic Structure of Europeans: A View from the North–East. PLoS ONE. 2009;4(5):e5472.

376. Luciano AL, Shorvon SD. Results of treatment changes in patients with apparently drug-resistant chronic epilepsy. Ann Neurol. 2007;62(4):375-81.

377. Giussani G, Canelli V, Bianchi E, Erba G, Franchi C, Nobili A, et al. Long-term prognosis of epilepsy, prognostic patterns and drug resistance: a population-based study. Eur J Neurol. 2016;23(7):1218-27.

378. Greenberg DA, Subaran R. Blinders, phenotype, and fashionable genetic analysis: A critical examination of the current state of epilepsy genetic studies. Epilepsia. 2011;52(1):1-9.

379. Kraft P, Zeggini E, Ioannidis JP. Replication in genome-wide association studies. Stat Sci. 2009;24(4):561-73.

380. Barendse W. The effect of measurement error of phenotypes on genome wide association studies. BMC Genomics. 2011;12:232.

381. de Toffol B, de Zelicourt M, Vespignani H, Fagnani F, Laurendeau C, Levy-Bachelot L, et al. ESPERA study: applicability of the new ILAE criteria for antiepileptic drug resistance of focal epilepsies in current clinical practice. Epilepsy Behav. 2012;25(2):166-9.

382. Berg AT, Levy SR, Testa FM, D'Souza R. Remission of epilepsy after two drug failures in children: a prospective study. Ann Neurol. 2009;65(5):510-9.

383. Sperling MR, Abou-Khalil B, Harvey J, Rogin JB, Biraben A, Galimberti CA, et al. Eslicarbazepine acetate as adjunctive therapy in patients with uncontrolled partial-onset seizures: Results of a phase III, double-blind, randomized, placebo-controlled trial. Epilepsia. 2015;56(2):244-53.

384. French JA, Krauss GL, Biton V, Squillacote D, Yang H, Laurenza A, et al. Adjunctive perampanel for refractory partial-onset seizures. Randomized phase III study 304. 2012;79(6):589-96.

385. Klein P, Schiemann J, Sperling MR, Whitesides J, Liang W, Stalvey T, et al. A randomized, double-blind, placebo-controlled, multicenter, parallel-group study to evaluate the efficacy and safety of adjunctive brivaracetam in adult patients with uncontrolled partial-onset seizures. Epilepsia. 2015;56(12):1890-8.

386. Biton V, Berkovic SF, Abou-Khalil B, Sperling MR, Johnson ME, Lu S. Brivaracetam as adjunctive treatment for uncontrolled partial epilepsy in adults: a phase III randomized, double-blind, placebo-controlled trial. Epilepsia. 2014;55(1):57-66.

387. Hao X, Goldberg D, Kelly K, Stephen L, Kwan P, Brodie MJ. Uncontrolled epilepsy is not necessarily the same as drug-resistant epilepsy: differences between populations with newly diagnosed epilepsy and chronic epilepsy. Epilepsy Behav. 2013;29(1):4-6.

388. Chadwick D, Smith D. The misdiagnosis of epilepsy. BMJ. 2002;324(7336):495-6.

389. Hao XT, Kwan P. Update and Overview of the International League Against Epilepsy Consensus Definition of Drug-resistant Epilepsy. European Neurological Review. 2011;6(1):57-9.

390. Gilioli I, Vignoli A, Visani E, Casazza M, Canafoglia L, Chiesa V, et al. Focal epilepsies in adult patients attending two epilepsy centers: classification of drug-resistance, assessment of risk factors, and usefulness of "new" antiepileptic drugs. Epilepsia. 2012;53(4):733-40.

391. Nesbitt G, McKenna K, Mays V, Carpenter A, Miller K, Williams M. The Epilepsy Phenome/Genome Project (EPGP) informatics platform. Int J Med Inform. 2013;82(4):248-59.

392. Epi4K Consortium. De Novo Mutations in SLC1A2 and CACNA1A Are Important Causes of Epileptic Encephalopathies. Am J Hum Genet. 2016;99(2):287-98.

393. Manolio TA, Bailey-Wilson JE, Collins FS. Genes, environment and the value of prospective cohort studies. Nat Rev Genet. 2006;7(10):812-20.

394. Fisher RS, Blum DE, DiVentura B, Vannest J, Hixson JD, Moss R, et al. Seizure diaries for clinical research and practice: Limitations and future prospects. Epilepsy Behav. 2012;24(3):304-10.

395. Hoppe C, Poepel A, Elger CE. Epilepsy: Accuracy of patient seizure counts. Arch Neurol. 2007;64(11):1595-9.

396. Rugg-Gunn FJ, Harrison NA, Duncan JS. Evaluation of the accuracy of seizure descriptions by the relatives of patients with epilepsy. Epilepsy Res. 2001;43(3):193-9.

397. Hong MKH, Yao HHI, Pedersen JS, Peters JS, Costello AJ, Murphy DG, et al. Error rates in a clinical data repository: lessons from the transition to electronic data transfer—a descriptive study. BMJ Open. 2013;3(5):e002406.

398. de Zélicourt M, de Toffol B, Vespignani H, Laurendeau C, Lévy-Bachelot L, Murat C, et al. Management of focal epilepsy in adults treated with polytherapy in France: the direct cost of drug resistance (ESPERA study). Seizure. 2014;23(5):349-56.

399. Gomez-Ibañez A, McLachlan RS, Mirsattari SM, Diosy DC, Burneo JG. Prognostic factors in patients with refractory idiopathic generalized epilepsy. Epilepsy Res. 2017;130:69-73.

400. Shorvon S, Luciano AL. Prognosis of chronic and newly diagnosed epilepsy: revisiting temporal aspects. Curr Opin Neurol. 2007;20(2):208-12.

401. Neligan A, Bell GS, Sander JW, Shorvon SD. How refractory is refractory epilepsy? Patterns of relapse and remission in people with refractory epilepsy. Epilepsy Res. 2011;96(3):225-30.

402. Berg AT, Vickrey BG, Testa FM, Levy SR, Shinnar S, DiMario F, et al. How long does it take for epilepsy to become intractable? A prospective investigation. Ann Neurol. 2006;60(1):73-9.

403. Callaghan BC, Anand K, Hesdorffer D, Hauser WA, French JA. Likelihood of seizure remission in an adult population with refractory epilepsy. Ann Neurol. 2007;62(4):382-9.

404. Choi H, Heiman GA, Munger Clary H, Etienne M, Resor SR, Hauser WA. Seizure remission in adults with long-standing intractable epilepsy: an extended follow-up. Epilepsy Res. 2011;93(2-3):115-9.

405. Choi H, Hayat MJ, Zhang R, Hirsch LJ, Bazil CW, Mendiratta A, et al. Drug-resistant epilepsy in adults: Outcome trajectories after failure of two medications. Epilepsia. 2016;57(7):1152-60.

406. Neligan A, Bell GS, Elsayed M, Sander JW, Shorvon SD. Treatment changes in a cohort of people with apparently drug-resistant epilepsy: an extended follow-up. J Neurol Neurosurg Psychiatry. 2012;83(8):810-3.

407. St Louis EK. Monitoring Antiepileptic Drugs: A Level-Headed Approach. Curr Neuropharmacol. 2009;7(2):115-9.

408. Kwan P, Brodie MJ. Effectiveness of First Antiepileptic Drug. Epilepsia. 2001;42(10):1255-60.

409. WHO Collaborating Centre for Drug Statistics Methodology. Guidelines for ATC classification and DDD assignment 2018. Oslo, Norway: WHO; 2017.

410. Perucca E, Dulac O, Shorvon S, Tomson T. Harnessing the clinical potential of antiepileptic drug therapy: dosage optimisation. CNS Drugs. 2001;15(8):609-21.

411. Genome-Wide DNA Analysis BeadChips [Internet]. 2010 [cited 1.3.2013]. Available from: https://www.illumina.com/content/dam/illumina-marketing/documents/products/brochures/datasheet_omni_whole-genome_arrays.pdf.

412. LaFramboise T. Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances. Nucleic Acids Res. 2009;37(13):4181-93.

413. Reed E, Nunez S, Kulp D, Qian J, Reilly MP, Foulkes AS. A guide to genome-wide association analysis and post-analytic interrogation. Stat Med. 2015;34(28):3769-92.

414. Marchini J, Howie B. Genotype imputation for genome-wide association studies. Nat Rev Genet. 2010;11(7):499-511.

415. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. Nat Genet. 2007;39(7):906-13.

416. Li Y, Abecasis GR. Mach 1.0: Rapid haplotype reconstruction and missing genotype inference. Am J Hum Genet. 2006:S79:2290.

417. Biernacka JM, Tang R, Li J, McDonnell SK, Rabe KG, Sinnwell JP, et al. Assessment of genotype imputation methods. BMC Proc. 2009;3(Suppl 7):S5-S.

418. Browning BL, Browning SR. A Unified Approach to Genotype Imputation and Haplotype-Phase Inference for Large Data Sets of Trios and Unrelated Individuals. Am J Hum Genet. 2009;84(2):210-23.

419. IMPUTE2; Output file options [Internet]. 2014 [cited 23.1.2015]. Available from: https://mathgen.stats.ox.ac.uk/impute/output_file_options.html.

420. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 2009;5(6):e1000529.

421. Turner S, Armstrong LL, Bradford Y, Carlson CS, Crawford DC, Crenshaw AT, et al. Quality control procedures for genome-wide association studies. Curr Protoc Hum Genet. 2011;68(1):1-24.

422. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. Nat Protoc. 2010;5(9):1564-73.

423. Teo YY. Exploratory data analysis in large-scale genetic studies. Biostatistics. 2010;11(1):70-81.

424. Lee S, Kasif S, Weng Z, Cantor CR. Quantitative analysis of single nucleotide polymorphisms within copy number variation. PLoS One. 2008;3(12):e3906.

425. Wigginton JE, Cutler DJ, Abecasis GR. A note on exact tests of Hardy-Weinberg equilibrium. Am J Hum Genet. 2005;76(5):887-93.

426. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81(3):559-75.

427. Neale BM, Purcell S. The positives, protocols, and perils of genome-wide association. Am J Med Genet B Neuropsychiatr Genet. 2008;147B(7):1288-94.

428. Verma SS, de Andrade M, Tromp G, Kuivaniemi H, Pugh E, Namjou-Khales B, et al. Imputation and quality control steps for combining multiple genome-wide datasets. Front Genet. 2014;5:370.

429. Chanda P, Yuhki N, Li M, Bader JS, Hartz A, Boerwinkle E, et al. Comprehensive evaluation of imputation performance in African Americans. J Hum Genet. 2012;57(7):411-21.

430. Ellinghaus D, Schreiber S, Franke A, Nothnagel M. Current software for genotype imputation. Human genomics. 2009;3(4):371-80.

431. de Bakker PI, Ferreira MA, Jia X, Neale BM, Raychaudhuri S, Voight BF. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. Hum Mol Genet. 2008;17(R2):R122-8.

432. Zheng HF, Rong JJ, Liu M, Han F, Zhang XW, Richards JB, et al. Performance of genotype imputation for low frequency and rare variants from the 1000 genomes. PLoS One. 2015;10(1):e0116487.

433. Krithika S, Valladares-Salgado A, Peralta J, Escobedo-de La Peña J, Kumate-Rodríguez J, Cruz M, et al. Evaluation of the imputation performance of the program IMPUTE in an admixed sample from Mexico City using several model designs. BMC Med Genomics. 2012;5:12.

434. SNPTEST v2 [Internet]. 2015 [cited 15.9.2015]. Available from: https://mathgen.stats.ox.ac.uk/genetics_software/snptest/old/snptest.html.

435. Nothnagel M, Ellinghaus D, Schreiber S, Krawczak M, Franke A. A comprehensive evaluation of SNP genotype imputation. Hum Genet. 2009;125(2):163-71.

436. Ahmad M, Sinha A, Ghosh S, Kumar V, Davila S, Yajnik CS, et al. Inclusion of Population-specific Reference Panel from India to the 1000 Genomes Phase 3 Panel Improves Imputation Accuracy. Sci Rep. 2017;7(1):6733.

437. Kelly M, Fortes MRS, Moore SS. Accuracy of imputation in a population of tropical composite cattle with particular emphasis on the use of allelic r2 as a quality control metric. Proc Assoc Advmt Anim Breed Genet. 2013;20:550-3.

438. Shaffer JR, Li J, Lee MK, Roosenboom J, Orlova E, Adhikari K, et al. Multiethnic GWAS Reveals Polygenic Architecture of Earlobe Attachment. Am J Hum Genet. 2017;101(6):913-24.

439. Adhikari K, Fuentes-Guajardo M, Quinto-Sanchez M, Mendoza-Revilla J, Camilo Chacon-Duque J, Acuna-Alonzo V, et al. A genome-wide association scan implicates DCHS2, RUNX2, GLI3, PAX1 and EDAR in human facial variation. Nat Commun. 2016;7:11616.

440. Laurie CC, Doheny KF, Mirel DB, Pugh EW, Bierut LJ, Bhangale T, et al. Quality control and quality assurance in genotypic data for genome-wide association studies. Genet Epidemiol. 2010;34(6):591-602.

441. Zondervan KT, Cardon LR. Designing candidate gene and genome-wide case-control association studies. Nat Protoc. 2007;2(10):2492-501.

442. Stevens EL, Heckenberg G, Roberson ED, Baugher JD, Downey TJ, Pevsner J. Inference of relationships in population data using identity-by-descent and identity-by-state. PLoS Genet. 2011;7(9):e1002287.

443. Price AL, Zaitlen NA, Reich D, Patterson N. New approaches to population stratification in genome-wide association studies. Nat Rev Genet. 2010;11(7):459-63.

444. Li Q, Yu K. Improved correction for population stratification in genome-wide association studies by identifying hidden population structures. Genet Epidemiol. 2008;32(3):215-26.

445. Lewis CM. Genetic association studies: design, analysis and interpretation. Brief Bioinform. 2002;3(2):163-53.

446. Lippert C, Listgarten J, Liu Y, Kadie CM, Davidson RI, Heckerman D. FaST linear mixed models for genome-wide association studies. Nat Methods. 2011;8(10):833-5.

447. Johnson RC, Nelson GW, Troyer JL, Lautenberger JA, Kessing BD, Winkler CA, et al. Accounting for multiple comparisons in a genome-wide association study (GWAS). BMC Genomics. 2010;11:724.

448. Spain SL, Barrett JC. Strategies for fine-mapping complex traits. Hum Mol Genet. 2015;24(R1):R111-9.

449. Cookson W, Liang L, Abecasis G, Moffatt M, Lathrop M. Mapping complex disease traits with global gene expression. Nat Rev Genet. 2009;10(3):184-94.

450. Zhu Y, Tazearslan C, Suh Y. Challenges and progress in interpretation of non-coding genetic variants associated with human disease. Exp Biol Med (Maywood). 2017;242(13):1325-34.

451. Nürnberg ST, Rendon A, Smethurst PA, Paul DS, Voss K, Thon JN, et al. A GWAS sequence variant for platelet volume marks an alternative DNM3 promoter in megakaryocytes near a MEIS1 binding site. Blood. 2012;120(24):4859-68.

452. Corradin O, Scacheri PC. Enhancer variants: evaluating functions in common disease. Genome Med. 2014;6(10):85.

453. Hrdlickova B, de Almeida RC, Borek Z, Withoff S. Genetic variation in the non-coding genome: Involvement of micro-RNAs and long non-coding RNAs in disease. Biochim Biophys Acta. 2014;1842(10):1910-22.

454. Smemo S, Tena JJ, Kim KH, Gamazon ER, Sakabe NJ, Gomez-Marin C, et al. Obesity-associated variants within FTO form long-range functional connections with IRX3. Nature. 2014;507(7492):371-5.

455. Claussnitzer M, Dankel SN, Klocke B, Grallert H, Glunk V, Berulava T, et al. Leveraging cross-species transcription factor binding site patterns: from diabetes risk loci to disease mechanisms. Cell. 2014;156(1-2):343-58.

456. Hormozdiari F, van de Bunt M, Segre AV, Li X, Joo JWJ, Bilow M, et al. Colocalization of GWAS and eQTL Signals Detects Target Genes. Am J Hum Genet. 2016;99(6):1245-60.

457. Gamazon ER, Wheeler HE, Shah KP, Mozaffari SV, Aquino-Michaels K, Carroll RJ, et al. A gene-based association method for mapping traits using reference transcriptome data. Nat Genet. 2015;47(9):1091-8.

458. Wooten EC, Huggins GS. Mind the dbGAP: The Application of Data Mining to Identify Biological Mechanisms. Mol Interv. 2011;11(2):95-102.

459. Howie B, Marchini J, Stephens M. Genotype imputation with thousands of genomes. G3 (Bethesda). 2011;1(6):457-70.

460. Pritchard JK, Stephens M, Donnelly P. Inference of Population Structure Using Multilocus Genotype Data. Genetics. 2000;155(2):945-59.

