Next-generation sequencing in the diagnosis of Dementia and Huntington's disease Phenocopy Syndromes

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy from University College London

Carolin Anna Maria Koriath

Institute of Neurology

Department of Molecular Neurodegeneration

University College London

2020

Declaration

I, Carolin Anna Maria Koriath, 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.

However, contemporary, large-scale projects require a complex set of skills and extensive work best provided by a team. Like many studies nowadays, the investigation and analysis of the MRC Dementia Gene Panel study were based on a collaborative effort and contributions from different members of the Human Genetics Programme at the MRC Prion Institute at UCL. I have personally performed a significant amount of the laboratory work and will highlight these contributions in the methods section; I have also performed all stages of data processing. Namely, I have analysed all next-generation sequencing data and developed criteria for the correct classification of identified variants. For time and cost advantages, samples for the Huntington's disease phenocopy study were whole-genome sequenced and aligned at Edinburgh Genomics, but I performed the data analysis. I have put this data into context with the clinical data for further analysis and the statistical workup, and, in close dialogue with my supervisors Prof Simon Mead and Prof Sarah Tabrizi, have drawn conclusions from all available data which are set out in this PhD thesis.

Abstract

Dementia is a major cause of disability worldwide, especially in the elderly. While Mendelian causes of dementia only account for a small proportion of cases, their role in elucidating the pathophysiology has been paramount. Genetically defined cohorts also offer opportunities for trials of disease-modifying treatments, even before the onset of symptoms. Previously, only a small number of genes could be selected for genetic testing because of cost-restrictions, but the advent of next-generation sequencing has enabled its more widespread use.

This thesis explored the use of next-generation sequencing in patients living with dementia and HD phenocopy (HDPC) syndromes, who include patients with mixed presentations of dementia and motor symptoms. Using a validated 17 gene Dementia Gene panel supplemented by *C9orf72* expansion testing and Apolipoprotein (ApoE) genotyping in over 3000 patients and controls, I determined the success rate of genetic panel testing in dementia; I developed an algorithm for the selection of patients for genetic testing based on the clinical presentation and common predictors of genetic causes of dementia. A detailed analysis of the ApoE data in the frontotemporal dementia cohort revealed strong effects of ApoE4 on age at onset in the subset with proven or suspected tau neuropathology, as well as opposite effects of amyloid-beta pathology.

In order to improve the definition and diagnostic rate of HDPC syndromes, patients who were referred for HD testing from two clinics were compared based on their clinical presentation; patients could not be distinguished based on clinical presentation alone, even if analysed as patterns. Given the low success rate of dementia gene panel testing in the HDPC cohort, 50 patients were selected for whole-genome sequencing based on their HD-likeness and their likelihood of harbouring a Mendelian variant. The results revealed a number of variants of interest but require replication.

Impact Statement

Due to the rising prevalence of dementia, interest in related research has been growing. Genetic causes of dementia have explained underlying processes and paved the way for trials of disease-modifying treatments; however genetic testing is still underutilized in clinic, due to longstanding perceptions of high costs long turn-around times, and lack of treatment options.

Based on over 3000 patient and control samples, the results of my research could improve clinical genetic testing and counselling. I established likely discovery rates for genetic testing in both rare and common dementias, and developed an algorithm for the selection of patients based on predictors of genetic disease. This has implications for diagnostic options of patients living with dementia and their counselling regarding the success rate and the implications for family members. The discovery rates would warrant more wide-spread testing, supporting future treatment trials. I also found many mutations in dementia genes not previously linked to the syndrome, many that had not been described previously, and that some patients carried more than one deleterious mutation. This means that picking genes based on clinical presentation risks missing mutations, that the known genes still harbour undescribed causal mutations, and that concurrent mutations may be missed if few genes are sequenced, with consequences for predictive testing of relatives. I could show that ApoE4, a risk factor for Alzheimer's disease, was linked to earlier disease onset in patients with frontotemporal dementia and tau neuropathology, which may make in relevant in other types of dementia.

Harnessing the statistical power of this dataset, I was able establish a frequency threshold in the population for mutations expected to cause early-onset dementia (EOD) in all patients who carry it (100% penetrance). The findings imply that not all variants published as deleterious will cause disease, and that mutations found in patients need to be compared to population databases to prove pathogenicity. This has direct impact on the assessment of novel variants and which ones are reported

back to patients. This work has been published to disseminate it to clinician's the world over.

Focusing on a subset of patients with poor diagnostic rates using focused genetic testing, I attempted to improve the definition and diagnostic rate of Huntington's disease phenocopy (HDPC) syndromes which include mixed presentations of dementia and motor symptoms. They are an excellent example of unclear neurodegenerative disease; clarifying their definition would improve the prediction of whose Huntington's disease (HD) test will be positive and, consequently, counselling of patients. However, HD and HDPC patients could not be distinguished based on clinical presentation alone, even if analysed as patterns. In a further attempt to better define these HDPC syndromes, I analysed whole-genome sequencing (WGS) of 50 HDPC patients to find a genetic link between these patients. However, with the exception of one known HDPC syndrome, none of the suspicious variants discovered in the dataset could be sufficiently verified as causal without replication. This requires further work, but could ultimately help elucidate the affected pathways both in patients with HD and HDPC syndromes.

Acknowledgements

I am very grateful to patients and volunteers who have donated blood samples and their time to research and have made this study possible. I would like to thank my PhD supervisors Professor Simon Mead, Professor Sarah Tabrizi, Dr Jonathan Rohrer and Dr Edward Wild for their continued support and guidance. I would also like to thank Gary Adamson, Ronald Druyeh, William Taylor, Athanasios Dimitriadis, Penny Norsworthy, Tracy Campbell, and Lee Darwent for their tireless work in the laboratory, as well as Dr Janna Kenny and Dr Jon Beck for the work they previously dedicated to the gene-panel sequencing project. I would like to thank Dr Holger Hummerich, Fernando Guntoro, and Liam Quinn for bio-informatic support, Irina Neves Simoes and Dr TzeHow Mok for their help gathering clinical information, and Helen Speedy for help accessing the Genomics England project.

I would like to thank the Leonard Wolfson Foundation for funding my PhD, the CHDI Foundation, and particularly Dr Thomas Vogt, for their funding of the HD phenocopy whole-genome sequencing project, the Biomedical Research Unit at UCL, the Medical Research Council UK and the MRC Institute for Prion diseases at UCL for their support of this project.

Finally, I would like to thank my parents Dres. Joachim and Elisabeth Cyran for their support over the years, my husband Alexander Koriath for his love and for his interest in my PhD project and thesis, as well as my sons Carl Oscar and Conrad Otto for the welcome distraction they provided.

Table of Contents

Table of Figures

Table of Tables

Abbreviations

ACMG	American College of Medical Genetics
AD	Alzheimer's disease
AAO	Age at onset
AMP	Association for Molecular Pathology
ApoE	Apolipoprotein E
CADD score	Combined Annotation Dependent Depletion score
CNS	Central nervous system
CNV	Copy number variant
DemMot	Dementia with motor symptoms (cohort)
DV	Deleterious and/or likely deleterious variant
EO	Early-onset
EOAD	Early-onset Alzheimer's disease
EOFTD	Early-onset frontotemporal dementia
ExAC	Exome Aggregation Database Consortium
FTD	Fronto-temporal dementia
GnomAD	Genome Aggregation Database
GOSH	Great Ormond Street Hospital
GS	Goldman score
GTEx	Genotype-Tissue Expression
HD	Huntington's disease
HDPC	Huntington's disease phenocopy
IVA	Ingenuity Variant Analysis ®
LO	Late-onset

MRC	Medical Research Council (UK)
NGS	Next-generation sequencing
NHNN	National Hospital for Neurology and Neurosurgery at Queen Square, London
OPRI	Octapeptide repeat insertion
PCR	Polymerase chain reaction
RVIS	Residual Variation Intolerance Score
UTR	Untranslated region
WES	Whole-exome sequencing
WGS	Whole-genome sequencing

I. Introduction

1 Dementia(s) through the ages

A. Traditional hurdles and new challenges

In recent years, interest in dementia research has been rising due to its increasing prevalence in an aging society[1]. Although most dementias appear sporadic, occur in old age[2], and are highly polygenic, in that risk is conferred by tens or hundreds of common, modest-effect loci[3,4], familial forms of early-onset dementia (EOD) with Mendelian inheritance (such as familial Alzheimer's disease (AD), familial fronto-temporal dementia (FTD) or inherited prion disease) have been crucial to furthering our understanding of the underlying clinical-pathological processes, and the ensuing development of animal models and experimental therapeutics[5]. One already established opportunity to make a precise diagnosis at the very early stages of disease or trials in presymptomatic patients is through genetic testing for these highly penetrant single gene disorders.

Because of a series of high-profile failures of advanced clinical trials, clinical research has focussed on testing therapies earlier in the disease using imaging and CSF biomarkers to support a diagnosis before the onset of cognitive symptoms[1]. Clinical genetic studies offer the potential for a presymptomatic diagnosis in at-risk individuals with a high degree of confidence about molecular pathology. Indeed, individuals carrying high penetrance mutations may be the most appropriate groups in whom to test experimental therapeutics to prevent or delay neurodegeneration – especially if those therapeutics were developed using animals expressing mutant human proteins[6]. Familial forms of dementia are now being leveraged to study cutting-edge pharmacological therapies in diseases where, thanks to the known genetic cause, the pathophysiological mechanism is better understood than in sporadic forms of disease. Therapeutic drug trials such as DIAN-TU in Alzheimer's disease (AD)[7], *GRN* replacement in frontotemporal dementia (FTD)[8], and the gene-silencing trial in HD[9] not only offer hope to the specific patient groups included in the trials, but the data gained from them will also improve our understanding of disease mechanisms more generally.

Several factors have historically stopped clinicians from considering a clinical genetic test in patients with dementia: perceptions that finding a genetic cause is unlikely and futile[10,11], high costs, the length of time to return results, genetic heterogeneity (multiple genes causing the same pathology or clinical syndrome), and the lack of disease-modifying treatment options or opportunities for trials. In dementia, obtaining informed consent and the timing of doing so may also be more complicated than in other diseases, since in the early stages when the wishes of the patient may be clearer, there may be uncertainty about which gene test(s) to request as the clinical picture has not fully evolved.

These problems have been exacerbated recently because of a high rate of gene discovery and heterogeneity, particularly in frontotemporal dementia, with many genes not becoming available for clinical testing in a timely manner[12]. Furthermore, other discoveries show marked pleiotropy, for example, the *C9orf72* expansion mutation being found in patients with clinically-diagnosed frontotemporal dementia (FTD), amyotrophic lateral sclerosis (ALS), Huntington's disease-like syndromes (HD-like syndromes), and Alzheimer's disease (AD)[13,14]. Rapid advances in technology, with next-generation sequencing (NGS)-based gene-panels now firmly established and whole-exome and whole-genome sequencing (WES and WGS, respectively) increasingly widely available, allow the circumvention of some of these problems by examining multiple genes, or even the whole exome or genome, simultaneously. However, numbers of patients living with dementia included in genetic studies have usually been limited and cannot provide the statistical power needed to support firm genetic evidence of pathogenicity of very rare variants or clinical decision making[13]. Most genetic research has also been focused on the most common dementia syndromes of AD and FTD, with little effort diverted to patients with less established presentations.