461. Delaneau O, Marchini J, Zagury J-F. A linear complexity phasing method for thousands of genomes. Nat Methods. 2011;9(2):179-81.

462. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat Genet. 2012;44(8):955-9.

463. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2013.

464. Menashe I, Rosenberg PS, Chen BE. PGA: power calculator for case-control genetic association analyses. BMC Genetics. 2008;9(1):36.

465. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics. 2010;26(18):2336-7.

466. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res. 2012;40(Database issue):D930-4.

467. HaploReg v4.1 [Internet]. Broad Institute. 2015 [cited 1.7.2018]. Available from: https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php.

468. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. Genome Biol. 2016;17(1):122.

469. Variant Effect Predictor [Internet]. EMBL-EBI 2018 [cited 1.10.2018]. Available from: http://www.ensembl.org/Homo_sapiens/Tools/VEP?db=core;tl=GxgGPPHWa6Ls4Mfv-4713691.

470. GTEx Consortium, Laboratory Data Analysis & Coordinating Center (LDAAC)—Analysis Working Group, Statistical Methods groups—Analysis Working Group, Enhancing GTEx (eGTEx) groups, NIH Common Fund, NIH/NCI, et al. Genetic effects on gene expression across human tissues. Nature. 2017;550(7675):204-13.

471. The Genotype-Tissue Expression (GTEx) Project [Internet]. [cited 15.2.2019]. Available from: https://gtexportal.org/home/.

472. Eppig JT, Blake JA, Bult CJ, Kadin JA, Richardson JE, Mouse Genome Database Group. The Mouse Genome Database (MGD): facilitating mouse as a model for human biology and disease. Nucleic Acids Res. 2015;43(Database issue):D726-D36.

473. Mouse Genome Informatics [Internet]. 2019 [cited 19.2.2019]. Available from: http://www.informatics.jax.org/.

474. Ng B, White CC, Klein H-U, Sieberts SK, McCabe C, Patrick E, et al. An xQTL map integrates the genetic architecture of the human brain's transcriptome and epigenome. Nat Neurosci. 2017;20(10):1418-26.

475. Brain xQTL Serve [Internet]. 2017 [cited 15.2.2019]. Available from: http://mostafavilab.stat.ubc.ca/xqtl/.

476. Ramasamy A, Trabzuni D, Guelfi S, Varghese V, Smith C, Walker R, et al. Genetic variability in the regulation of gene expression in ten regions of the human brain. Nat Neurosci. 2014;17(10):1418-28.

477. Braineac - The Brain eQTL Almanac [Internet]. UK Brain Expression Consortium (UKBEC). 2013 [cited 1.10.2018].

478. Lopez-Garcia MA, Feria-Romero IA, Serrano H, Rayo-Mares D, Fagiolino P, Vazquez M, et al. Influence of genetic variants of CYP2D6, CYP2C9, CYP2C19 and CYP3A4 on antiepileptic drug metabolism in pediatric patients with refractory epilepsy. Pharmacol Rep. 2017;69(3):504-11.

479. Bao Y, Liu X, Xiao Z. Association between two SCN1A polymorphisms and resistance to sodium channel blocking AEDs: a meta-analysis. Neurol Sci. 2018;39(6):1065-72.

480. Shi L, Zhu M, Li H, Wen Z, Chen X, Luo J, et al. SCN1A and SCN2A polymorphisms are associated with response to valproic acid in Chinese epilepsy patients. Eur J Clin Pharmacol. 2019;75(5):655-63.

481. Sinnott JA, Kraft P. Artifact due to differential error when cases and controls are imputed from different platforms. Hum Genet. 2012;131(1):111-9.

482. Xu J, Zhou W, Yang F, Chen G, Li H, Zhao Y, et al. The β-TrCP-FBXW2-SKP2 axis regulates lung cancer cell growth with FBXW2 acting as a tumour suppressor. Nat Commun. 2017;8:14002-.

483. Gstaiger M, Jordan R, Lim M, Catzavelos C, Mestan J, Slingerland J, et al. Skp2 is oncogenic and overexpressed in human cancers. Proceedings of the National Academy of Sciences of the United States of America. 2001;98(9):5043-8.

484. Chen F, Lin X, Xu P, Zhang Z, Chen Y, Wang C, et al. Nuclear Export of Smads by RanBP3L Regulates Bone Morphogenetic Protein Signaling and Mesenchymal Stem Cell Differentiation. Mol Cell Biol. 2015;35(10):1700-11.

485. Ohashi K, Kawai S, Murata K. Identification and characterization of a human mitochondrial NAD kinase. Nat Commun. 2012;3:1248-.

486. Takai S, Yamada K, Kawakami H, Tanaka K, Nakamura S. Localization of the gene (SLC1A3) encoding human glutamate transporter (GluT-1) to 5p13 by fluorescence in situ hybridization. Cytogenet Cell Genet. 1995;69(3-4):209-10.

487. Jen JC, Wan J, Palos TP, Howard BD, Baloh RW. Mutation in the glutamate transporter EAAT1 causes episodic ataxia, hemiplegia, and seizures. Neurology. 2005;65(4):529-34.

488. Park JH, Wacholder S, Gail MH, Peters U, Jacobs KB, Chanock SJ, et al. Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. Nat Genet. 2010;42(7):570-5.

489. Gibson G. Rare and common variants: twenty arguments. Nat Rev Genet. 2012;13(2):135-45.

490. Leu C, Balestrini S, Maher B, Hernández-Hernández L, Gormley P, Hämäläinen E, et al. Genome-wide Polygenic Burden of Rare Deleterious Variants in Sudden Unexpected Death in Epilepsy. EBioMedicine. 2015;2(9):1063-70.

491. Schork NJ, Murray SS, Frazer KA, Topol EJ. Common vs. rare allele hypotheses for complex diseases. Curr Opin Genet Dev. 2009;19(3):212-9.

492. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic localization of common disease-associated variation in regulatory DNA. Science. 2012;337(6099):1190-5.

493. Schaub MA, Boyle AP, Kundaje A, Batzoglou S, Snyder M. Linking disease associations with regulatory information in the human genome. Genome research. 2012;22(9):1748-59.

494. Nakayama Y, Nara N, Kawakita Y, Takeshima Y, Arakawa M, Katoh M, et al. Cloning of cDNA Encoding a Regeneration-Associated Muscle Protease Whose Expression Is Attenuated in Cell Lines Derived from Duchenne Muscular Dystrophy Patients. Am J Pathol. 2004;164(5):1773-82.

495. Lo PH, Tanikawa C, Katagiri T, Nakamura Y, Matsuda K. Identification of novel epigenetically inactivated gene PAMR1 in breast carcinoma. Oncol Rep. 2015;33(1):267-73.

496. Visser M, Kayser M, Palstra R-J. HERC2 rs12913832 modulates human pigmentation by attenuating chromatin-loop formation between a long-range enhancer and the OCA2 promoter. Genome research. 2012;22(3):446-55.

497. Tanaka K, Watase K, Manabe T, Yamada K, Watanabe M, Takahashi K, et al. Epilepsy and Exacerbation of Brain Injury in Mice Lacking the Glutamate Transporter GLT-1. Science. 1997;276(5319):1699-702.

498. Colton CK, Kong Q, Lai L, Zhu MX, Seyb KI, Cuny GD, et al. Identification of translational activators of glial glutamate transporter EAAT2 through cell-based high-throughput screening: an approach to prevent excitotoxicity. J Biomol Screen. 2010;15(6):653-62.

499. Ulu MO, Tanriverdi T, Oz B, Biceroglu H, Isler C, Eraslan BS, et al. The expression of astroglial glutamate transporters in patients with focal cortical dysplasia: an immunohistochemical study. Acta Neurochir (Wien). 2010;152(5):845-53.

500. Zhang Y, Dong H, Duan L, Yuan G, Liang W, Li Q, et al. SLC1A2 mediates refractory temporal lobe epilepsy with an initial precipitating injury by targeting the glutamatergic synapse pathway. IUBMB Life. 2019;71(2):213-22.

501. Rothstein JD, Dykes-Hoberg M, Pardo CA, Bristol LA, Jin L, Kuncl RW, et al. Knockout of Glutamate Transporters Reveals a Major Role for Astroglial Transport in Excitotoxicity and Clearance of Glutamate. Neuron. 1996;16(3):675-86.

502. Jiang J, Amara SG. New views of glutamate transporter structure and function: advances and challenges. Neuropharmacology. 2011;60(1):172-81.

503. Fairman WA, Vandenberg RJ, Arriza JL, Kavanaught MP, Amara SG. An excitatory amino-acid transporter with properties of a ligand-gated chloride channel. Nature. 1995;375(6532):599-603.

504. Winter N, Kovermann P, Fahlke C. A point mutation associated with episodic ataxia 6 increases glutamate transporter anion currents. Brain. 2012;135(11):3416-25.

505. Kumar RG, Breslin KB, Ritter AC, Conley YP, Wagner AK. Variability with Astroglial Glutamate Transport Genetics Is Associated with Increased Risk for Post-Traumatic Seizures. J Neurotrauma. 2019;36(2):230-8.

506. de Vries B, Mamsa H, Stam AH, Wan J, Bakker SLM, Vanmolkot KRJ, et al. Episodic Ataxia Associated With EAAT1 Mutation C186S Affecting Glutamate Reuptake. Archives of neurology. 2009;66(1):97-101.

507. Rethoré MO, de Blois MC, Peeters M, Popowski P, Pangalos C, Lejeune J. Pure partial trisomy of the short arm of chromosome 5. Hum Genet. 1989;82(3):296-8.

508. Lorda-Sánchez I, Urioste M, Villa A, Carrascosa MdC, Vásquez MS, Martínez A, et al. Proximal partial 5p trisomy resulting from a maternal (19;5) insertion. Am J Med Genet. 1997;68(4):476-80.

509. Avansino JR, Dennis TR, Spallone P, Stock AD, Levin ML. Proximal 5p trisomy resulting from a marker chromosome implicates band 5p13 in 5p trisomy syndrome. Am J Med Genet. 1999;87(1):6-11.

510. Yasutomo K, Suzue T, Nishioka A, Kozan H, Sekiguchi T, Ohara K, et al. Partial trisomy for short arm of chromosome 5. Am J Med Genet. 1993;35(4):336-9.

511. van Amen-Hellebrekers CJ, Jansen S, Pfundt R, Schuurs-Hoeijmakers JH, Koolen DA, Marcelis CL, et al. Duplications of SLC1A3: Associated with ADHD and autism. Eur J Med Genet. 2016;59(8):373-6.

512. Laurin N, Lee J, Ickowicz A, Pathare T, Malone M, Tannock R, et al. Association study for genes at chromosome 5p13-q11 in attention deficit hyperactivity disorder. Am J Med Genet B Neuropsychiatr Genet. 2008;147B(5):600-5.

513. Armstrong ME, Weaver DD, Lah MD, Vance GH, Landis BJ, Ware SM, et al. Novel phenotype of 5p13.3-q11.2 duplication resulting from supernumerary marker chromosome 5: implications for management and genetic counseling. Mol Cytogenet. 2018;11:23.

514. Velagaleti GVN, Morgan DL, Tonk VS. Trisomy 5p. A case report and review. Ann Genet. 2000;43(3):143-5.

515. Camerota L, Pitzianti M, Postorivo D, Nardone AM, Ligas C, Moretti C, et al. A Small Supernumerary Marker Derived from the Pericentromeric Region of Chromosome 5: Case Report and Delineation of Partial Trisomy 5p Phenotype. Cytogenet Genome Res. 2017;153(1):22-8.

516. Gilby KL, O'Brien TJ. Epilepsy, autism, and neurodevelopment: kindling a shared vulnerability? Epilepsy Behav. 2013;26(3):370-4.

517. Williams AE, Giust JM, Kronenberger WG, Dunn DW. Epilepsy and attention-deficit hyperactivity disorder: links, risks, and challenges. Neuropsychiatr Dis Treat. 2016;12:287-96.

518. McCusker CG, Kennedy PJ, Anderson J, Hicks EM, Hanrahan D. Adjustment in children with intractable epilepsy: importance of seizure duration and family factors. Dev Med Child Neurol. 2002;44(10):681-7.

519. Sansa G, Carlson C, Doyle W, Weiner HL, Bluvstein J, Barr W, et al. Medically refractory epilepsy in autism. Epilepsia. 2011;52(6):1071-5.

520. Mattick JS, Rinn JL. Discovery and annotation of long noncoding RNAs. Nat Struct Mol Biol. 2015;22(1):5-7.

521. Peschansky VJ, Wahlestedt C. Non-coding RNAs as direct and indirect modulators of epigenetic regulation. Epigenetics. 2014;9(1):3-12.

522. Penny GD, Kay GF, Sheardown SA, Rastan S, Brockdorff N. Requirement for Xist in X chromosome inactivation. Nature. 1996;379(6561):131-7.

523. Anttila V, Bulik-Sullivan B, Finucane HK, Walters RK, Bras J, Duncan L, et al. Analysis of shared heritability in common disorders of the brain. Science. 2018;360(6395):eaap8757.

524. Whelan CD, Altmann A, Botia JA, Jahanshad N, Hibar DP, Absil J, et al. Structural brain abnormalities in the common epilepsies assessed in a worldwide ENIGMA study. Brain. 2018;141(2):391-408.

525. Speed D, Hemani G, Johnson MR, Balding DJ. Improved heritability estimation from genome-wide SNPs. Am J Hum Genet. 2012;91(6):1011-21.

526. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011;88(1):76-82.

527. Yang J, Weedon MN, Purcell S, Lettre G, Estrada K, Willer CJ, et al. Genomic inflation factors under polygenic inheritance. Eur J Hum Genet. 2011;19(7):807-12.

528. Browning SR, Browning BL. Population structure can inflate SNP-based heritability estimates. Am J Hum Genet. 2011;89(1):191-3; author reply 3-5.

529. Speed D, Cai N, UCLEB Consortium, Johnson MR, Nejentsev S, Balding DJ. Reevaluation of SNP heritability in complex human traits. Nat Genet. 2017;49(7):986-92.

530. Guggenheim JA, St Pourcain B, McMahon G, Timpson NJ, Evans DM, Williams C. Assumption-free estimation of the genetic contribution to refractive error across childhood. Mol Vis. 2015;21:621-32.

APPENDICES

Appendix 1. EpiPGX CRF

1. **General data**

- Site code (e.g. UCL, ULB…):
- Date of CRF completion:
- Person entering data:
- Data source (tick one):
    - o medical records
    - o database (specify):
    - o other (specify):
- DNA nr:
- DNA source (tick one):
    - o blood
    - o saliva
    - o brain tissue
    - o other (specify):
- Genotyped : yes / no
  If yes: -Platform:
  -Imputed: yes / no
- Gender: male / female
- DOB:
- Ethnicity: European, African, Asian, other (specify), mixed (specify), unclassified
- Date of recruitment / DNA collection:
- Date of epilepsy diagnosis:
- Start date of continuous contemporary clinical records:
- Status at start of continuous contemporary clinical records (tick one):
    - o new epilepsy
    - o existing epilepsy, off treatment
    - o existing epilepsy, off treatment but previous AED treatment (<12 months / >12 months)

- existing epilepsy, on treatment

- Date of latest recorded visit:

2. Epilepsy diagnosis

- Epilepsy syndrome according to 1989 ILAE classification:
- Hippocampal sclerosis: yes / no
  - If yes: - left / right
    - confirmed by: MRI / histology / unknown

3. Known progressive neurological disorder

- Yes / no
- If yes: - Type (tick one):
  - Neoplastic/paraneoplastic
  - Metabolic
  - Infectious
  - Inflammatory
  - Degenerative
  - Genetic
  - Other
  - Details:
  - Onset date:

4. Neurological examination

- Normal / abnormal / NA
  - If abnormal: tick one or more:
    - Higher cortical functions
    - Speech disturbance
    - Cranial nerve abnormalities

- Motor abnormalities
- Sensory abnormalities
- Coordination
- Other

- Details:

5. Seizures

- Seizure types (tick any that apply):
  - primary generalized tonic clonic (GTC)
  - absence
  - clonic
  - tonic
  - atonic
  - myoclonic
  - simple partial
  - complex partial
  - secondarily GTC
  - unclassified partial
  - unclassified GTC
  - uncertain epileptic
  - non-epileptic

6. Seizure frequency

- Date/year of first ever seizure:
- Total number of seizures prior to first ever AED
  1) GTC seizures:
     - Absolute number (if known):
     - Categorical: 1-2, 3-5, 6-10, 11-20, 21+, unknown
  2) Non-GTC seizures (any type):

- Absolute number (if known):

- Categorical: 1-2, 3-5, 6-10, 11-20, 21+, unknown

3) Combined (if type unknown):

- Absolute number (if known):

- Categorical: 1-2, 3-5, 6-10, 11-20, 21+, unknown

- Did the patient experience at least one seizure in the 12 months prior to starting the first AED? Yes / No / Unknown

- Any periods of ≥12 months remission: yes / no / unknown

If yes: - Remission 1: - start date:

- stop date:

- Remission 2: - start date:

- stop date:

- Remission 3: - start date:

- stop date:

- Remission 4: - start date:

- stop date:

- Total number of seizures in the last 12 months prior to latest recorded visit

1) GTC seizures:

- Absolute number (if known):

- Categorical: 1-3, ≥4, unknown

2) Non-GTC seizures (any type):

- Absolute number (if known):

- Categorical: 1-3, ≥4, unknown

3) Combined (if type unknown):

- Absolute number (if known):

- Categorical: 1-3, ≥4, unknown

- Total number of seizures in the last 12 months prior to non-medical epilepsy treatment (if applicable)

1) GTC seizures:

- Absolute number (if known):

- Categorical: 1-3, ≥4, unknown

2) Non-GTC seizures (any type):

- Absolute number (if known):
- Categorical: 1-3, ≥4, unknown

3) Combined (if type unknown):

- Absolute number (if known):
- Categorical: 1-3, ≥4, unknown

7. Non-medical epilepsy treatment (surgery, VNS…)

- Date of procedure:
- Type:

8. Investigations 1 (first ever)

- EEG:
  - normal
  - abnormal (epileptiform)
  - abnormal (non-specific)
  - not done
  - not known
- Imaging (MRI / CT – delete as appropriate):
  - normal
  - abnormal (focal)
  - abnormal (non-specific)
  - not done
  - not known

9. Investigations 2 (most relevant, if different from above)

- EEG:
  - normal

- abnormal (epileptiform)
- abnormal (non-specific)
- not done
- not known

- Imaging (MRI / CT – delete as appropriate):
  - normal
  - abnormal (focal)
  - abnormal (non-specific)
  - not done
  - not known

10. AED history per AED (*Fill out 1 form per AED tried, in chronological order)*

- AED generic name:
- Start date:
- Stop date (NA if patient still on AED):
- Patient adherent for this AED: yes / no / unknown
- Initiated as monotherapy / add-on / unknown (delete as appropriate)
- Maximum dose reached (mg/d):
- Serum level at maximum dose (µg/ml):
- Average monthly seizure frequency during ≥3 months before starting this AED
    1) GTC seizures:
    2) Non-GTC seizures (any type):
    3) Combined (if type unknown):
- Average monthly seizure frequency while on this AED (and prior to any subsequent change in epilepsy treatment)
    1) GTC seizures:
    2) Non-GTC seizures (any type):
    3) Combined (if type unknown):
- Reason for stopping:
    - inadequate seizure control
    - unacceptable adverse effects
    - both inadequate seizure control and unacceptable adverse effects
    - remission
    - other (specify):
    - unknown
    - NA
- Can this AED trial be considered appropriate and adequate: yes / no / unknown
- Outcome of this AED trial: response / failure / extreme late response / unclassified / unknown

- **Adverse drug reactions**: yes / no / unknown

  If yes (tick any, label "ADR1, 2 …" and fill out details below for each ADR):

  - Behavioural disorder (agitation / aggression / irritability / confusion). Ascertained through:
    - Prospective neuropsychological assessment
    - Retrospective contemporary data
  - Cognitive impairment (amnesia / forgetfulness / concentration difficulties / slowed mentation). Ascertained through:
    - Prospective neuropsychological assessment
    - Retrospective contemporary data
  - Hepatic dysfunction
    - Highest GOT/AST (IU/l + reference values):
    - Highest GPT/ALT (IU/l + reference values):
    - First elevated AP (IU/l + reference values):
    - Highest AP (IU/l + reference values):
    - First elevated bilirubin (mg/dl + reference values):
    - Highest bilirubin (mg/dl + reference values):
    - Highest PT (seconds):
  - Hyponatraemia / SIADH
    - Lowest plasma Na+ (mEq/l):
  - Neutropenia / agranulocytosis
    - Lowest absolute neutrophil count (/µl):
  - Psychosis
    - Psychosis according to ICD10 definition
    - Confirmed by psychiatrist: yes / no / unknown
  - Cutaneous adverse reactions
    - Confirmed by lymphocyte transformation test: yes / no / unknown
    - Confirmed by dermatologist: yes / no / unknown

    Type:
    - Maculopapular exanthema
    - Hypersensitivity syndrome (tick any that apply):

- ☐ Prolonged recovery phase (despite AED withdrawal)
- ☐ Fever
- ☐ Internal organ involvement (tick + give details):
  - ☐ Liver
  - ☐ Gastro-intestinal
  - ☐ Kidney
  - ☐ Lung
  - ☐ Central nervous system
  - ☐ Heart
  - ☐ Muscle
  - ☐ Thyroid
  - ☐ Haematological
  - ☐ Lymphoid system
- o Stevens-Johnson syndrome
- o Toxic epidermal necrolysis

o Speech disorder
- o Speech difficulties witnessed by physician

o Thrombocytopenia
- o Lowest thrombocyte count (/µl):

o Tremor
- o Mild / moderate / severe
- o Family history of tremor: yes / no / unknown

o Visual field constriction
- o Confirmed by : Goldmann / Humphrey / OCT

o Weight change
- o Weight before AED: kg
- o Weight after AED: kg
- o Weight change: kg

o Miscellaneous
- o Cardiac conduction abnormality / syncope
- o Depressed mood / depression

- o Encephalopathy
- o Other:

Fill out details for each of the above ADRs:

- ADR1 :

- ADR dosing at time of reaction (mg/d):
- ADR start date:
- ADR stop date (NA if ongoing):
- Can ADR be reasonably attributed to AED: yes / no / unknown
- Did ADR lead to dose reduction: yes / no / unknown
- Did ADR lead to ADR withdrawal: yes / no / unknown

- ADR2 :

- ADR dosing at time of reaction (mg/d):
- ADR start date:
- ADR stop date (NA if ongoing):
- Can ADR be reasonably attributed to AED: yes / no / unknown
- Did ADR lead to dose reduction: yes / no / unknown
- Did ADR lead to ADR withdrawal: yes / no / unknown

- ADR3 :

- ADR dosing at time of reaction (mg/d):
- ADR start date:
- ADR stop date (NA if ongoing):
- Can ADR be reasonably attributed to AED: yes / no / unknown
- Did ADR lead to dose reduction: yes / no / unknown
- Did ADR lead to ADR withdrawal: yes / no / unknown

- ADR4 :

- ADR dosing at time of reaction (mg/d):

- ADR start date:
- ADR stop date (NA if ongoing):
- Can ADR be reasonably attributed to AED: yes / no / unknown
- Did ADR lead to dose reduction: yes / no / unknown
- Did ADR lead to ADR withdrawal: yes / no / unknown

11. <u>Pregnancy</u> (*Fill out 1 form per pregnancy, in chronological order)*

- Date of outcome of pregnancy:
- Number of this pregnancy:
- Child number (A, B…):
- Personal history of major congenital malformation: yes / no / unknown
  - If yes: description / classification of MCM:
- Family history of MCM: yes / no / unknown
  - If yes: description / classification of MCM:
- Folic acid taken: yes / no / unknown
  - If yes: - Preconceptual folic acid: yes / no / unknown
    - Dose (mg/d):
- List + dose of AEDs taken prior to conception and during first 3 months:
  -
  -
  -
  -
- Other drugs taken during pregnancy: yes / no / unknown
  - If yes, provide drug (generic) names and doses:
    -
    -
    -
    -
- Other drugs taken in 3 months before conception: yes / no / unknown
  - If yes, provide drug (generic) names and doses:
    -
    -
    -
    -
- Alcohol use in 1st trimester: yes / no / unknown
- Smoking in pregnancy: yes / no / unknown

- Gestational age at birth (weeks):
- Birth weight (grams):
- Sex: male / female / unknown
- Outcome (tick one):
    - completed
    - miscarriage
    - induced abortion
    - stillbirth
    - ongoing
- Major congenital malformation: yes / no / unknown
    - If yes (tick any):
        - cardiac malformation
        - cleft palate
        - facial dysmorphism (other than cleft palate)
        - gastro-intestinal tract defect
        - genito-urinary tract defect
        - neural tube defect / spina bifida
        - skeletal malformation
        - other (specify):
        - unspecified
- Neurodevelopmental delay: yes / no / unknown

12. Summary *(tick any that apply)*

- ☐ Patient has generalized epilepsy
- ☐ Patient has focal epilepsy
- ☐ WP02: Patient has newly-diagnosed epilepsy: yes / no / unknown

  If yes:
  - Immediate remission
  - Deferred remission
  - No remission
  - First AED failed due to inefficacy
  - First AED failed due to ADRs
- ☐ Patient can be included in WP03: yes / no / unknown
  - Is drug resistant
    - Total number of adequate, appropriate and tolerated AEDs:
  - Is drug responsive
  - Is extremely drug resistant
  - Is a control for extremely drug resistant
- ☐ Patient can be included in WP04 (list AED names and outcome below): yes / no / unknown

  Include only patients who have failed at least one AED trial due to lack of efficacy!

  1) ..... : response / failure / extreme late response
  2) ..... : response / failure / extreme late response
  3) ..... : response / failure / extreme late response
  4) ..... : response / failure / extreme late response
  5) ..... : response / failure / extreme late response
  6) ..... : response / failure / extreme late response
  7) ..... : response / failure / extreme late response
  8) ..... : response / failure / extreme late response
  9) ..... : response / failure / extreme late response
  10) ..... : response / failure / extreme late response
  11) ..... : response / failure / extreme late response

12) …… : response / failure / extreme late response

Responder to VPA + LTG in combination only: yes / no / unknown / NA

☐ Patient can be included in WP05 (list AED names and ADRs below)

AED name	ADR

☐ Patient can be included in WP06 (list AEDs below)

1) ……: case / control
2) ……: case / control
3) ……: case / control
4) ……: case / control
5) ……: case / control

Appendix 2. Consensus definitions for phenotypic categories

Phenotype definitions

WP 02 – Genome-based biomarkers of early treatment response in newly-diagnosed epilepsy

Only patients with newly diagnosed epilepsy will be included for this purpose.
- **Newly diagnosed epilepsy** is defined as the occurrence of ≥2 clinically definite unprovoked epileptic seizures in the previous year, *or* the occurrence of one seizure and the clinician decides to start AEDs.
    - Prospective data are preferred, but retrospective data are allowed if based on contemporary evidence (i.e. continuous records from initiation of the first AED onwards from a specialist epilepsy centre).
    - Patients with known progressive neurological disorders at time of first AED initiation are excluded
    - Make a note of those patients with prior AED exposure or rescue treatment + the indication.
- **Focal & generalized epilepsy** are defined as in the 1989 ILAE classification

Task 1: Identifying genome-based biomarkers of remission with *first well-tolerated* drug.

- **Remission** is defined as any continuous period of ≥12 months complete seizure freedom. A titration period of 2 months is taken into account.
- **Immediate remission** is defined as remission within 14 months of starting the first well-tolerated AED.
- **Deferred remission** is defined as remission that is first recorded later than 14 months after starting the first well-tolerated AED (and prior to initiation of another AED).
- **No remission** is defined as continuing seizures after starting the first *well-tolerated*, *adequately applied* and *appropriate* AED (see Appendix for guidance).

Task 2: Identifying genome-based biomarkers that distinguish general and selective drug responsiveness.

No additional definitions needed here.

Task 3: Identifying genomic biomarkers of first drug failure.

- **Treatment failure (withdrawal) due to lack of efficacy** is defined as continuing seizures after the first *appropriate* AED has been *adequately* applied. Where seizure frequency data is not available, there should be clear written evidence that the AED was withdrawn specifically because it failed to control seizures.
- **Treatment failure due to ADRs**: In order to be attributed to the AED in question, ADRs should (i) occur within 6 months of initiation of the AED (not applicable for visual field defects), (ii) lead to withdrawal of the AED, and (iii) not be attributed to another cause by the treating clinician or the phenotyping clinician.

ADRs are sub-classified into 'on-target' and 'off-target' reactions:

- **On-target** : neurological in origin, related to dose or concentration, associated with dose increase, resolve on dose reduction or drug withdrawal.
- **Off-target:** non-neurological, not necessarily related to dose or concentration, not necessarily associated with dose increase, do not necessarily resolve on dose reduction or drug withdrawal.

WP 03 – Genome-based biomarker discovery for broad AED resistance

Task 2: Undertaking GWAS for broad drug resistance

- **Broad AED resistance** is defined as seizures recurring at a frequency of ≥4/year over the last year till latest recorded visit, despite *adequate* trials of ≥2 *tolerated* and *appropriately chosen and used* AED schedules, whether as monotherapies or in combination (see Appendix for guidance).
- **Drug responsiveness** is defined as freedom from seizures for ≥12 months up to latest recorded visit.

  - Patients who have had epilepsy surgery and fulfilled the above criteria for broad AED resistance before surgery can also be included.
  - Patients who have had epilepsy surgery can never be classified as drug responsive thereafter
  - Patients known to be systematically non-adherent should be excluded. (1 seizure a year due to non-adherence may be disregarded)
  - Make a note of patients with a history of alternating remissions (≥12 months) and relapses on any AED + specify the number of remissions.
  - During its long and sometimes fluctuating course a person's epilepsy may not fulfil the definition criteria for either drug resistant or drug-responsive epilepsy at certain time points. In such circumstances, drug responsiveness should be temporarily classified as ''undefined.''

Task 3: Search for rare variants causing broad drug resistance

- **Extreme AED resistance** is defined as:
  - Clearly identifiable (MRI- or histologically-confirmed), stable lesion
  - Follow-up of ≥5 years
  - Seizures recurring at a frequency of ≥4/year over the last year till latest data entry, despite *adequate* trials of ≥5 *tolerated* and *appropriately chosen and used* AED schedules (whether as monotherapies or in combination).
  - Never been seizure-free for ≥12 months
  - Patients who have had epilepsy surgery and fulfilled the above criteria before surgery can also be included
- **Drug responsiveness** is defined as:
  - Clearly identifiable, stable lesion
  - Free from seizures for ≥5 years up to latest data entry
  - Patients cannot have had surgery for their epilepsy.

WP 04 – Genome-based biomarker discovery for late response to specific AEDs

Only patients who have failed at least one AED trial due to lack of efficacy will be included for this purpose (although this AED does not necessarily have to be withdrawn).

Task 1: Search for variants for late response to specific AEDs in focal epilepsies

- AEDs to be included: mainly LTG, LEV, CBZ/OXC, TPM, probably others (LCM, RTG…).
- **Response to specific AEDs** is defined as freedom from seizures lasting for ≥12 months which according to the treating clinician and/or the phenotyper can be attributed to the AED, e.g. after an increase of dose (and prior to initiation of another treatment for epilepsy).
- **Failure of specific AEDs** is defined as seizures recurring at >50% of the pretreatment seizure frequency after the *appropriate* AED has been *adequately* applied (see Appendix for guidance).
    - Patients known to be systematically non-adherent should be excluded. (1 seizure a year due to non-adherence may be disregarded)

- **Extreme late AED response** is defined as:
    - Patients who failed *adequate* trials of ≥2 *tolerated* and *appropriately chosen and used* AED schedules (whether as monotherapies or in combination).
    - Became seizure free for ≥12 months after reaching the minimum therapeutic dose of the AED 'X' (see Appendix 1)
    - Where appropriate, concomitant AEDs have been withdrawn to leave the patient on AED 'X' in monotherapy.
    - Patients cannot have had surgery for their epilepsy.

Task 2: Search for variants for late response to specific AEDs in generalised epilepsies

- AEDs to be included: LTG & VPA, maybe others (e.g. LEV)
- Same definitions as above
- Make a note of patients responding to VPA+LTG combination therapy, when both AEDs alone have failed, and to those not responding to VPA+LTG

- Absolute number (if known):

- Categorical: 1-2, 3-5, 6-10, 11-20, 21+, unknown

3) Combined (if type unknown):

- Absolute number (if known):

- Categorical: 1-2, 3-5, 6-10, 11-20, 21+, unknown

- Did the patient experience at least one seizure in the 12 months prior to starting the first AED? Yes / No / Unknown

- Any periods of ≥12 months remission: yes / no / unknown

If yes: - Remission 1: - start date:

- stop date:

- Remission 2: - start date:

- stop date:

- Remission 3: - start date:

- stop date:

- Remission 4: - start date:

- stop date:

- Total number of seizures in the last 12 months prior to latest recorded visit

1) GTC seizures:

- Absolute number (if known):

- Categorical: 1-3, ≥4, unknown

2) Non-GTC seizures (any type):

- Absolute number (if known):

- Categorical: 1-3, ≥4, unknown

3) Combined (if type unknown):

- Absolute number (if known):

- Categorical: 1-3, ≥4, unknown

- Total number of seizures in the last 12 months prior to non-medical epilepsy treatment (if applicable)

1) GTC seizures:

- Absolute number (if known):

- Categorical: 1-3, ≥4, unknown

WP 06 – Genome-based biomarker discovery for valproate teratogenesis

- **Cases** are women with epilepsy who were taking any AED (either in monotherapy or polytherapy) during a period including the first trimester of pregnancy where the infant was identified to have any major congenital malformation (MCM).
- **Controls** are women with epilepsy who were taking any AED (either in monotherapy or polytherapy) during a period including the first trimester of pregnancy where the infant did not have a MCM.
- A **MCM** is defined as any structural abnormality with surgical, medical, functional or cosmetic importance
- MCMs resulting in spontaneous or induced abortion are also included.
- Make a note in case of siblings with MCM.