B. Clinical heterogeneity in autosomal dominant dementia

a. Genetic factors in dementia syndromes

Dementia is characterized by the loss of previously held cognitive skills compromising social and occupational functioning; it is a major cause of disability worldwide, especially in the elderly[15]. Depending on the syndrome, the disease may ravage disparate abilities from memory through speech and language to social cognition[16,17]. Familial forms of early-onset dementia (EOD) with Mendelian inheritance (such as familial AD, FTD or HD) have been crucial to furthering our understanding of the underlying clinical-pathological processes[18].

Based on the strong predictive value of a history of dementia in the family, hereditary factors are thought to have a considerable impact on the presentation of the common types of dementia[19,20]. In recent years, the number of genes known to cause dementia has expanded considerably (genetic heterogeneity), as has our understanding of the spectrum of phenotypes caused by mutations in a single gene (pleiotropy) and the clinical overlap between different dementia syndromes.

However, as explained above, genetic testing is currently underutilised in clinical practice[21], leaving deleterious variants with unusual presentations or incomplete penetrance[22] most likely underdiagnosed. Genetic causes for dementia can have diverse underpinnings (see Table I-1), not all of which can be picked up by all testing methods. A detailed neurological examination and history, including family history, must form the basis of any assessment. The family history can be quantified using for example the Goldman score (GS)[23,24,25] and helps to distinguish the patients most likely to carry a genetic causal variant[24] and to decide who, when and what to test[25]. Relatives' symptoms may potentially aid an earlier diagnosis[26], making eventual treatment more effective. An important pitfall to be avoided is mistaking a censored family history for a negative one as neither precludes a genetic diagnosis: this may be due to factors including the difficulty of accurate diagnosis or misdiagnosis, early deaths from other causes in previous generations, as well as anticipation in some

diseases. A genetic test may therefore be warranted even in cases without a significant family history[27].

Numerous familial cases remain unexplained; many contributing genetic factors remain to be identified and not all variants may be fully-penetrant Mendelian causes of disease[28]. With the number of genes known to cause dementia, our understanding of gene pleiotropy and a clinical overlap between dementia syndromes expanding considerably[13,29,30], it has become clear that testing only one or even a handful of cherry-picked genes may miss significant numbers of known deleterious variants and concurrent mutations.

On the other hand, some genetic variants may only confer very small increases in risk. Genome-wide association studies (GWAS) have begun to elucidate some of the likely interactions and pathways involved; for example, genes involved in the lipid pathway, inflammation and endocytosis appear to be implicated in Alzheimer's disease[31]. When combined as polygenic risk scores, multiple GWAS-discovered common variants of weak effects may have useful predictive functions, but as yet, these tools have not entered clinical decision making as the implications for an individual's risk remain too uncertain to be of clinical benefit[32].

Table I-1: Mutation types found in different forms of dementia

Mutation type	Found in (in exemplum)	Effect
Single nucleotide variation (SNV)	Missense mutations in *PSEN1* and *PSEN2* causing AD	Increases the amyloid-beta burden[33]
Small insertions and deletions (Indels)	Indels in *PSEN1* causing AD	Increases the amyloid-beta burden[33]
Large exon deletions / insertions/ duplications/copy number variants	*APP* duplication causing AD	An extra copy of *APP* causes an increase in amyloid-beta[34]
Splice-site mutations	Splice site mutations in *GRN* and *MAPT* causing FTD	*GRN*: loss of function mutations[35] *MAPT*: differential splicing causing an imbalance between 3-repeat and 4-repeat tau[36]
Triplet repeat expansions	CAG expansion in *HTT* causing HD; 'GGGGCC' repeat expansion in *C9orf72* causing FTD/ALS/HD-like syndromes	Gain-of-function effects and differential splicing for *HTT*[37,38]; Gain and loss-of-function effects for the *C9orf72* expansions[39].
Chromosomal region inversions	The genomic inversion of *MAPT* is linked to tangle-dominant dementia[40]	Causes extracellular 3-repeat and 4-repeat tau tangles[41]

b. *Alzheimer's disease*

Overall, Alzheimer's disease (AD) is the most common dementia, both for early onset (<65years)[18] as well as late onset disease. Patients typically present initially with symptoms affecting their memory or orientation; this reflects the early, even presymptomatic impact of the disease on the patient's hippocampus[42-44]. The majority

of AD cases are diagnosed in patients developing symptoms after the age of 65[33,45]; these late-onset cases are most often considered sporadic for lack of a strong history of dementia in the family, nonetheless hereditary factors still appear to influence familial incidence and disease manifestation[19] and tau heterogeneity modifies the speed of progression[46]. While rare, early-onset AD (EOAD) in patients who develop symptoms before reaching 60 years of age has a devastating impact on patients and families often hit in the rush hour of life[47]. In addition, the low prevalence of EOAD[48] compared to more common differential diagnoses like depression, and an initial presentation with often less typical symptoms, frequently leads to a delayed diagnosis[49].