Appendix

- **Adequate:** the AED has been administered during an adequate time period and at an adequate dose (see also Tables 1 & 2 below).
- **Appropriate:** previously shown to be effective, preferably in randomized controlled studies (e.g. ethosuximide for focal seizures is considered inappropriate and therefore does not count). Please note that some patients may ''fail'' several AEDs before they fail one that is ''appropriate'' and in a way that is ''informative.''

Table 1. AEDs, appropriate seizure types, minimum and defined daily doses for AED monotherapy in adult patients

Antiepileptic drug	Focal seizures	Primary generalised tonic-clonic seizures	Absence seizures	Other primary generalised or unclassified seizures	Minimum therapeutic dose (mg)	Defined daily dose (mg)
Carbamazepine	✓	✓	×	×	600	1000
Clobazam	✓	✓	✓	✓	10	20
Clonazepam	✓	✓	✓	✓	4	8
Eslicarbazepine	✓	✓	×	×	800	800
Ethosuximide	×	×	✓	×	1000	1250
Felbamate	✓	✓	✓	✓	1200	2400
Gabapentin	✓	×	×	×	1200	1800
Lacosamide	✓	×	×	×	200	300
Lamotrigine	✓	✓	✓	✓	150	300
Levetiracetam	✓	✓	✓	✓	1000	1500
Oxcarbazepine	✓	✓	×	×	900	1050
Phenobarbital	✓	✓	×	✓	60	100
Phenytoin	✓	✓	×	×	200	300
Pregabalin	✓	×	×	×	300	300
Primidone	✓	✓	×	✓	750	1250
Retigabine	✓	×	×	×	600	900
Rufinamide	×	×	×	✓	1200	1400
Tiagabine	✓	×	×	×	30	30
Topiramate	✓	✓	✓	✓	100	300
Valproate	✓	✓	✓	✓	1000	1500
Vigabatrin	✓	×	×	×	1000	2000
Zonisamide	✓	✓	✓	✓	150	200

The above doses are given for guidance only. Final judgment of adequacy of any AED trial is left to the discretion of the treating clinician and/or phenotyper.

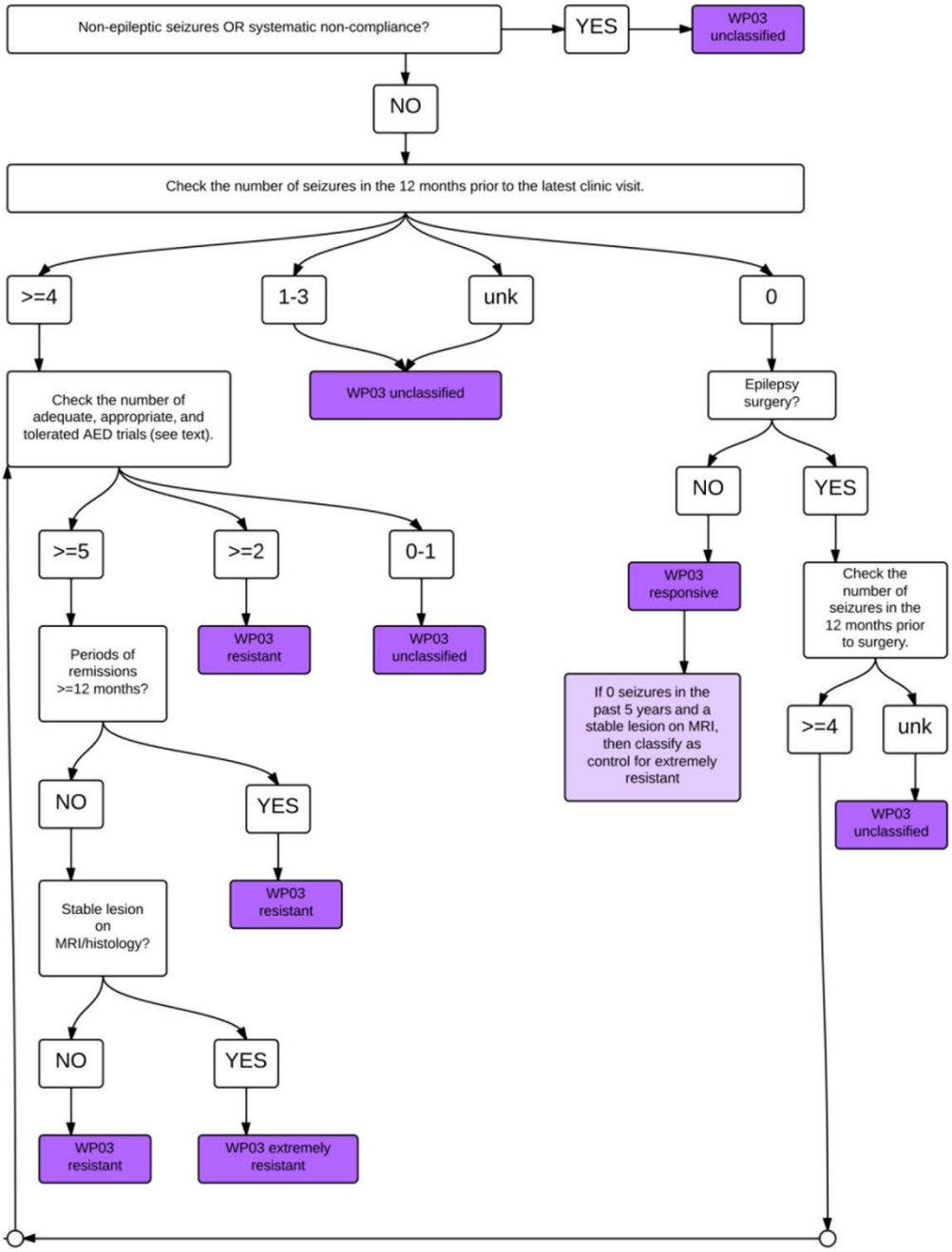
Table 2. Minimum dataset required to determine whether the trial of a therapeutic intervention is informative
Nature of the intervention (e.g., type of drug, in the case of antiepileptic drug treatment)
Mode of application (e.g., formulation, dose, dosing interval, and patient's compliance in case of an antiepileptic drug)
Duration of exposure
Occurrence of seizures and adverse effects during the trial period
Whether there was any effort to optimize dose
Reason(s) for discontinuation (if applicable)
Unsatisfactory seizure control
Adverse effects
Long-term seizure freedom
Psychosocial reasons, for example, planning for pregnancy
Administrative reasons, for example, lost to follow up
Financial issues, for example, cannot afford treatment
Patient/caretaker preference
Other reasons

Kwan P. et al, Epilepsia 2010 (http://www.ncbi.nlm.nih.gov/pubmed/19889013)

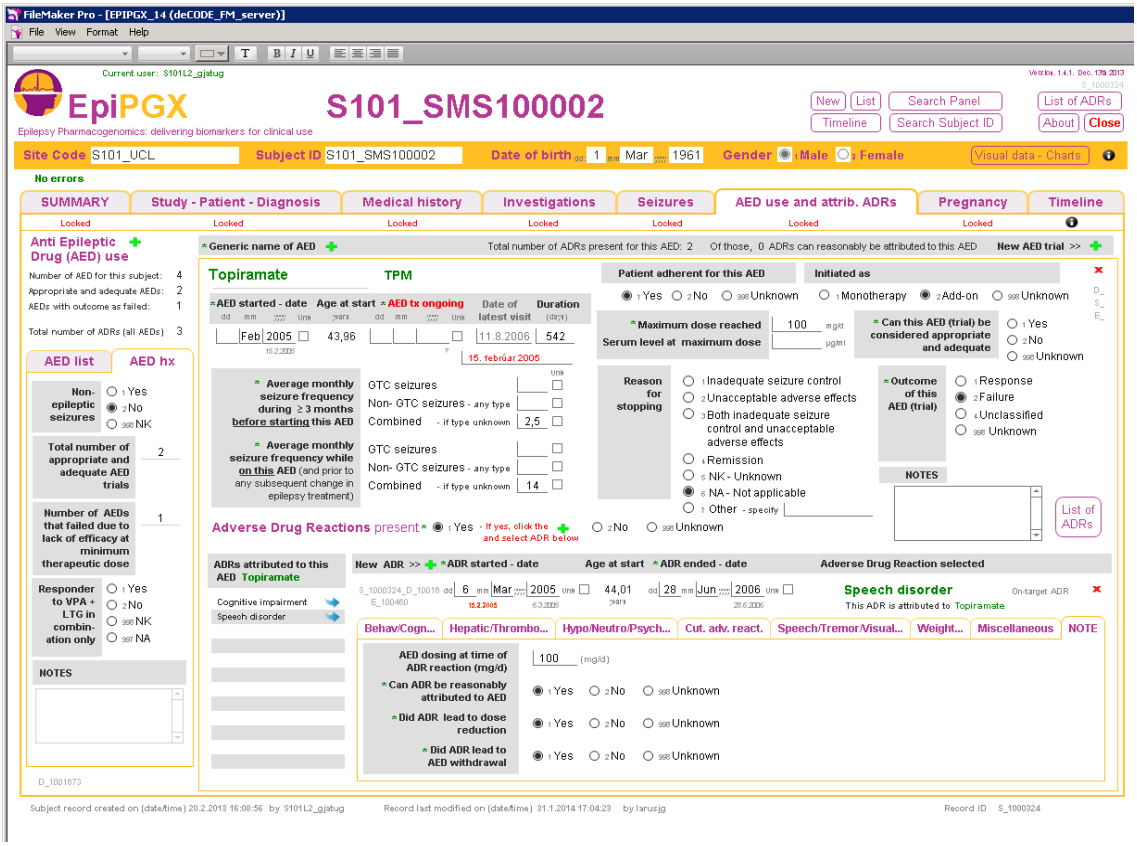
Table 3. Involvement of internal organs in HSS

Organ	Manifestation
Liver	Abnormal LFTs, hepatitis, cholestasis
GI tract	Colitis
Kidney	Nephritis
Lung	Pneumonitis
CNS	Aseptic meningitis, encephalitis, inappropriate anti-diuretic hormone syndrome
Heart	Myocarditis
Muscle	Myositis
Thyroid	Thyroiditis
Haematological	Eosinophilia, atypical lymphocytes, agranulocytosis, thrombocytopenia, haemolytic anaemia, aplastic anemia
Lymphoid system	Lymphadenopathy, pseudolymphoma

Appendix 3. WP03 decision algorithm for the classification of cases



Appendix 4. EpiPGX database view



Appendix 5. National and international ethics regulations relevant for EpiPGX

Country (EpiPGX sites)	National Regulations for clinical studies
UK (EpiPGX sites: UCL, BHSCT, ULIV, Imperial, UGLA)	Local Research Ethics approval under UK clinical research guidelines and European Clinical Trials Directive Application through the Integrated Research Application System (IRAS) which is automatically forwarded to the Multi-centre NHS Research Ethics, NHS Research and Development Committees and MHRA UK Data Protection Act, 1998 UK Human Tissue Act, 2004
Belgium (EpiPGX site ULB)	Arrêté Royal instituant les CEM ou Comité d'Ethique Médicaux Hospitaliers (12 août 1994) - Loi sur la protection de la vie privée (décembre 1992; Arrêté Royal février 2001) définit, entre autres, les conditions dans lesquelles des données de santé, considérées comme sensibles, peuvent être traitées dans le cadre de la recherche clinique. - Loi relative aux droits du patient (août 2002) consacre le droit à l'autonomie du patient et définit la représentation du patient mineur / majeur incapable. - Loi relative à l'expérimentation humaine (mai 2004) définit le cadre dans lequel toute recherche clinique sur la personne humaine doit s'organiser. Loi relative à l'obtention et à l'utilisation de matériel corporel humain destiné à des applications médicales humaines ou à des fins de recherche scientifique (décembre 2008) définit la cadre dans lequel ce matériel peut être prélevé, conservé (biobanques) et utilisé.

Italy (EpiPGX site IGG)	Legislative Decree no.211 of June 24, 2003 “Transposition of Directive 2001/20/EC relating to the implementation of good clinical practice in the conduct of clinical trials on medicinal products for clinical use”. (Decreto Legislativo n. 211 del 24 giugno 2003 Attuazione della direttiva 2001/20/CE relative all'applicazione della buona pratica clinica nell'esecuzione delle sperimentazioni cliniche di medicinali per uso clinico)
The Netherlands (EpiPGX sites: SEIN, UMCU)	Research is approved under the Medical Research Involving Human Subjects Act and/or the Embryos Act by an accredited Medical Research Ethics Committee (MREC) (Medical Research (Human Subjects) Act March 1, 2006. - Wet bescherming persoonsgegevens (WBP) (1 September 2001) [Personal Data Protection Act] - Wet geneeskundige behandelovereenkomst (WGBO) (17 November 1994) [Medical Treatment Agreement Act] - Wet medisch onderzoek (WMO) (1Maart 2006) [Medical-Scientific Research Act] - Gedragscode Gezondheidsonderzoek (19 April 2004) [Code of Behaviour for Health Research] The research that will be conducted under this project is covered by the protocol METC 09-352K: The Genetics of epilepsy.
Irenland (EpiPGX site RCSI)	The World Medical Association Declaration of Helsinki 2008, the [Irish] Data Protection Act 1988, the [Irish] Data Protection (Amendment) Act 2003, the [Irish] Disability Act (2005), the [Irish] Data Protection Guidelines on Research in the Health Sector (Data Protection Commissioner 2007), and Human Biological Material: recommendations for Collection, Use and Storage in Research (Irish Council for Bioethics; 2005)
Iceland (EpiPGX site deCODE)	Samples are kept in the deCODE Biobank, which has an operation license from the Ministry of Health and Social Security according to the Act on Biobanks no. 110/2000. deCODE has a licence Reg. no.134/2000 on the keeping and utilisation of biological samples in biobanks

Luxembourg (EpiPGX site UL)	There are two national commissions in Luxembourg that issue decrees and guidelines concerning ethical issues. These are taken into account in biomedicine research performed in Luxumbourg. The commissions are: Committée Nationale d'Ethique (CNE, www.cne.public.lu) And Commission Nationale Pour la Protection des Données (CNPD, wwwcnpd.public.lu).

Relevant EU legislation and directives

Directive 2005/28/EC or Good Clinical Practice Directive, of 8 April 2005 of the European Parliament and of the Council, lays down principles and detailed guidelines for good clinical practice as regards investigational medicinal products for human use, as well as the requirements for authorisation of the manufacturing or importation of such products.

Directive 2001/20/EC or Clinical Trials Directive of 4 April 2001, of the European Parliament and of the Council on the approximation of the laws, Regulations and administrative provisions of the Member States relating to implementation of good clinical practice in the conduct of clinical trials on medicinal products for human use.

Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data

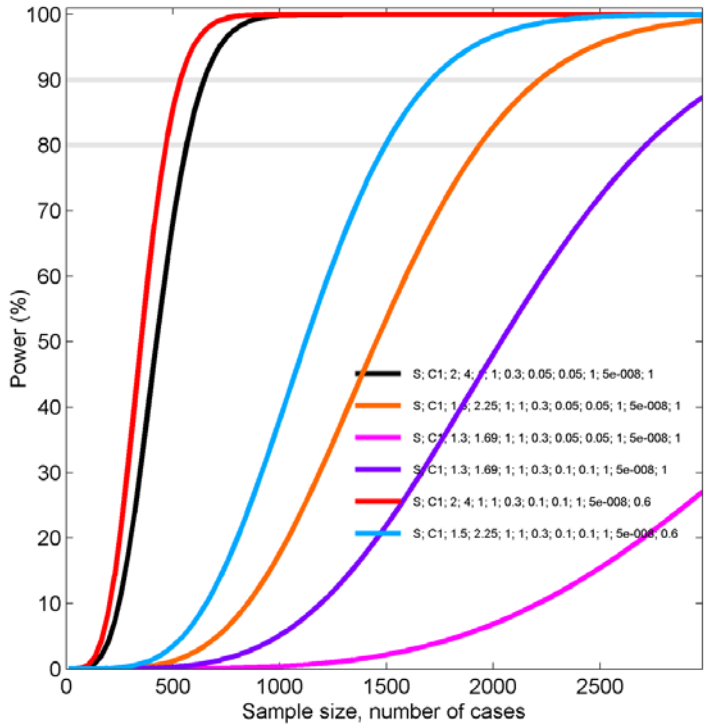
International conventions and declarations

WORLD MEDICAL ASSOCIATION DECLARATION OF HELSINKI: Ethical Principles for Medical Research Involving Human Subjects (Adopted by the 18th WMA General Assembly, Helsinki, Finland, June 1964, and amended by the 29th WMA General Assembly, Tokyo, Japan, October 1975; 35th WMA General Assembly, Venice, Italy, October 1983; 41st WMA General Assembly, Hong Kong, September 1989; 48th WMA General Assembly, Somerset West, Republic of South Africa, October 1996; and the 52nd WMA General Assembly, Edinburgh, Scotland, October 2000; Note of Clarification on Paragraph 29 added by the WMA General Assembly, Washington 2002; Note of Clarification on Paragraph 30 added by the WMA General Assembly, Tokyo 2004)

Site	Ethics committee approving the study
UCL	Joint Research Ethics Committee, National Hospital for Neurology and Neurosurgery and Institute of Neurology, approval 00/N081
ULB	Comité d'Ethique de l'Hôpital Erasme
IGG	Ethics Committee for Scientific and Biomedical Research and for Clinical Experimentation, Instituto Gianina Gaslini
EKUT	Die Ethik-Kommission an der Medizinischen Fakultät der Universität Tübingen
SEIN	Medisch-Ethische Commissie SEIN
UKB	Ethikkommission an der Medizinischen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn
RCSI	Research protocol has been approved by Beaumont Hospital Ethics Committee (study code 02/44, title “Pharmacogenetics of Epilepsy”)
BHSCT	BHSCT Research Ethics Committee
UMCU	CCMO / Medisch Ethische Toetsingsommissie (METC) van het UMCU (study code 09/352, title “The Genetics of epilepsy”, approval ref: AvG/vb/10/15096
ULIV	North West 3 Research Ethics Committee, approval ref: 10/H1002/5
Imperial	North-West Multicentre Research Ethics Committee, approval ref: MREC 02/8/45
UGLA	West Ethics Committee, North Glasgow University Hospitals NHS Trust, approval ref: 02/119(2)

Appendix 6. GWAS1 and GWAS2 power and sample size curves

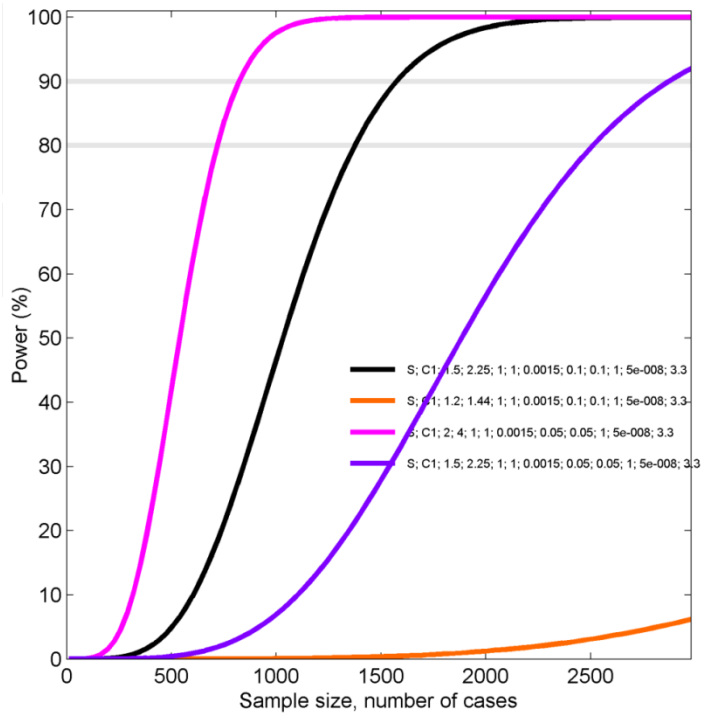
GWAS1 power and sample size curves



All calculations assume a co-dominant model, prevalence of DRE in the epilepsy population 0.3, r^2 of 1.0 between a causal variant and a genotyped marker, and a genome-wide statistical significance threshold $5*10^{-8}$.
Specific parameters:

- RR 2; disease allele frequency 0.1; marker allele frequency 0.1; control to case ratio 0.6
- RR 2; disease allele frequency 0.05; marker allele frequency 0.05; control to case ratio 0.6
- RR 1.5; disease allele frequency 0.1; marker allele frequency 0.05; control to case ratio 0.6
- RR 1.5; disease allele frequency 0.05; marker allele frequency 0.05; control to case ratio 1
- RR 1.3; disease allele frequency 0.1; marker allele frequency 0.1; control to case ratio 1
- RR 1.3; disease allele frequency 0.1; marker allele frequency 0.1; control to case ratio 1

GWAS2 power and sample size curves

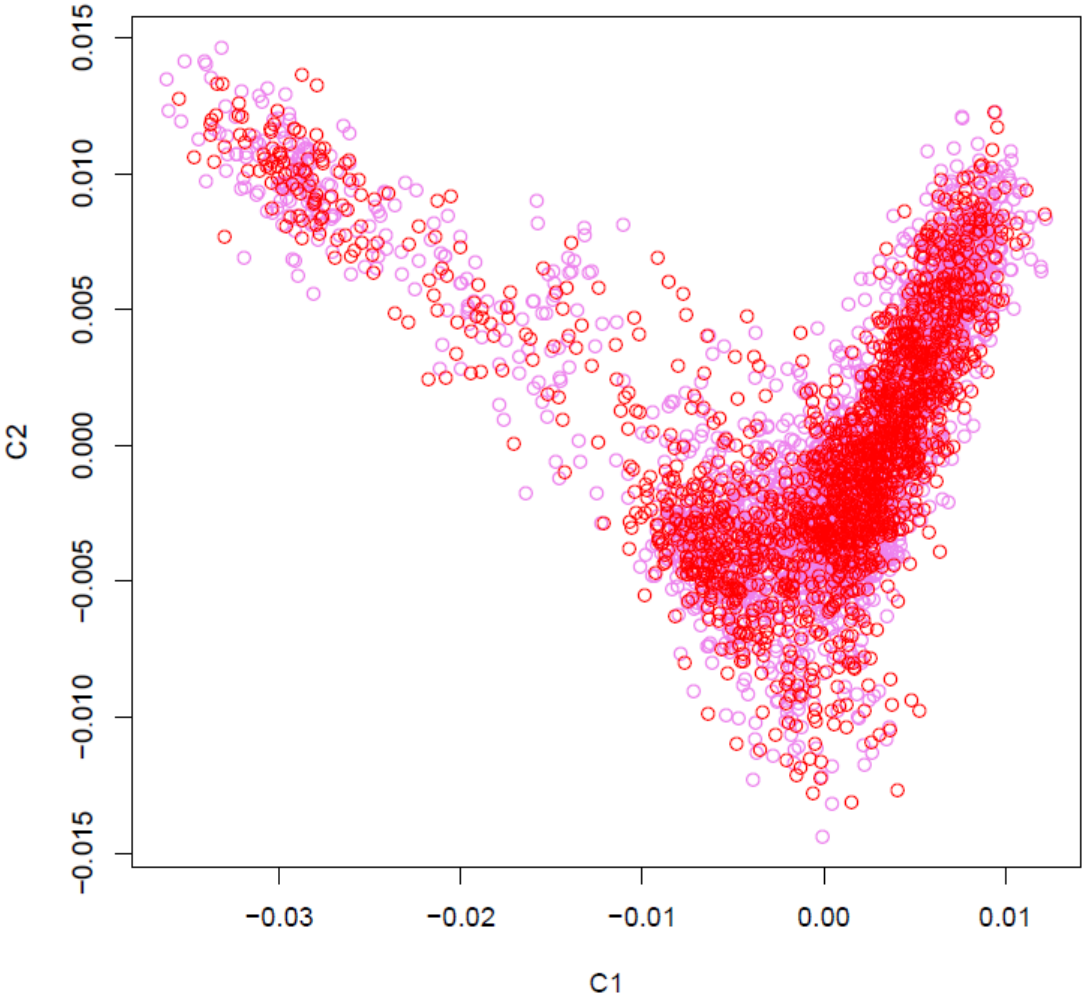


The following assumptions were used for all power calculations: a co-dominant model, prevalence of DRE in the epilepsy population 0.0015, r^2 of 1.0 between a causal variant and a genotyped marker, and a genome-wide statistical significance threshold $5*10^{-8}$.
Parameters specific to individual calculations:

- RR 2; disease allele frequency 0.05; marker allele frequency 0.05; control to case ratio 3.3
- RR 1.5; disease allele frequency 0.1; marker allele frequency 0.1; control to case ratio 3.3
- RR 1.5; disease allele frequency 0.05; marker allele frequency 0.05; control to case ratio 3.3
- RR 1.2; disease allele frequency 0.1; marker allele frequency 0.1; control to case ratio 3.3

Appendix 7. MDS analyses of cases and controls included in GWAS1 and GWAS2

MDS analysis of drug-resistant cases and drug-responsive controls included in GWAS1

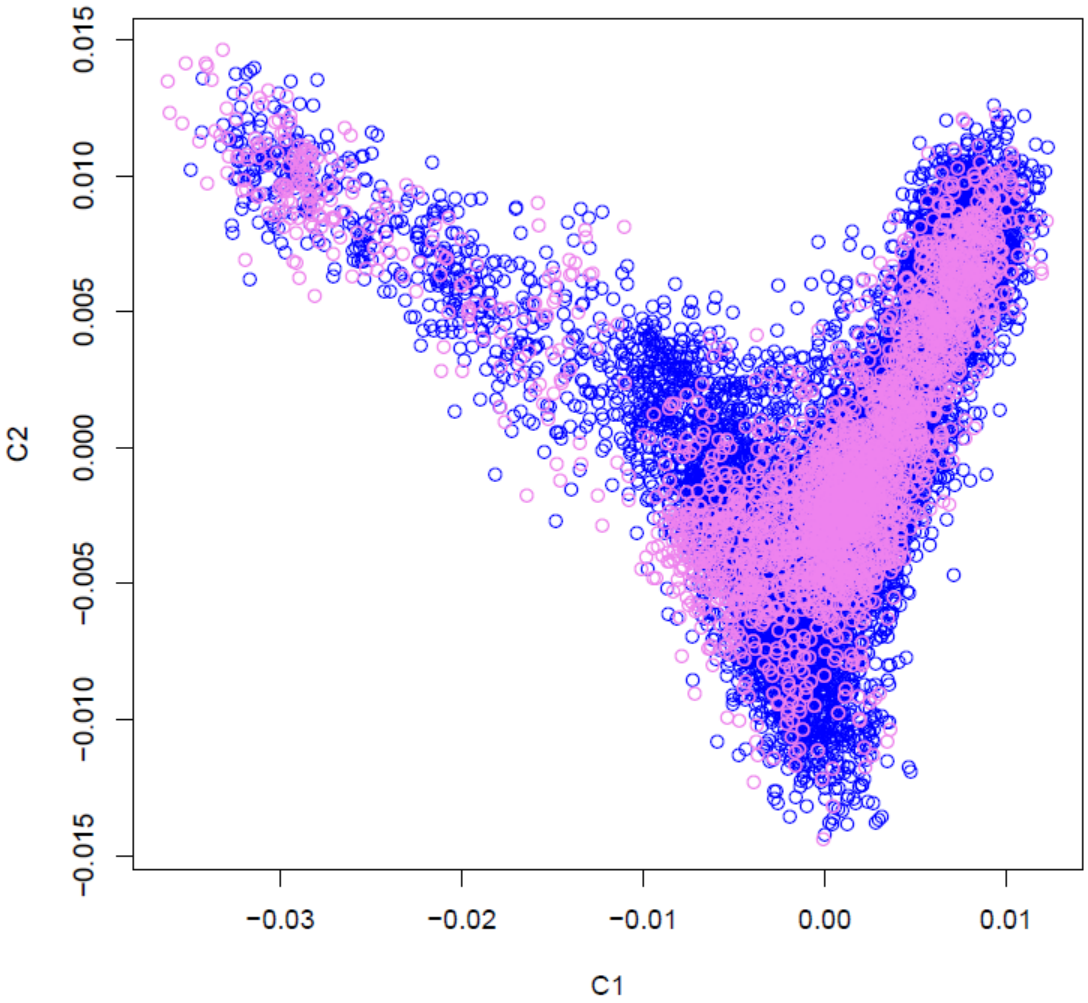


o Drug-resistant epilepsy cases

o Drug-responsive epilepsy controls

C1 = principal coordinate 1, C2 = principal coordinate 2.

MDS analysis of drug-resistant cases and healthy controls included in GWAS2



o Drug-resistant epilepsy cases

o Healthy controls

C1 = principal coordinate 1, C2 = principal coordinate 2.

Appendix 8. GWAS1 and GWAS2 statistics for SNPs in genes previously associated with drug resistance in epilepsy

Gene	*SNP*	*Base pair position*	*Chr*	*Allele 1, allele 2*	*MAF (minor allele)*	*P-value (FaSTLMM) GWAS1*	*OR (95% CI)*	*P-value (FaSTLMM) GWAS2*	*OR (95% CI)*
CYP2C9	*rs9332104*	*96698690*	*10*	*T, C*	*0.2165 (C)*	$8.57*10^{-01}$	*1.00 (0.97-1.02)*	$2.90*10^{-01}$	*0.99 (0.98-1.00)*
	rs9332108	*96699980*	*10*	*T, C*	*0.0724 (C)*	$9.42*10^{-01}$	*1.00 (0.96-1.04)*	$3.38*10^{-01}$	*0.99 (0.97-1.01)*
	rs2253635	*96700537*	*10*	*C, T*	*0.3766 (C)*	$2.12*10^{-01}$	*0.99 (0.97-1.01)*	$4.67*10^{-01}$	*1.00 (0.99-1.01)*
	rs1799853	*96702047*	*10*	*C, T*	*0.1213 (T)*	$4.23*10^{-01}$	*0.99 (0.96-1.02)*	$2.69*10^{-01}$	*1.01 (0.99-1.02)*
	rs4086116	*96707202*	*10*	*C, T*	*0.1939 (T)*	$5.46*10^{-01}$	*0.99 (0.97-1.02)*	$7.72*10^{-01}$	*1.00 (0.99-1.01)*
	rs17443251	*96707890*	*10*	*T, C*	*0.1232 (C)*	$4.25*10^{-01}$	*0.99 (0.96-1.02)*	$1.99*10^{-01}$	*1.01 (1.00-1.02)*
	rs10509679	*96708226*	*10*	*G, A*	*0.1734 (A)*	$8.82*10^{-01}$	*1.00 (0.98-1.03)*	$9.01*10^{-01}$	*1.00 (0.99-1.01)*
	rs4918766	*96711884*	*10*	*G, A*	*0.3692 (A)*	$6.37*10^{-01}$	*1.00 (0.97-1.02)*	$7.35*10^{-01}$	*1.00 (0.99-1.01)*
	rs2475376	*96712400*	*10*	*A, G*	*0.1573 (A)*	$6.99*10^{-01}$	*0.99 (0.97-1.02)*	$9.09*10^{-01}$	*1.00 (0.99-1.01)*
	rs7897079	*96720518*	*10*	*A, G*	*0.2163 (G)*	$8.18*10^{-01}$	*1.00 (0.97-1.02)*	$2.75*10^{-01}$	*0.99 (0.98-1.01)*
	rs2153628	*96723424*	*10*	*A, G*	*0.2163 (G)*	$8.18*10^{-01}$	*1.00 (0.97-1.02)*	$2.75*10^{-01}$	*0.99 (0.98-1.01)*
	rs4917639	*96725535*	*10*	*A, C*	*0.1940 (C)*	$5.65*10^{-01}$	*0.99 (0.97-1.02)*	$7.54*10^{-01}$	*1.00 (0.99-1.01)*
	rs12569850	*96727160*	*10*	*A, G*	*0.1734 (G)*	$8.75*10^{-01}$	*1.00 (0.97-1.03)*	$9.02*10^{-01}$	*1.00 (0.99-1.01)*
	rs9332168	*96731292*	*10*	*C, T*	*0.2166 (T)*	$8.59*10^{-01}$	*1.00 (0.97-1.02)*	$2.50*10^{-01}$	*0.99 (0.98-1.01)*
	rs9332169	*96731310*	*10*	*A, G*	*0.0726 (G)*	$9.25*10^{-01}$	*1.00 (0.96-1.04)*	$3.20*10^{-01}$	*0.99 (0.97-1.01)*
	rs9332172	*96731788*	*10*	*A, G*	*0.1946 (G)*	$5.63*10^{-01}$	*0.99 (0.97-1.02)*	$6.92*10^{-01}$	*1.00 (0.99-1.01)*
	rs9332174	*96732097*	*10*	*A, G*	*0.2166 (G)*	$8.59*10^{-01}$	*1.00 (0.97-1.02)*	$2.50*10^{-01}$	*0.99 (0.98-1.01)*
	rs17110288	*96732599*	*10*	*G, A*	*0.0726 (A)*	$9.25*10^{-01}$	*1.00 (0.96-1.04)*	$3.15*10^{-01}$	*0.99 (0.97-1.01)*
	rs1856908	*96732731*	*10*	*T, G*	*0.3758 (T)*	$2.54*10^{-01}$	*0.99 (0.97-1.01)*	$4.81*10^{-01}$	*1.00 (0.99-1.01)*
	rs9325473	*96734582*	*10*	*G, A*	*0.0726 (A)*	$9.25*10^{-01}$	*1.00 (0.96-1.04)*	$3.11*10^{-01}$	*0.99 (0.97-1.01)*
	rs1057910	*96741053*	*10*	*A, C*	*0.0726 (C)*	$9.25*10^{-01}$	*1.00 (0.96-1.04)*	$3.15*10^{-01}$	*0.99 (0.97-1.01)*
	rs1934967	*96741426*	*10*	*C, T*	*0.2157 (T)*	$3.72*10^{-01}$	*1.01 (0.99-1.04)*	$3.47*10^{-01}$	*1.01 (0.99-1.02)*