The most common causes of familial AD are *PSEN1*, *PSEN2* and *APP* genes and *APP* duplications - with *PSEN1* being the most common gene with a younger age at onset (AAO) than *APP* and *PSEN2*, although mutations in each of the genes may cause disease into old age. The discovery of these causal mutations and genes has contributed immeasurably to our understanding of the amyloid beta protein metabolism[5] and its links the innate immune system[50,51]. In addition, the fact that AAO for a given mutation, while relatively consistent within families[52], may vary between families[53], as well as reports of reduced penetrance in some cases[54-56], suggests the influence of other, as yet unknown, genetic modifiers.

AD is often used as a default diagnosis in patients presenting with dementia and therefore a family history of "Alzheimer's disease" should be treated cautiously in the absence of biomarker or pathological confirmation. Similarly, the typical progressive amnestic syndrome of AD may also occasionally be mimicked by non-AD genetic dementias such as FTD due to *MAPT* mutations[57] or inherited prion disease due to octapeptide repeat insertions (OPRI) in *PRNP*[58]. Conversely, autosomal dominant AD may present atypically, including with seizures and myoclonus[59,60]; *PSEN1* more than *APP* mutations are linked to atypical symptoms resembling FTD, prion disease or dementia with motor symptoms[24,60-63], and *PSEN2* mutations may lead to delusions and hallucinations[64].

In the case of AD, in addition to the autosomal dominant mutations, apolipoprotein E (ApoE, encoded by *APOE*) deserves a mention. It remains by far the most important single genetic risk factor for LOAD[32] and exists in three isoforms – ApoE2 (population prevalence approx. 8-10%), ApoE3 (approx. 70-78%), and ApoE4 (approx. 14-20%)[65,66], although allele prevalences and their deleterious effects vary across populations[67]. Possession of an ApoE2 allele is protective against AD, while an ApoE4 confers risk, with odds ratios ranging from 1.8 to 9.9[68-70] and the risk being additive[71]. However, possession of an ApoE4 allele is neither necessary nor sufficient to cause AD; clinical genetic testing is therefore not recommended.

c. *Frontotemporal dementia*

FTD is the second commonest early-onset dementia[72,73] and has always been known to be a heterogeneous disorder, comprising multiple clinical and pathological conditions[73]. FTD patients most usually suffer from one of the canonical FTD syndromes (behavioural FTD (BVFTD) and the primary progressive aphasias (PPA)), depending on whether their frontal or temporal lobes are most affected[74], but may also present with an FTD spectrum disorder (amyotrophic lateral sclerosis (ALS), corticobasal syndrome (CBS)[75], progressive supranuclear palsy (PSP)) or non-FTD disorders including 'AD', 'Parkinson's disease (PD)' and 'Huntington's disease (HD)' phenotypes[76].

Around 30% of FTD is familial[20,77], but heritability of FTD varies by phenotype and family history[20]; the most heritable form is bvFTD, followed by FTD-ALS, then PPA and the atypical parkinsonian syndromes[20,78-80]. *MAPT*, *GRN, C9orf72* are the most commonly involved genes[81-83]; rarer mutations have been identified in *TBK1, VCP, CHMP2B, FUS, SQSTM1, TARDBP, CHCHD10, TIA1* and *CCNF*[83], which have been linked to FTD's diverse neuropathological features and molecular mechanisms including the, DNA metabolism and the lysosomal pathway[84].

Some clinical presentations are more typically associated with certain genes, such as ALS and FTD-ALS overlap with *C9orf72* expansions rather than *GRN* or *MAPT*

mutations, and PPA with *GRN* mutations. Typical cognitive profiles also differ between *GRN*, *MAPT* and *C9orf72* expansion carriers, with *GRN* linked to worsening attention, *MAPT* to impaired memory function, and *C9orf72* to global impairment[75]. Age at onset is variable in all of the forms of genetic FTD but particularly so in *GRN* mutations and *C9orf72* expansion carriers[76]. As an additional complicating factor, there have been reports of families with pathogenic mutations in both *C9orf72* and one of the other FTD-related genes[24,85]. Despite these discoveries and neuropathological findings in post-mortem brain tissue, our understanding how all these different pathways converge on a very similar cerebral network, and thereby clinical presentation, remains limited.

d. Prion disease

Overall, prion diseases are rare. Unlike other in other dementias, patients typically progress very rapidly from symptom onset to ultimately death in a mute and bedridden state; the disease course is often no longer than six months compared to the 10 years associated with other forms of dementia[72,86,87]. Prion diseases comprise several different phenotypes: Creutzfeldt-Jakob-Disease (CJD), Gerstman-Scheinker-Straussler disease (GSS), fatal familial insomnia (FFI), octapeptide repeat insertions (OPRI)-related disease, and prion protein systemic amyloidosis. They are defined by the accumulation of abnormal prion protein and typical histopathological changes[28]. While CJD is the most frequent presentation of prion disease and is defined by rapidly progressive dementia, myoclonus and ataxia, GSS combines a frontal syndrome and cerebellar ataxia with peripheral loss of sensation, and FFI patients present with the eponymous fragmented sleep, gait abnormalities and autonomic symptoms. While suggestive features such as myoclonus or insomnia, as well as the rapid pace of deterioration, offer clinical cues to a diagnosis that can be confirmed by highly specific investigations[88] and permit a focused genetic test, atypical presentations of prion disease may still mimic other neurodegenerative disease or even reversible causes of dementia[89].

10-15% of the total incidence of prion disease is being caused by a mutation in *PRNP*[58], but most patients suffer from sporadic forms of the disease where the reason for the

misfolding and spread of the prion protein, and therefore the pathogenesis of the disease, remain unclear; other genes are likely to act as susceptibility factors[90]. Prion diseases have also been shown to be transmissible, as the Kuru epidemic in Papua-New Guinea, the BSE crisis of the 1990s, and a number of iatrogenic cases of prion disease have unfortunately shown[90-93]. These transmissions, however, also led to the discovery of the protective M129V variant in the *PRNP* gene, which all but protects individuals from developing the disease[92,93].

e. Huntington's disease and Huntington's disease phenocopy syndromes

Huntington's disease (HD) is the one of the commonest fatal, autosomal dominant neurodegenerative disorders[38,94] typically defined by a progressive triad of movement, cognitive, and psychiatric symptoms[95]. For many clinicians, chorea is the defining feature of HD, but clinical manifestation is variable; symptoms can range from hyperkinetic to hypokinetic and while unequivocal motor extrapyramidal symptoms are required for the diagnosis of disease onset, patients often experience a preceding prodromal phase with cognitive impairment, anxiety, and depression[96].

HD was the first neurodegenerative disease for which genetic linkage was identified[97]; it is caused by a CAG triplet repeat expansion in the *HTT* gene on chromosome 4[98]. The number of repeats is inversely correlated with the age at onset and often increases from one generation to the next, especially when paternally inherited, a process termed anticipation. In the presence of over 40 repeats, penetrance is complete, with incomplete penetrance between 36 and 39 repeats[99]; age at onset in part depends on the repeat number but is also influenced by other genetic modifiers, such as DNA repair genes[100] and gene promoters[101]. Thanks to the suggestive combination of symptoms, the diagnostic rate for the HD test is high, and historically, when the test first became available only approximately 1% of patients with HD-compatible symptoms tested negative for the *HTT* expansion[102].

However, more recently the low cost and ready availability of the HD test has increased the negative test rate as clinicians may wish to exclude the disorder even if the clinical syndrome is atypical and a proportion of those in whom HD is suspected do

not carry the pathognomonic mutation. These patients are said to have an HD phenocopy syndrome and range from those mimicking HD exactly, to those with partially overlapping clinical features. If the HD test is negative, the differential diagnosis is wide[95] and testament to the existing overlap between different dementia syndromes; recently, *C9orf72* has been identified as their most frequent identified genetic cause[13] while other cases have been identified to carry other known mutations: SCA17, HDL2 (familial prion disease caused by a *PRNP* octapeptide expansion), and Friedreich's ataxia at a lower frequency[103].

Despite these efforts, the cause and pathogenesis of a large proportion of these cases with dementia and motor symptoms remains to be elucidated. HD phenocopy syndromes (HDPC syndromes) are an excellent real-world example of unclear neurodegenerative disease presenting with symptoms spanning cognitive, psychiatric and motor disorders, and not being the main focus of any of the specialists concerned with these respective domains. This is unfortunate as there is much that can be learnt from looking at sets of genes which are linked to the same pathology or clinical syndrome. In related neurodegenerative diseases like AD or FTD, associated with mutations in several different genes, it has been insightful to study the functional categories or biological pathways which link disease genes.

Overall, clinical diagnoses in patients with dementia syndromes are often not clear cut because of significant overlap and because, based on the evidence so far, also seem to share some common genetic pathways. All the while, dependable clinical predictors of genetic disease are proving elusive. Despite the need for accurate diagnoses for future clinical trials of disease-modifying therapies, genetic testing in dementia is severely underutilized; only a fraction of patients undergo genetic testing and even fewer receive a certain genetic diagnosis.

2 Advances in genetic sequencing methods

Recent increases in the availability of genetic testing, including whole-genome sequencing, and expanding knowledge about implicated genes and pathways are transforming diagnostic genetic testing from a niche to a more mainstream

application. However, little consensus has been reached on which patients living with dementia should be offered a genetic test.

A. Single gene testing

Testing single genes for point mutations and short frame-shift caused by small insertions or deletions can be performed on most genes with good results; however, it can be laborious, depending on the size of the gene, and typically does not detect larger insertions or deletions, intronic mutations or copy number variations. Where it comes to diagnostic testing of most dementia conditions, single gene testing has much lower odds of detecting a causal variant; it is therefore increasingly being supplanted by other, newer sequencing methods (see below). In addition, testing only one or even a handful of cherry-picked genes does not enable the detection of concurrent mutations; recent evidence demonstrates the significant occurrence of concurrent mutations in AD, FTD, prion disease and dementia with motor symptoms, with a notable excess in FTD[24], as previously suggested by case reports of concurrent pathogenic mutations in patients with FTD [85,104,105].

Since the 1980, the polymerase chain reaction (PCR) based sequencing developed by Sanger *et al.* in 1977[106] ("Sanger sequencing") has been the method of choice to generate genomic data, but while it has been constantly improved upon and offers good precision and reliability, it still remains relatively slow and expensive; it also requires fairly large amounts of template DNA. Sanger sequencing generates different length sequences with terminal dedioxynucletotides, which are then separated out and read using capillary electrophoresis. It is labour-intensive, slow if used on a larger scale and prohibitively expensive for larger numbers of exons: sequencing of the first complete human genome took approximately ten years using this method, and cost $2.7 billion[107,108]. Because of its accuracy it is still the preferred Gold standard method for confirmatory testing and testing for a known mutation. Due to methodological constraints, Sanger sequencing is unsuited to the detection of complex variants, such as copy number variants (CNVs), insertions and deletions; this can be achieved using multiplex ligation-dependent probe amplification (MLPA). MLPA differs from standard PCR approaches in that it is the probes that are amplified, not the subject's DNA, which

only serves as a template for the fragmented probe to be hybridized. This means that a single primer pair can amplify dozens of sequences, within one or several genes; the fragments can then be distinguished by their length using capillary electrophoresis[109].