	rs1934968	*96741817*	*10*	*A, G*	*0.1101 (A)*	*$7.95*10^{-01}$*	*1.00 (0.97-1.04)*	*$8.80*10^{-01}$*	*1.00 (0.98-1.01)*
	rs9332214	*96743108*	*10*	*T, C*	*0.0726 (C)*	*$9.25*10^{-01}$*	*1.00 (0.96-1.04)*	*$3.20*10^{-01}$*	*0.99 (0.97-1.01)*
	rs9332217	*96743228*	*10*	*A, C*	*0.0726 (C)*	*$9.25*10^{-01}$*	*1.00 (0.96-1.04)*	*$3.15*10^{-01}$*	*0.99 (0.97-1.01)*
	rs9332220	*96743943*	*10*	*G, A*	*0.1939 (A)*	*$5.80*10^{-01}$*	*0.99 (0.97-1.02)*	*$7.08*10^{-01}$*	*1.00 (0.99-1.01)*
	rs9332222	*96744064*	*10*	*G, A*	*0.1220 (A)*	*$3.99*10^{-01}$*	*0.99 (0.95-1.02)*	*$3.23*10^{-01}$*	*1.01 (0.99-1.02)*
	rs9332227	*96745180*	*10*	*T, G*	*0.0726 (G)*	*$9.25*10^{-01}$*	*1.00 (0.96-1.04)*	*$3.14*10^{-01}$*	*0.99 (0.97-1.01)*
	rs2298037	*96746078*	*10*	*C, T*	*0.1732 (T)*	*$8.52*10^{-01}$*	*1.00 (0.97-1.03)*	*$8.87*10^{-01}$*	*1.00 (0.99-1.01)*
	rs9332238	*96748492*	*10*	*G, A*	*0.1940 (A)*	*$5.74*10^{-01}$*	*0.99 (0.97-1.02)*	*$8.14*10^{-01}$*	*1.00 (0.99-1.01)*
CYP2C19	*rs12768009*	*96525865*	*10*	*G, A*	*0.1511 (A)*	*$5.27*10^{-01}$*	*1.01 (0.98-1.04)*	*$8.64*10^{-01}$*	*1.00 (0.99-1.01)*
	rs6583954	*96534263*	*10*	*C, T*	*0.1534 (T)*	*$4.51*10^{-01}$*	*1.01 (0.98-1.04)*	*$9.42*10^{-01}$*	*1.00 (0.99-1.01)*
	rs7916649	*96534584*	*10*	*G, A*	*0.4417 (A)*	*$9.63*10^{-01}$*	*1.00 (0.98-1.02)*	*$2.56*10^{-01}$*	*0.99 (0.98-1.00)*
	rs4388808	*96536056*	*10*	*A, G*	*0.1767 (G)*	*$8.06*10^{-01}$*	*1.00 (0.97-1.02)*	*$8.37*10^{-01}$*	*1.00 (0.99-1.01)*
	rs7068577	*96536708*	*10*	*C, T*	*0.2185 (T)*	*$7.60*10^{-01}$*	*1.00 (0.97-1.02)*	*$3.87*10^{-01}$*	*1.00 (0.98-1.01)*
	rs17878673	*96539144*	*10*	*A, G*	*0.0695 (G)*	*$8.00*10^{-01}$*	*1.01 (0.96-1.05)*	*$4.28*10^{-01}$*	*0.99 (0.97-1.01)*
	rs4304697	*96540889*	*10*	*G, A*	*0.0695 (A)*	*$8.15*10^{-01}$*	*1.00 (0.96-1.05)*	*$4.43*10^{-01}$*	*0.99 (0.97-1.01)*
	rs7088784	*96541373*	*10*	*A, G*	*0.0715 (G)*	*$8.75*10^{-01}$*	*1.00 (0.96-1.04)*	*$4.63*10^{-01}$*	*0.99 (0.97-1.01)*
	rs4244285	*96541616*	*10*	*G, A*	*0.1511 (A)*	*$5.27*10^{-01}$*	*1.01 (0.98-1.04)*	*$9.94*10^{-01}$*	*1.00 (0.99-1.01)*
	rs12571421	*96541982*	*10*	*A, G*	*0.1512 (G)*	*$5.00*10^{-01}$*	*1.01 (0.98-1.04)*	*$9.96*10^{-01}$*	*1.00 (0.99-1.01)*
	rs35390752	*96543823*	*10*	*T, G*	*0.1511 (G)*	*$5.27*10^{-01}$*	*1.01 (0.98-1.04)*	*$9.79*10^{-01}$*	*1.00 (0.99-1.01)*
	rs12767583	*96547463*	*10*	*C, T*	*0.1511 (T)*	*$5.27*10^{-01}$*	*1.01 (0.98-1.04)*	*$9.67*10^{-01}$*	*1.00 (0.99-1.01)*
	rs4494250	*96563757*	*10*	*G, A*	*0.3423 (A)*	*$4.85*10^{-01}$*	*1.01 (0.99-1.03)*	*$1.53*10^{-01}$*	*1.01 (1.00-1.02)*
	rs12772672	*96566889*	*10*	*A, G*	*0.1510 (G)*	*$5.32*10^{-01}$*	*1.01 (0.98-1.04)*	*$9.64*10^{-01}$*	*1.00 (0.99-1.01)*
	rs4641393	*96567386*	*10*	*C, T*	*0.1512 (T)*	*$5.06*10^{-01}$*	*1.01 (0.98-1.04)*	*$9.94*10^{-01}$*	*1.00 (0.99-1.01)*
	rs1322179	*96575242*	*10*	*C, T*	*0.1511 (T)*	*$5.27*10^{-01}$*	*1.01 (0.98-1.04)*	*$9.84*10^{-01}$*	*1.00 (0.99-1.01)*
	rs10509678	*96576190*	*10*	*A, C*	*0.0692 (C)*	*$7.53*10^{-01}$*	*1.01 (0.97-1.05)*	*$4.59*10^{-01}$*	*0.99 (0.98-1.01)*
	rs10786172	*96581094*	*10*	*A, G*	*0.3423 (G)*	*$4.77*10^{-01}$*	*1.01 (0.99-1.03)*	*$1.50*10^{-01}$*	*1.01 (1.00-1.02)*
	rs11592737	*96603414*	*10*	*A, G*	*0.2189 (G)*	*$8.43*10^{-01}$*	*1.00 (0.97-1.02)*	*$3.92*10^{-01}$*	*1.00 (0.98-1.01)*
	rs1322181	*96609064*	*10*	*G, A*	*0.4420 (A)*	*$8.81*10^{-01}$*	*1.00 (0.98-1.02)*	*$2.36*10^{-01}$*	*0.99 (0.98-1.00)*
	rs4917623	*96609568*	*10*	*T, C*	*0.4809 (T)*	*$6.68*10^{-01}$*	*1.00 (0.98-1.03)*	*$2.51*10^{-01}$*	*1.01 (1.00-1.02)*

	rs17878382	*96610631*	*10*	*T, C*	*0.0692 (C)*	*$7.53*10^{-01}$*	*1.01 (0.97-1.05)*	*$5.76*10^{-01}$*	*1.00 (0.98-1.01)*
	rs12779363	*96612040*	*10*	*G, A*	*0.2191 (A)*	*$8.28*10^{-01}$*	*1.00 (0.97-1.02)*	*$4.26*10^{-01}$*	*1.00 (0.98-1.01)*
	rs12268020	*96612371*	*10*	*C, T*	*0.2189 (T)*	*$8.43*10^{-01}$*	*1.00 (0.97-1.02)*	*$4.00*10^{-01}$*	*1.00 (0.98-1.01)*
ABCB1	*rs17149699*	*87141751*	*7*	*C, T*	*0.0367 (T)*	*$9.67*10^{-01}$*	*1.00 (0.95-1.06)*	*$4.93*10^{-01}$*	*1.01 (0.98-1.03)*
	rs4148751	*87143153*	*7*	*T, C*	*0.0364 (C)*	*$9.29*10^{-01}$*	*1.00 (0.95-1.06)*	*$4.37*10^{-01}$*	*1.01 (0.99-1.04)*
	rs4148750	*87143275*	*7*	*T, C*	*0.0370 (C)*	*$9.71*10^{-01}$*	*1.00 (0.94-1.06)*	*$4.66*10^{-01}$*	*1.01 (0.98-1.03)*
	rs1922243	*87143504*	*7*	*C, T*	*0.0370 (T)*	*$9.71*10^{-01}$*	*1.00 (0.94-1.06)*	*$4.88*10^{-01}$*	*1.01 (0.98-1.03)*
	rs4148740	*87152103*	*7*	*A, G*	*0.1283 (G)*	*$7.46*10^{-01}$*	*1.00 (0.96-1.03)*	*$7.69*10^{-01}$*	*1.00 (0.99-1.01)*
	rs2373588	*87153160*	*7*	*G, A*	*0.0325 (A)*	*$6.12*10^{-01}$*	*1.02 (0.96-1.08)*	*$2.62*10^{-01}$*	*1.02 (0.99-1.04)*
	rs10280101	*87153585*	*7*	*A, C*	*0.1287 (C)*	*$6.97*10^{-01}$*	*0.99 (0.96-1.03)*	*$8.25*10^{-01}$*	*1.00 (0.99-1.02)*
	rs10225473	*87154646*	*7*	*A, G*	*0.1286 (G)*	*$7.39*10^{-01}$*	*0.99 (0.96-1.03)*	*$8.07*10^{-01}$*	*1.00 (0.99-1.02)*
	rs7787082	*87157051*	*7*	*G, A*	*0.1615 (A)*	*$8.65*10^{-01}$*	*1.00 (0.97-1.03)*	*$4.50*10^{-01}$*	*1.00 (0.99-1.02)*
	rs2032583	*87160561*	*7*	*A, G*	*0.1289 (G)*	*$6.85*10^{-01}$*	*0.99 (0.96-1.03)*	*$8.66*10^{-01}$*	*1.00 (0.99-1.02)*
	rs2032582	*87160618*	*7*	*A, C*	*0.4532 (A)*	*$2.16*10^{-01}$*	*1.01 (0.99-1.04)*	*$7.00*10^{-01}$*	*1.00 (0.99-1.01)*
	rs4148739	*87161049*	*7*	*T, C*	*0.1289 (C)*	*$6.85*10^{-01}$*	*0.99 (0.96-1.03)*	*$8.66*10^{-01}$*	*1.00 (0.99-1.02)*
	rs11983225	*87161520*	*7*	*T, C*	*0.1289 (C)*	*$6.85*10^{-01}$*	*0.99 (0.96-1.03)*	*$8.66*10^{-01}$*	*1.00 (0.99-1.02)*
	rs11760837	*87163016*	*7*	*T, C*	*0.1289 (C)*	*$6.85*10^{-01}$*	*0.99 (0.96-1.03)*	*$8.66*10^{-01}$*	*1.00 (0.99-1.02)*
	rs10274587	*87164483*	*7*	*G, A*	*0.1289 (A)*	*$6.85*10^{-01}$*	*0.99 (0.96-1.03)*	*$8.73*10^{-01}$*	*1.00 (0.99-1.02)*
	rs10248420	*87164986*	*7*	*A, G*	*0.1615 (G)*	*$9.18*10^{-01}$*	*1.00 (0.97-1.03)*	*$4.23*10^{-01}$*	*1.01 (0.99-1.02)*
	rs2235040	*87165750*	*7*	*C, T*	*0.1289 (T)*	*$6.85*10^{-01}$*	*0.99 (0.96-1.03)*	*$8.73*10^{-01}$*	*1.00 (0.99-1.02)*
	rs12668877	*87167004*	*7*	*C, T*	*0.0186 (T)*	*$2.39*10^{-01}$*	*0.95 (0.88-1.03)*	*$6.64*10^{-01}$*	*0.99 (0.96-1.03)*
	rs3789246	*87168027*	*7*	*C, T*	*0.0186 (T)*	*$2.40*10^{-01}$*	*0.96 (0.88-1.03)*	*$6.57*10^{-01}$*	*0.99 (0.96-1.03)*
	rs7795817	*87169037*	*7*	*C, T*	*0.0187 (T)*	*$2.24*10^{-01}$*	*0.95 (0.88-1.03)*	*$6.32*10^{-01}$*	*0.99 (0.96-1.03)*
	rs12720067	*87169356*	*7*	*C, T*	*0.1277 (T)*	*$8.35*10^{-01}$*	*1.00 (0.97-1.03)*	*$7.16*10^{-01}$*	*1.00 (0.99-1.02)*
	rs4148737	*87171152*	*7*	*T, C*	*0.4168 (C)*	*$4.86*10^{-01}$*	*1.01 (0.99-1.03)*	*$4.67*10^{-01}$*	*0.99 (0.99-1.01)*
	rs4148736	*87171383*	*7*	*G, A*	*0.4169 (A)*	*$4.96*10^{-01}$*	*1.01 (0.99-1.03)*	*$4.53*10^{-01}$*	*1.00 (0.99-1.01)*
	rs10276603	*87171527*	*7*	*T, C*	*0.1286 (C)*	*$7.33*10^{-01}$*	*0.99 (0.96-1.03)*	*$8.35*10^{-01}$*	*1.00 (0.99-1.02)*
	rs6961419	*87172136*	*7*	*T, C*	*0.4167 (C)*	*$4.95*10^{-01}$*	*1.01 (0.99-1.03)*	*$4.63*10^{-01}$*	*1.00 (0.99-1.01)*
	rs4148735	*87172881*	*7*	*C, T*	*0.4167 (T)*	*$5.04*10^{-01}$*	*1.01 (0.99-1.03)*	*$4.69*10^{-01}$*	*1.00 (0.99-1.01)*

	rs2235046	*87174066*	*7*	*T, C*	*0.4545 (T)*	*$6.40*10^{-01}$*	*1.01 (0.98-1.03)*	*$5.97*10^{-01}$*	*1.00 (0.99-1.01)*
	rs2091766	*87174504*	*7*	*C, T*	*0.3856 (T)*	*$8.34*10^{-01}$*	*1.02 (1.00-1.04)*	*$7.81*10^{-01}$*	*1.00 (0.99-1.01)*
	rs2235013	*87178626*	*7*	*C, T*	*0.4834 (T)*	*$8.07*10^{-01}$*	*1.00 (0.98-1.02)*	*$8.11*10^{-01}$*	*1.00 (0.99-1.01)*
	rs2235035	*87179086*	*7*	*G, A*	*0.3147 (A)*	*$8.54*10^{-02}$*	*1.02 (1.00-1.04)*	*$4.46*10^{-01}$*	*1.00 (0.99-1.01)*
	rs2235033	*87179143*	*7*	*A, G*	*0.4834 (G)*	*$8.07*10^{-01}$*	*1.00 (0.98-1.02)*	*$8.12*10^{-01}$*	*1.00 (0.99-1.01)*
	rs1128503	*87179601*	*7*	*A, G*	*0.4382 (A)*	*$8.54*10^{-01}$*	*1.00 (0.98-1.02)*	*$4.07*10^{-01}$*	*1.00 (0.99-1.01)*
	rs10276036	*87180198*	*7*	*C, T*	*0.4381 (C)*	*$8.32*10^{-01}$*	*1.00 (0.98-1.02)*	*$4.01*10^{-01}$*	*1.00 (0.99-1.01)*
	rs12704364	*87181175*	*7*	*C, T*	*0.4833 (T)*	*$8.25*10^{-01}$*	*1.00 (0.98-1.02)*	*$8.08*10^{-01}$*	*1.00 (0.99-1.01)*
	rs6961665	*87181418*	*7*	*C, A*	*0.4832 (A)*	*$8.50*10^{-01}$*	*1.00 (0.98-1.02)*	*$8.08*10^{-01}$*	*1.00 (0.99-1.01)*
	rs1922240	*87183354*	*7*	*T, C*	*0.3138 (C)*	*$1.04*10^{-01}$*	*1.02 (1.00-1.04)*	*$4.46*10^{-01}$*	*1.00 (0.99-1.01)*
	rs1922241	*87185894*	*7*	*G, A*	*0.3146 (A)*	*$8.19*10^{-02}$*	*1.02 (1.00-1.04)*	*$4.83*10^{-01}$*	*1.00 (0.99-1.01)*
	rs868755	*87189930*	*7*	*T, G*	*0.4233 (T)*	*$5.36*10^{-01}$*	*1.01 (0.99-1.03)*	*$6.90*10^{-01}$*	*1.00 (0.99-1.01)*
	rs2235023	*87190452*	*7*	*C, T*	*0.0783 (T)*	*$9.04*10^{-01}$*	*1.00 (0.96-1.04)*	*$2.49*10^{-01}$*	*0.99 (0.97-1.01)*
	rs11975994	*87192731*	*7*	*G, A*	*0.4374 (G)*	*$8.89*10^{-01}$*	*1.00 (0.98-1.02)*	*$4.26*10^{-01}$*	*1.00 (0.99-1.01)*
	rs4148734	*87193597*	*7*	*G, A*	*0.2910 (A)*	*$3.44*10^{-01}$*	*1.03 (1.00-1.05)*	*$3.45*10^{-01}$*	*1.01 (0.99-1.02)*
	rs1202170	*87195106*	*7*	*C, T*	*0.4783 (C)*	*$8.95*10^{-01}$*	*1.00 (0.98-1.02)*	*$8.28*10^{-01}$*	*1.00 (0.99-1.01)*
	rs1202169	*87195850*	*7*	*T, C*	*0.4371 (C)*	*$9.38*10^{-01}$*	*1.00 (0.98-1.02)*	*$4.34*10^{-01}$*	*1.00 (0.99-1.01)*
	rs1202168	*87195962*	*7*	*G, A*	*0.4370 (A)*	*$9.15*10^{-01}$*	*1.00 (0.98-1.02)*	*$4.54*10^{-01}$*	*1.00 (0.99-1.01)*
	rs1202167	*87197059*	*7*	*C, T*	*0.4371 (T)*	*$8.73*10^{-01}$*	*1.00 (0.98-1.02)*	*$5.03*10^{-01}$*	*1.00 (0.99-1.01)*
	rs1024409	*87198367*	*7*	*G, A*	*0.3768 (A)*	*$2.04*10^{-01}$*	*0.95 (0.88-1.03)*	*$9.75*10^{-01}$*	*1.00 (0.97-1.04)*
	rs2235015	*87199564*	*7*	*C, A*	*0.2058 (A)*	*$9.27*10^{-01}$*	*1.00 (0.97-1.03)*	*$7.43*10^{-01}$*	*1.00 (0.99-1.01)*
	rs10259849	*87200842*	*7*	*C, T*	*0.4150 (C)*	*$3.70*10^{-01}$*	*0.98 (0.95-1.00)*	*$4.03*10^{-01}$*	*1.00 (0.99-1.01)*
	rs2520464	*87201086*	*7*	*C, T*	*0.0187 (T)*	*$8.72*10^{-01}$*	*1.00 (0.98-1.02)*	*$5.96*10^{-01}$*	*1.00 (0.99-1.01)*
	rs10280623	*87202544*	*7*	*T, C*	*0.1980 (C)*	*$9.16*10^{-01}$*	*1.00 (0.97-1.03)*	*$8.67*10^{-01}$*	*1.00 (0.99-1.01)*
	rs10264990	*87202615*	*7*	*C, T*	*0.2929 (C)*	*$5.22*10^{-01}$*	*0.99 (0.97-1.02)*	*$5.90*10^{-01}$*	*1.00 (0.99-1.01)*
	rs1202180	*87203840*	*7*	*C, T*	*0.4370 (C)*	*$1.79*10^{-01}$*	*0.99 (0.96-1.01)*	*$9.09*10^{-01}$*	*1.00 (0.99-1.01)*
	rs1202179	*87204279*	*7*	*C, T*	*0.1957 (T)*	*$2.00*10^{-01}$*	*0.99 (0.96-1.01)*	*$9.16*10^{-01}$*	*1.00 (0.99-1.01)*
	rs1989830	*87205663*	*7*	*A, G*	*0.3324 (A)*	*$2.11*10^{-01}$*	*0.99 (0.96-1.01)*	*$9.39*10^{-01}$*	*1.00 (0.99-1.01)*
	rs1202175	*87209150*	*7*	*G, A*	*0.3467 (G)*	*$2.23*10^{-01}$*	*0.99 (0.96-1.01)*	*$9.17*10^{-01}$*	*1.00 (0.99-1.01)*
	rs1202172	*87210974*	*7*	*C, A*	*0.3432 (C)*	*$2.64*10^{-01}$*	*0.99 (0.97-1.01)*	*$9.52*10^{-01}$*	*1.00 (0.99-1.01)*

	rs4148733	*87213232*	*7*	*A, G*	*0.3422 (G)*	*$4.49*10^{-01}$*	*1.01 (0.98-1.04)*	*$5.56*10^{-01}$*	*1.00 (0.99-1.02)*
	rs1202185	*87213384*	*7*	*C, T*	*0.3416 (C)*	*$2.31*10^{-01}$*	*0.99 (0.96-1.01)*	*$9.54*10^{-01}$*	*1.00 (0.99-1.01)*
	rs1202184	*87213901*	*7*	*C, T*	*0.3429 (C)*	*$1.94*10^{-01}$*	*0.99 (0.97-1.01)*	*$7.66*10^{-01}$*	*1.00 (0.99-1.01)*
	rs1202182	*87215304*	*7*	*G, A*	*0.1431 (G)*	*$2.31*10^{-01}$*	*0.99 (0.96-1.01)*	*$9.49*10^{-01}$*	*1.00 (0.99-1.01)*
	rs1202181	*87216150*	*7*	*G, A*	*0.3418 (G)*	*$2.19*10^{-01}$*	*0.99 (0.96-1.01)*	*$9.26*10^{-01}$*	*1.00 (0.99-1.01)*
	rs12535512	*87220334*	*7*	*T, C*	*0.4990 (C)*	*$8.80*10^{-01}$*	*1.00 (0.98-1.02)*	*$9.57*10^{-01}$*	*1.00 (0.99-1.01)*
	rs1858923	*87221216*	*7*	*A, G*	*0.3418 (G)*	*$2.22*10^{-01}$*	*0.99 (0.97-1.01)*	*$3.51*10^{-01}$*	*1.00 (0.99-1.01)*
	rs17149792	*87224251*	*7*	*C, T*	*0.3419 (T)*	*$8.68*10^{-02}$*	*0.94 (0.88-1.01)*	*$4.18*10^{-01}$*	*0.99 (0.96-1.02)*
	rs2214104	*87224429*	*7*	*A, G*	*0.4350 (G)*	*$6.00*10^{-02}$*	*0.93 (0.88-0.98)*	*$4.70*10^{-01}$*	*0.98 (0.95-1.00)*
	rs2188525	*87224772*	*7*	*C, A*	*0.4739 (A)*	*$4.58*10^{-01}$*	*0.92 (0.88-0.98)*	*$2.16*10^{-01}$*	*0.97 (0.95-1.00)*
	rs3213619	*87230193*	*7*	*A, G*	*0.0235 (G)*	*$4.58*10^{-01}$*	*0.92 (0.87-0.98)*	*$3.99*10^{-01}$*	*0.97 (0.95-1.00)*
	rs4728709	*87233602*	*7*	*G, A*	*0.0390 (A)*	*$7.96*10^{-02}$*	*0.93 (0.89-0.97)*	*$3.90*10^{-01}$*	*0.97 (0.95-1.00)*
	rs17149810	*87233989*	*7*	*C, T*	*0.0389 (T)*	*$7.90*10^{-02}$*	*0.93 (0.89-0.97)*	*$3.88*10^{-01}$*	*0.98 (0.96-1.00)*
	rs17328288	*87237187*	*7*	*T, C*	*0.0389 (C)*	*$3.59*10^{-01}$*	*0.92 (0.87-0.97)*	*$1.82*10^{-01}$*	*0.97 (0.95-0.99)*
	rs28381787	*87237759*	*7*	*C, T*	*0.0621 (T)*	*$3.56*10^{-01}$*	*0.92 (0.87-0.97)*	*$2.10*10^{-01}$*	*0.97 (0.95-1.00)*
	rs4148731	*87239329*	*7*	*G, A*	*0.0621 (A)*	*$6.06*10^{-02}$*	*0.94 (0.87-1.00)*	*$3.52*10^{-01}$*	*0.99 (0.95-1.02)*
	rs4148730	*87239351*	*7*	*A, G*	*0.0390 (G)*	*$6.06*10^{-02}$*	*0.94 (0.87-1.00)*	*$3.65*10^{-01}$*	*0.99 (0.95-1.02)*
	rs28381780	*87246595*	*7*	*A, G*	*0.0390 (G)*	*$3.47*10^{-01}$*	*0.92 (0.87-0.97)*	*$1.66*10^{-01}$*	*0.97 (0.95-0.99)*
	rs28381779	*87246983*	*7*	*G, A*	*0.0232 (A)*	*$8.22*10^{-01}$*	*1.00 (0.97-1.04)*	*$2.45*10^{-01}$*	*1.01 (0.99-1.02)*
	rs28381775	*87247873*	*7*	*T, C*	*0.0232 (C)*	*$6.84*10^{-02}$*	*0.94 (0.87-1.00)*	*$3.76*10^{-01}$*	*0.99 (0.95-1.02)*
	rs17328447	*87260146*	*7*	*T, C*	*0.0391 (C)*	*$3.30*10^{-01}$*	*0.92 (0.87-0.97)*	*$1.61*10^{-01}$*	*0.97 (0.95-0.99)*
	rs28381767	*87260242*	*7*	*G, A*	*0.1264 (A)*	*$3.30*10^{-01}$*	*0.92 (0.97-0.87)*	*$1.61*10^{-01}$*	*0.97 (0.95-0.99)*
	rs10231033	*87261580*	*7*	*A, G*	*0.0232 (G)*	*$7.77*10^{-02}$*	*0.94 (0.88-1.01)*	*$3.99*10^{-01}$*	*0.99 (0.96-1.02)*
	rs10276499	*87261736*	*7*	*T, C*	*0.0391 (C)*	*$7.39*10^{-02}$*	*0.93 (0.89-0.97)*	*$2.79*10^{-01}$*	*0.98 (0.96-1.00)*
	rs10264856	*87262581*	*7*	*G, A*	*0.0391 (A)*	*$7.35*10^{-02}$*	*0.93 (0.89-0.97)*	*$2.88*10^{-01}$*	*0.98 (0.96-1.00)*
	rs17250003	*87262723*	*7*	*A, G*	*0.0233 (G)*	*$3.28*10^{-01}$*	*0.92 (0.87-0.93)*	*$1.59*10^{-01}$*	*0.97 (0.94-0.99)*
	rs17149840	*87262943*	*7*	*G, A*	*0.0626 (G)*	*$7.38*10^{-02}$*	*0.94 (0.87-1.01)*	*$3.81*10^{-01}$*	*0.99 (0.95-1.02)*
	rs2214101	*87269497*	*7*	*T, C*	*0.0626 (C)*	*$4.73*10^{-01}$*	*0.92 (0.88-0.97)*	*$1.31*10^{-01}$*	*0.98 (0.96-0.99)*
	rs2188532	*87273134*	*7*	*C, T*	*0.0391 (T)*	*$3.25*10^{-01}$*	*0.92 (0.87-0.97)*	*$1.94*10^{-01}$*	*0.97 (0.95-1.00)*
	rs10275831	*87275107*	*7*	*C, T*	*0.0233 (T)*	*$7.68*10^{-02}$*	*0.94 (0.88-1.01)*	*$3.86*10^{-01}$*	*0.99 (0.96-1.02)*

	rs10267099	*87278760*	*7*	*G, A*	*0.0624 (G)*	*$3.16*10^{-01}$*	*0.99 (0.96-1.01)*	*$8.95*10^{-01}$*	*1.00 (0.99-1.01)*
	rs12539395	*87279856*	*7*	*C, T*	*0.0391 (T)*	*$7.15*10^{-02}$*	*0.93 (0.89-0.97)*	*$2.90*10^{-01}$*	*0.98 (0.96-1.00)*
	rs28746492	*87287880*	*7*	*A, G*	*0.0233 (G)*	*$7.66*10^{-02}$*	*0.94 (0.88-1.01)*	*$3.86*10^{-01}$*	*0.99 (0.96-1.02)*
	rs6951067	*87288324*	*7*	*T, C*	*0.2413 (C)*	*$7.14*10^{-02}$*	*0.93 (0.89-0.97)*	*$2.90*10^{-01}$*	*0.98 (0.96-1.00)*
	rs6957599	*87315881*	*7*	*G, A*	*0.0626 (A)*	*$7.53*10^{-02}$*	*0.94 (0.88-1.01)*	*$3.81*10^{-01}$*	*0.99 (0.95-1.02)*
	rs17250255	*87322386*	*7*	*A, G*	*0.0233 (G)*	*$2.90*10^{-01}$*	*0.92 (0.87-0.97)*	*$1.41*10^{-01}$*	*0.97 (0.95-0.99)*
	rs17328880	*87323145*	*7*	*C, T*	*0.0626 (T)*	*$2.91*10^{-01}$*	*0.92 (0.87-0.97)*	*$1.26*10^{-01}$*	*0.97 (0.94-0.99)*
	rs7796247	*87324386*	*7*	*G, A*	*0.0233 (A)*	*$7.53*10^{-02}$*	*0.94 (0.88-1.01)*	*$4.34*10^{-01}$*	*0.99 (0.96-1.02)*
	rs6465118	*87330423*	*7*	*G, A*	*0.0392 (A)*	*$4.33*10^{-01}$*	*0.92 (0.88-0.97)*	*$1.84*10^{-01}$*	*0.98 (0.96-1.00)*
SCN1A	*rs10208529*	*27786188*	*2*	*A, T*	*0.2774 (T)*	*$7.24*10^{-01}$*	*1.00 (0.98-1.03)*	*$2.09*10^{-01}$*	*0.99 (0.98-1.00)*
	rs4989185	*10123918*	*2*	*A, T*	*0.2025 (T)*	*$5.04*10^{-01}$*	*0.99 (0.97-1.02)*	*$2.83*10^{-01}$*	*0.99 (0.98-1.01)*
	rs55722221	*101611283*	*2*	*G, C*	*0.1355 (C)*	*$3.90*10^{-01}$*	*1.01 (0.98-1.05)*	*$3.11*10^{-01}$*	*1.01 (0.99-1.02)*
	rs4849269	*114486907*	*2*	*T, C*	*0.4422 (T)*	*$1.93*10^{-01}$*	*0.99 (0.97-1.01)*	*$1.54*10^{-01}$*	*0.99 (0.98-1.00)*
	rs4849266	*114455116*	*2*	*C, T*	*0.4424 (C)*	*$2.13*10^{-01}$*	*0.99 (0.97-1.01)*	*$1.58*10^{-01}$*	*0.99 (0.98-1.00)*
	rs4849264	*114454908*	*2*	*T, G*	*0.4422 (T)*	*$2.26*10^{-01}$*	*0.99 (0.97-1.01)*	*$1.76*10^{-01}$*	*0.99 (0.98-1.00)*
	rs4849265	*114455006*	*2*	*T, C*	*0.4423 (T)*	*$2.36*10^{-01}$*	*0.99 (0.97-1.01)*	*$1.79*10^{-01}$*	*0.99 (0.98-1.00)*
	rs4849268	*111694330*	*2*	*G, A*	*0.2931 (G)*	*$3.84*10^{-01}$*	*1.01 (0.99-1.03)*	*$1.88*10^{-01}$*	*1.00 (0.99-1.01)*
	rs4849267	*114469325*	*2*	*A, T*	*0.1285 (T)*	*$7.49*10^{-01}$*	*1.01 (0.97-1.04)*	*$2.67*10^{-01}$*	*1.00 (0.99-1.01)*
	rs4849262	*114452256*	*2*	*G, A*	*0.4219 (A)*	*$9.05*10^{-01}$*	*1.00 (0.98-1.02)*	*$6.47*10^{-01}$*	*1.01 (1.00-1.02)*
	rs4849260	*114449695*	*2*	*T, C*	*0.3738 (C)*	*$9.34*10^{-01}$*	*1.00 (0.98-1.02)*	*$8.81*10^{-01}$*	*1.00 (0.99-1.02)*
	rs4849263	*114454721*	*2*	*A, T*	*0.4102 (T)*	*$9.40*10^{-01}$*	*1.00 (0.98-1.02)*	*$9.10*10^{-01}$*	*0.99 (0.98-1.00)*
	rs1813502	*166846016*	*2*	*G, A*	*0.4190 (A)*	*$6.98*10^{-01}$*	*1.00 (0.97-1.02)*	*$5.65*10^{-02}$*	*1.01 (1.00-1.02)*
	rs10497275	*166846730*	*2*	*A, G*	*0.1354 (G)*	*$2.25*10^{-01}$*	*0.98 (0.95-1.01)*	*$5.29*10^{-01}$*	*1.00 (0.99-1.02)*
	rs552878	*166853054*	*2*	*T, C*	*0.2901 (C)*	*$9.19*10^{-01}$*	*1.00 (0.98-1.02)*	*$3.11*10^{-02}$*	*1.01 (1.00-1.02)*
	rs10497276	*166857000*	*2*	*C, A*	*0.1355 (A)*	*$2.40*10^{-01}$*	*0.98 (0.95-1.01)*	*$5.27*10^{-01}$*	*1.00 (0.99-1.02)*
	rs16851356	*166860634*	*2*	*T, C*	*0.1348 (C)*	*$2.58*10^{-01}$*	*0.98 (0.95-1.01)*	*$3.94*10^{-01}$*	*1.01 (0.99-1.02)*
	rs577306	*166871905*	*2*	*C, A*	*0.3032 (A)*	*$7.81*10^{-01}$*	*1.00 (0.98-1.03)*	*$3.36*10^{-02}$*	*1.01 (1.00-1.02)*
	rs565348	*166873299*	*2*	*A, C*	*0.3016 (C)*	*$7.97*10^{-01}$*	*1.00 (0.98-1.03)*	*$2.40*10^{-02}$*	*1.01 (1.00-1.02)*
	rs10182473	*166873723*	*2*	*T, C*	*0.2713 (C)*	*$6.16*10^{-01}$*	*1.00 (0.97-1.02)*	*$9.14*10^{-01}$*	*1.00 (0.99-1.01)*