B. Multiple gene testing

In order to cut down on time and labour, newer methods of sequencing have been developed, which are capable of generating information on multiple genes at the same time. Declared method of the year 2007[110], next-generation sequencing allows for high-throughput, accurate and in- parallel amplification of DNA samples, which is also comparatively cost-effective. Targeted next-generation sequencing has become increasingly common; however, most datasets for dementia so far remain small in patient number[28]. Currently one of the most popular options for diagnostic genetic testing are next-generation sequencing gene panels targeted at genes relevant to a certain phenotype, often including rarer causes of disease[111]. Several next-generation sequencing platforms share a common principle: a selection of amplicons of interest is selected using primers; each sample is then barcoded and cleaned before being combined into a sample library. The library is amplified further and finally sequenced and read on the sequencing platform[24]. They are cheap and easy to run, as well as quick to analyse, however, newly discovered genes will only be incorporated with a delay, and gene panels are unable to detect repeat disorders or copy number variants (CNVs). Gene panels are most suitable for conditions where most cases are caused by a limited number of known genes, but can cater to diagnostic uncertainty by covering several related conditions (for example AD and FTD[24]). Most laboratories will use focused gene panels, either by sequencing amplicon-selected regions of interest or by performing whole-exome sequencing (WES)[112] or whole-genome sequencing (WGS) (see below) and subsequently restricting their analysis to genes selected based on the referral diagnosis. For a dementia gene panel, these should include most genes linked to the phenotype of interest, i.e. the AD genes *APP*, *PSEN1*, and *PSEN2*, the canonical FTD genes *GRN* and *MAPT*, but also genes associated with rarer causes of familial dementia and leukencephalopathy[113].

Next-generation sequencing also encompasses whole-exome sequencing (WES)[112] and whole-genome sequencing (WGS). Similar to gene panels, WES relies on PCR-generated amplicons, but they cover all protein-coding regions of the genome[112], thought to contain the vast majority of mutations causing inherited human disease. After sample preparation and barcoding, samples are sequenced on a high-throughput platform. WGS is usually also amplicon-based, but can forego the PCR step with genomes instead sheared into fragments; barcoded libraries are then sequenced on the same platforms as WES samples[114].

Costs for WES and WGS have been falling rapidly in the last years, bringing them into the realm of financially feasible clinical tests, but WGS remains substantially more expensive. Notwithstanding, WGS offers the advantage of interrogating not only the exome, but also the non-coding regions, such as splice sites, promoters, and regulatory sequences, where an increasing number of examples of pathogenic mutations are being identified[24,36,115,116]. In addition, WGS has been shown to provide more homogeneous coverage with higher genotyping quality, to be less susceptible to allelic drop outs[117], and to identify more variants[117,118]. Rates of diagnosis for WES of above 25% in neurological disease in adult simplex cases have been reported[116,119-121], and are likely to be higher for WGS. This is particularly pertinent in dementia, where due to the late age at onset (AAO) trios (a patient and both parents) are rarely available. Still, recruitment of additional unaffected and affected family members may increase the genetic yield by allowing segregation analysis of variants.

WES/WGS is particularly useful for highly heterogeneous disorders[116,122]. Benefits include removed bias and added diagnostic capabilities, such as analysis for structural variation (e.g. duplications, deletions, insertions and inversions of DNA), as well as the possibility of a re-analysis should new information become available or a new gene-disease association be discovered. They may also allow for novel gene discovery. However, they are still significantly more expensive, especially to analyse and store large amounts of data. The analysis of genome sequencing data especially is demanding, involving variants in known causative genes (e.g. a dementia panel of

genes) being identified and interpreted, and the analysis of copy number variations and repeat expansions. Given the usual lack of parent-child trios in late-onset diseases such as dementia, analysis has come to rely on online population databases such as *GnomAD*[123], which have been critical in judgements about the pathogenicity of variants[28,124]. However, it is important to remember that, unlike in the case of rare paediatric disease, pathogenic variants in Mendelian dementia will inevitably end up in population databases, because of the late AOO of disease and the possibility of reduced penetrance. Population databases therefore reflect the healthy young adult and middle-aged population they recruit, and a variant's rare presence in this population does not preclude pathogenicity. In addition, both WES and WGS also introduce the issue of incidental findings in genes unrelated to the syndrome being diagnosed[125,126]. These issues are exacerbated in the case of WGS, with approximately 100 times more data. WGS nonetheless offers additional information with regards to larger insertions and deletions, especially non-coding[24,115], as well as regulatory sequences and technical advantages including more even coverage[127].

C. Testing for nucleotide repeat expansions

Most common modern sequencing platforms struggle to precisely sequence repeat expansions, which can run into hundreds of nucleotide repeats. The current sequencing techniques excel at detecting SNPs and smaller insertions and deletions; however, due to sequence assembly by aligning overlapping fragments with a reference genome, the detection of larger deletions and insertions, CNVs, and tandem repeat disorders is limited by the achievable read length in standard WGS[128] and the recognition of structural variants if their breakpoints are located in a repeat region[128,129]. Modern bioinformatics analysis methods try to overcome these difficulties, with some success especially in modern PCR-free WGS datasets[130], but require high coverage and resolution[127] and are not yet reliable enough to be employed on a larger scale. For clinical use, fragment analysis, repeat prime PCR, and Southern blotting are still needed to confirm expansions and their size.

Fragment analysis, while labour intensive, allows the screening of samples for nucleotide repeat expansions, which can cause neurodegenerative conditions such as

Huntington's disease and frontotemporal dementia/amyotrophic lateral sclerosis (*HTT* and *C9orf72*, respectively), as well as many of the spinocerebellar ataxias. Repeat-primed PCR producing DNA fragments of the repeat with the reverse primer binding at different points within the repeat permits the generation of electropherograms of the region[131]. Samples with mismatched alleles can then be further examined for size using Southern blotting[111]. Fragment analysis[131] and southern blotting[111] have become the commonest methods for detecting these, but may soon be superseded by long-read sequencing platforms (such as those offered by Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT), which can achieve several thousand base pairs and can detect larger deletions and insertions, CNVs, and tandem repeat disorders[128] but cannot a yet match the accuracy and reliability of more established platforms for SNP detection; PacBio also remains prohibitively expensive.

3 Challenges and questions in genetic dementia

D. Variant identification and classification

a. Classification criteria and variants of uncertain significance

With the more widespread use of next-generation sequencing technology the number of variants identified in patients has increased massively. All these variants need to be classified based on the evidence that they cause disease; the most thorough, and internationally recognized are those recently published by the American College of Medical Genetics and Genomics (ACMG) and Association for Molecular Pathology (AMP)[132]. These ACMG-AMP guidelines aim to use all available evidence to classify variants into one of five categories- benign, likely benign, variant of uncertain significance (VUS), likely pathogenic or pathogenic. Variants that are classified as likely pathogenic or pathogenic can be fed back to the patient and used to offer diagnostic, predictive or prenatal testing to other family members. However, many variants end up classified as VUS, either because there is insufficient or conflicting evidence. The criteria are designed to ensure that unless there is compelling evidence for a variant being pathogenic or benign, it will remain a VUS, because the risks of misclassifying a variant are considerable. However, the criteria are not applied in a universal way[133,134]; in addition, they also take into account variant information published prior to the

development of population databases such as the Exome Aggregation Database (*ExAC*) and the Genome Aggregation Database (*GnomAD*)[123], which allow the determination of allele frequency in many populations.

b. Secondary findings

With genetic testing less and less focused on genes known to be associated with the syndrome, the rate of secondary findings of clinical significance is increasing. Secondary findings, or incidental findings, are mutations in genes that are unrelated to the condition the patient is being tested for and can be relatively frequent[135,136]. Patients should be specifically consented for the reporting of secondary findings but are generally in favour of disclosure[137-139]. Limiting the analysis to relevant genes only can help avoid this problem but the ACMG currently recommends that in the case of clinical WES or WGS, some secondary findings for which screening or early treatment might improve patient outcomes should be reported[140]; these may increase as available treatment options increase.

E. Ethical considerations

In patients living with dementia, obtaining informed consent for genetic testing may be particularly challenging when the patient has impaired cognition. The decision to test an individual who lacks capacity should be based on their best interests and should ideally aim to ascertain what their wishes from when they had capacity. Other factors to consider include whether the patient has the capacity to understand the results, and how these may affect the patient, i.e. whether receiving results might exacerbate the patient's condition, or whether they would modify the management, treatment or monitoring of the patient. Since genetic testing clarifies the risks to other family members and predictive testing for relatives, this may be a motivation for some patients to undergo testing and as such, confidentiality is rarely an issue. On occasion however, a patient may decline to tell their relatives of a genetic diagnosis; a recent UK judgement states that clinicians need to weigh right of the patient to confidentiality against the right of an interested third party to be informed of results that impact upon them[141.142]. Up to 20% of relatives are not informed of relevant genetic results,

either because the patient wanted to protect relatives or was unaware that the information was relevant[143,144].

A genetic diagnosis offers many benefits, such as diagnostic certainty or access to support groups, but also carries risks, such as psychosocial problems, and problems with job and insurance prospects. The decision for or against a predictive test can therefore be fraught, and uptake for predictive testing for Huntington's Disease, for example, hovers in the 5-20% range for at-risk relatives, and 25% for those at-risk of prion disease[145]. This may change as more treatment options become available.

F. Project Hypotheses

Based on clinical observations and evidence of genetic pleiotropy, such as the discovery of genetic links between very different neurodegenerative syndromes[13,146], dementia syndromes are increasingly understood to feature a larger overlap between clinical syndromes than previously though (clinical heterogeneity), while diverse clinical phenotypes may be caused by mutations in the same gene (pleiotropy). The research project underlying this thesis aimed to help solve some of the challenges facing genetic testing services and clinicians treating patients living with dementia. Starting from analysing a large dataset of samples from diverse dementia patients using a targeted dementia gene panel, I subsequently zoomed in on Huntington's disease phenocopy syndromes in particular.

Examining over 3000 samples patient and control samples using a validated NGS panel for dementia[13] supplemented with assays of the *C9orf72* and *PRNP* expansion mutations, in the first part of this project, I sought to establish statistically meaningful prevalences of genetically determined dementias. In order to help the counselling of patients, I sought to provide figures of likely genetic discovery rates based on the clinical syndrome and establish the validity of easily determined clinical predictors of genetic disease in stratifying genetic risk I analysed rates of concurrent mutations to assess the impact of epistasis, as well as rates of novel variants and deleterious mutations in genes not typically associated with the clinical diagnosis in order to examine to what extent next-generation sequencing (NGS) increases the yield of

traditional Sanger sequencing of a few hand-picked genes and how frequently a dementia syndrome may be caused by mutations in genes unexpected for the phenotype.

The size of the dataset subsequently allowed a number of statistical calculations that would not have been possible with fewer samples. Using a recently published methodology[28], it facilitated the assessment of variants with regards to their likely penetrance based on their presence and frequency in the population via online population databases such as *GnomAD*[123,124]*;* it also allowed me to determine a threshold for what may constitute a rare variant with likely Mendelian effects in EOAD and EOFTD. In order to effectively counsel patients before genetic testing, information on how likely they are to receive a genetic diagnosis, and whether variants are fully penetrant and likely to manifest in family members, is essential.