	rs498631	*166877177*	*2*	*C, T*	*0.4194 (C)*	*$7.03*10^{-01}$*	*1.00 (0.97-1.02)*	*$3.84*10^{-02}$*	*1.01 (1.00-1.02)*
	rs1020853	*166879782*	*2*	*G, T*	*0.2711 (T)*	*$6.26*10^{-01}$*	*1.00 (0.97-1.02)*	*$6.65*10^{-01}$*	*1.00 (0.99-1.01)*
	rs10208532	*168818426*	*2*	*C, T*	*0.1414 (T)*	*$9.44*10^{-01}$*	*1.00 (0.97-1.03)*	*$9.06*10^{-01}$*	*1.00 (0.98-1.01)*
	rs536744	*166881890*	*2*	*T, C*	*0.3571 (C)*	*$7.31*10^{-01}$*	*1.00 (0.98-1.03)*	*$2.63*10^{-02}$*	*1.01 (1.00-1.02)*
	rs478389	*166882828*	*2*	*A, G*	*0.4512 (G)*	*$6.67*10^{-01}$*	*1.00 (0.97-1.02)*	*$4.62*10^{-02}$*	*1.01 (1.00-1.02)*
	rs1834840	*166885720*	*2*	*C, T*	*0.1414 (T)*	*$6.41*10^{-01}$*	*1.00 (0.97-1.02)*	*$1.29*10^{-02}$*	*0.99 (0.98-1.00)*
	rs692995	*166885949*	*2*	*C, A*	*0.3008 (A)*	*$2.31*10^{-01}$*	*0.97 (0.91-1.02)*	*$5.99*10^{-01}$*	*0.99 (0.97-1.02)*
	rs2298771	*166892788*	*2*	*C, T*	*0.4181 (T)*	*$5.68*10^{-01}$*	*0.99 (0.97-1.02)*	*$1.16*10^{-02}$*	*0.99 (0.98-1.00)*
	rs6732655	*166895066*	*2*	*T, A*	*0.2200 (A)*	*$5.86*10^{-01}$*	*0.99 (0.97-1.02)*	*$1.22*10^{-02}$*	*0.97 (0.94-1.01)*
	rs2126152	*166896143*	*2*	*G, A*	*0.3013 (A)*	*$5.67*10^{-01}$*	*0.99 (0.97-1.02)*	*$1.16*10^{-02}$*	*0.99 (0.98-1.00)*
	rs6432860	*166897864*	*2*	*G, A*	*0.0352 (A)*	*$5.67*10^{-01}$*	*0.99 (0.97-1.02)*	*$1.16*10^{-02}$*	*0.99 (0.98-1.00)*
	rs6731591	*166898249*	*2*	*T, C*	*0.3017 (C)*	*$3.71*10^{-01}$*	*1.01 (0.99-1.04)*	*$7.80*10^{-01}$*	*1.00 (0.99-1.01)*
	rs538921	*166903188*	*2*	*A, C*	*0.3017 (C)*	*$2.20*10^{-01}$*	*0.97 (0.91-1.02)*	*$6.31*10^{-01}$*	*0.99 (0.97-1.02)*
	rs13421166	*166903756*	*2*	*C, A*	*0.3017 (A)*	*$5.65*10^{-01}$*	*0.99 (0.97-1.02)*	*$1.21*10^{-02}$*	*0.99 (0.98-1.00)*
	rs1461193	*166904346*	*2*	*G, A*	*0.1922 (A)*	*$5.65*10^{-01}$*	*0.99 (0.97-1.02)*	*$1.21*10^{-02}$*	*0.99 (0.98-1.00)*
	rs1542484	*166905375*	*2*	*A, G*	*0.0353 (G)*	*$8.47*10^{-01}$*	*1.00 (0.98-1.02)*	*$3.52*10^{-03}$*	*0.99 (0.98-1.00)*
	rs3812718	*166909544*	*2*	*C, T*	*0.3017 (T)*	*$8.45*10^{-01}$*	*1.00 (0.98-1.02)*	*$3.46*10^{-03}$*	*0.99 (0.98-1.00)*
	rs3812719	*166909559*	*2*	*C, A*	*0.3017 (A)*	*$5.39*10^{-01}$*	*0.99 (0.97-1.02)*	*$1.07*10^{-02}$*	*1.00 (0.98-1.00)*
	rs1972445	*166910209*	*2*	*G, A*	*0.4379 (A)*	*$8.45*10^{-01}$*	*1.00 (0.98-1.02)*	*$3.46*10^{-03}$*	*0.99 (0.98-1.00)*
	rs922224	*166911912*	*2*	*G, A*	*0.4379 (A)*	*$8.44*10^{-01}$*	*1.00 (0.98-1.02)*	*$3.40*10^{-03}$*	*0.99 (0.98-1.00)*
	rs16851382	*166913475*	*2*	*G, A*	*0.3018 (A)*	*$4.53*10^{-01}$*	*0.99 (0.96-1.02)*	*$5.49*10^{-01}$*	*1.00 (0.99-1.02)*
	rs545331	*166913962*	*2*	*G, A*	*0.4379 (A)*	*$9.28*10^{-01}$*	*1.00 (0.98-1.03)*	*$1.16*10^{-02}$*	*1.01 (1.00-1.03)*
	rs10930201	*166915335*	*2*	*C, A*	*0.4379 (A)*	*$5.30*10^{-01}$*	*0.99 (0.97-1.02)*	*$1.02*10^{-02}$*	*1.00 (0.98-1.00)*
	rs10930202	*166915605*	*2*	*T, C*	*0.1614 (C)*	*$2.65*10^{-01}$*	*0.98 (0.95-1.01)*	*$4.49*10^{-01}$*	*1.01 (1.00-1.02)*
	rs10188577	*166915897*	*2*	*T, C*	*0.2665 (C)*	*$7.72*10^{-01}$*	*1.00 (0.98-1.02)*	*$1.13*10^{-02}$*	*0.99 (0.98-1.00)*
	rs4667866	*166916033*	*2*	*A, G*	*0.3017 (G)*	*$2.40*10^{-01}$*	*0.98 (0.95-1.01)*	*$4.79*10^{-01}$*	*1.01 (0.99-1.02)*
	rs4667867	*166916043*	*2*	*A, G*	*0.1361 (G)*	*$5.30*10^{-01}$*	*0.99 (0.97-1.02)*	*$1.02*10^{-02}$*	*0.99 (0.98-1.00)*
	rs7609055	*166916253*	*2*	*T, C*	*0.3615 (C)*	*$7.25*10^{-01}$*	*1.00 (0.98-1.03)*	*$6.37*10^{-03}$*	*0.99 (0.98-1.00)*
	rs16851400	*166919528*	*2*	*T, G*	*0.1355 (G)*	*$2.07*10^{-01}$*	*0.98 (0.95-1.01)*	*$4.72*10^{-01}$*	*1.01 (1.00-1.02)*
	rs11674130	*166919643*	*2*	*A, C*	*0.3017 (C)*	*$2.18*10^{-01}$*	*0.98 (0.95-1.01)*	*$4.43*10^{-01}$*	*1.01 (1.00-1.02)*

	rs2217198	*166919976*	*2*	*G, A*	*0.4437 (A)*	*$6.82*10^{-01}$*	*1.00 (0.97-1.02)*	*$9.21*10^{-01}$*	*1.00 (0.99-1.01)*
	rs1841548	*166920430*	*2*	*A, G*	*0.1355 (G)*	*$5.59*10^{-01}$*	*1.01 (0.98-1.03)*	*$5.77*10^{-01}$*	*1.00 (0.98-1.00)*
	rs7607455	*166920752*	*2*	*T, C*	*0.1360 (C)*	*$5.69*10^{-01}$*	*1.01 (0.98-1.03)*	*$6.20*10^{-01}$*	*1.00 (0.99-1.00)*
	rs1427651	*166920813*	*2*	*T, C*	*0.3365 (C)*	*$2.31*10^{-01}$*	*0.98 (0.95-1.01)*	*$4.15*10^{-01}$*	*1.01 (0.99-1.02)*
	rs557222	*166920909*	*2*	*G, A*	*0.2005 (A)*	*$9.01*10^{-01}$*	*1.00 (0.98-1.02)*	*$7.78*10^{-03}$*	*1.01 (1.00-1.02)*
	rs492299	*166921698*	*2*	*A, G*	*0.2153 (G)*	*$9.19*10^{-01}$*	*1.00 (0.98-1.03)*	*$1.20*10^{-02}$*	*1.01 (1.00-1.03)*
	rs13397210	*166922028*	*2*	*C, T*	*0.1360 (T)*	*$5.59*10^{-01}$*	*1.01 (0.98-1.03)*	*$5.85*10^{-01}$*	*1.00 (0.98-1.01)*
	rs12998913	*166922755*	*2*	*T, C*	*0.1355 (C)*	*$1.98*10^{-01}$*	*0.98 (0.95-1.01)*	*$5.24*10^{-01}$*	*1.00 (0.99-1.02)*
	rs1824549	*166924971*	*2*	*T, G*	*0.3731 (G)*	*$6.71*10^{-01}$*	*1.00 (0.97-1.02)*	*$9.89*10^{-01}$*	*1.00 (0.99-1.01)*
	rs11692675	*166926428*	*2*	*T, C*	*0.2669 (C)*	*$7.62*10^{-01}$*	*1.00 (0.97-1.02)*	*$3.02*10^{-03}$*	*0.98 (0.97-0.99)*
	rs1020852	*166927659*	*2*	*C, T*	*0.2005 (T)*	*$5.39*10^{-01}$*	*1.01 (0.98-1.04)*	*$5.91*10^{-01}$*	*1.00 (0.98-1.01)*
	rs2169312	*166927896*	*2*	*A, G*	*0.1363 (G)*	*$3.71*10^{-01}$*	*1.01 (0.99-1.04)*	*$7.58*10^{-01}$*	*1.00 (0.99-1.01)*
	rs12613942	*166933023*	*2*	*A, C*	*0.3367 (C)*	*$2.48*10^{-01}$*	*0.98 (0.95-1.01)*	*$4.99*10^{-01}$*	*1.00 (0.99-1.02)*
	rs484926	*166937584*	*2*	*G, A*	*0.3605 (A)*	*$4.63*10^{-01}$*	*0.99 (0.96-1.02)*	*$6.08*10^{-01}$*	*1.00 (0.99-1.02)*
	rs11884723	*166938268*	*2*	*C, T*	*0.2774 (T)*	*$4.86*10^{-01}$*	*1.01 (0.98-1.04)*	*$2.99*10^{-03}$*	*1.02 (1.01-1.03)*
	rs497594	*166941773*	*2*	*A, G*	*0.2003 (G)*	*$6.01*10^{-01}$*	*0.99 (0.97-1.02)*	*$9.07*10^{-01}$*	*1.00 (0.99-1.01)*
	rs13004083	*166945367*	*2*	*A, G*	*0.1013 (G)*	*$9.25*10^{-01}$*	*1.00 (0.98-1.03)*	*$4.61*10^{-03}$*	*1.02 (1.01-1.03)*
	rs580041	*166950510*	*2*	*C, T*	*0.0497 (T)*	*$3.87*10^{-01}$*	*0.99 (0.97-1.01)*	*$2.62*10^{-03}$*	*0.99 (0.98-1.00)*
	rs498918	*166951160*	*2*	*G, A*	*0.1941 (A)*	*$5.88*10^{-01}$*	*0.99 (0.97-1.02)*	*$1.07*10^{-02}$*	*0.99 (0.98-1.00)*
	rs12469308	*166953372*	*2*	*C, T*	*0.1366 (T)*	*$8.49*10^{-01}$*	*1.00 (0.98-1.03)*	*$5.74*10^{-01}$*	*1.00 (0.99-1.01)*
	rs10930204	*166955439*	*2*	*T, C*	*0.2931 (C)*	*$8.40*10^{-01}$*	*1.00 (0.98-1.03)*	*$5.75*10^{-01}$*	*1.00 (0.99-1.01)*
	rs535533	*166959505*	*2*	*T, C*	*0.3738 (C)*	*$5.42*10^{-01}$*	*1.01 (0.99-1.03)*	*$3.00*10^{-01}$*	*1.00 (0.98-1.00)*
	rs12999167	*166969347*	*2*	*C, A*	*0.4219 (A)*	*$6.55*10^{-02}$*	*1.04 (1.00-1.08)*	*$5.22*10^{-02}$*	*1.02 (1.00-1.03)*
	rs666833	*166982290*	*2*	*C, T*	*0.4102 (C)*	*$5.11*10^{-02}$*	*0.97 (0.94-1.00)*	*$9.27*10^{-03}$*	*0.98 (0.97-1.00)*
	rs533202	*166983181*	*2*	*A, C*	*0.4422 (A)*	*$1.05*10^{-01}$*	*0.97 (0.94-1.01)*	*$1.32*10^{-02}$*	*0.98 (0.97-1.00)*
	rs515090	*166986354*	*2*	*T, C*	*0.4423 (T)*	*$1.10*10^{-01}$*	*0.97 (0.94-1.01)*	*$1.48*10^{-02}$*	*0.98 (0.97-1.00)*
	rs16851582	*166989841*	*2*	*A, G*	*0.4424 (G)*	*$9.56*10^{-01}$*	*1.00 (0.93-1.08)*	*$3.63*10^{-01}$*	*1.01 (0.98-1.05)*
	rs7593275	*166991386*	*2*	*G, T*	*0.1285 (T)*	*$7.40*10^{-01}$*	*1.01 (0.95-1.07)*	*$4.11*10^{-01}$*	*1.01 (0.98-1.04)*
	rs4233806	*166996841*	*2*	*A, G*	*0.4422 (G)*	*$8.67*10^{-01}$*	*1.00 (0.97-1.02)*	*$5.11*10^{-01}$*	*1.00 (0.99-1.01)*
	rs6746010	*167004057*	*2*	*A, C*	*0.1921 (C)*	*$8.95*10^{-01}$*	*1.00 (0.97-1.02)*	*$5.14*10^{-01}$*	*1.00 (0.99-1.01)*

	rs6749736	*167004685*	*2*	*A, G*	*0.2303 (A)*	*$8.16*10^{-01}$*	*1.00 (0.98-1.03)*	*$8.38*10^{-01}$*	*1.00 (0.99-1.01)*