Subsequently, and inspired by Shi *et al*., who described the nefarious effects of apolipoprotein 4 (ApoE4) in a tau mouse model and in a cohort of AD patients[147], I used the dataset acquired for the MRC Dementia Gene Panel study, to explore the effects of ApoE4 on FTD patients with proven or suspected tau pathology. While ApoE4 has long been known to be the strongest known risk factor for AD[71,148,149], it is not routinely clinically tested for.

Taking a closer look at the HD phenocopy (HDPC) syndromes, which are an excellent example of unclear neurodegenerative disease, I first attempted to improve the clinical definition of HDPC syndromes based on adult patients who presented to two clinics for Neurogenetics at the National Hospital for Neurology and Neurosurgery (NHNN) and Great Ormond Street Hospital and had gone on to have a genetic test for HD. Presently, these patients are defined by what they do NOT have – a positive genetic test for HD – and I therefore proceeded to explore whether I would be able to establish other causal mutations in a subset of 50 HDPC patients, that may help elucidate genetic links between these patients who present with a syndrome similar or identical to HD. Finding different genes that converge on the same pathway and cause

similar symptoms would further our understanding of the underlying processes and could help guide future biological *in vitro* and *in vivo* studies.

In summary, this thesis aimed to address the following questions using the MRC Dementia Gene Panel study:

- How are clinical syndromes and genes with deleterious variants (DVs) linked, i.e. is the clinical presentation a reliable predictor of the genes involved and what are the likely rates of genetic discovery in a given clinical syndrome? How and on what basis should variants be classified? Depending on the results, this may indicate that there remains work to do to generate data that would allow more definitive classification of variants to allow meaningful feedback to patients. In addition, if DVs in genes not typically associated with the clinical syndrome were to contribute substantially to the disease burden this would imply that the single gene "cherry picking" approach is problematic.
- How common are novel mutations overall and in the cohorts? A substantial proportion would imply that there remains substantial causal variation in known genes to be discovered and described.
- How frequent are concurrent mutations in the cohorts and overall? If epistasis were to exceed random distribution in cases would mean that this possibility needs to be considered when doing genetic testing as if missed can lead to erroneous predictive genetic testing in families, and interesting biological implications.
- Which patient groups are likely to have the highest genetic discovery rates? How useful are clinical predictors of genetic disease such as AAO and family history and are the results of this study transferable to a clinical context?
- Can all relevant mutations, both previously reported and those discovered in this dataset be considered fully penetrant when the statistical power of the present dataset and online population databases is harnessed for dedicated calculations?

To address these questions, statistical analyses were carried out with a pre-defined statistical threshold of 0.01 to account for testing these five main hypotheses;

subsequent secondary tests such as post-hoc analyses were then carried out without further corrections for multiple statistical testing.

Subsequently, and inspired by Shi *et al*., I would examine the effect of ApoE4 on tau in FTD based on the available data from the MRC Dementia Gene panel study: would patients with proven or presumed tau pathology who carried an ApoE4 allele have an earlier disease onset than those with other ApoE alleles?

Based on the HDPC syndrome results from the MRC Dementia Gene Panel study, I would consequently turn my focus to these cryptic patients as an example of unclear neurodegenerative disease.

- As a first step, I would survey what HD clinician experts deemed typical of HD. What symptom(s) would make them order an HD test and what would make them think that a test may come back negative?
- Next, I would attempt to clarify the clinical definition of HD phenocopy syndromes. Would certain symptoms, or symptom combinations, may be more common in HD or HDPC syndromes helping to distinguish them clinically? Would these symptoms match up with what HD clinician experts were expecting? This could help improve prediction of whose HD test will be positive or negative based on clinical symptoms or, on the contrary, show that HD and HDPC syndromes are not easily clinically distinguishable.

Analysing whole-genome sequencing (WGS) of 50 HDPC patients included in the MRC Dementia Gene Panel study, but without a genetic diagnosis, I would explore these patients genetically:

- Could other known causes of HDPC syndromes be identified screening the whole genome?
- Could variants found in two or more cases be used to link new genes and pathways likely to be involved in causing HD-like symptoms? This would potentially offer a genetic explanation for a larger proportion of HD lookalikes than before.

- Could functional connections be made between the set of genes responsible for the syndrome and how these pathways may be linked to HD-like symptoms?

II. Materials and Methods

1 Cohort stratification scores

A. The Goldman Score

Family history has traditionally been viewed as valuable when assessing a patient for a genetic diagnosis. To quantify the strength of a patient's family history and assess the associated risk of carrying a deleterious variant, each case was assigned a modified Goldman score (GS)[78,150] (see Figure II-1). A relevant family history included first or second degree relatives with dementia or symptoms compatible with the syndrome of the index case, such as a history of personality change in a deceased parent in a patient with fronto-temporal dementia, or chorea in the relative of a patient with dementia with motor symptoms. A GS of 1 requires a patient to present with an autosomal history with at least three affected family members over two generations linked by a first degree relative of the other two. A GS of 2 would be assigned to a family with three affected family members but not fulfilling the criteria for GS1. One other affected family member would be noted as a GS3 if the relative was younger than 65 years of age at symptom onset, or a GS of 3.5 if onset was later than 65 years. Cases with a known blank family history were catalogued as GS4, while cases with a censored, incomplete or unknown family history were categorized as GS4.5.

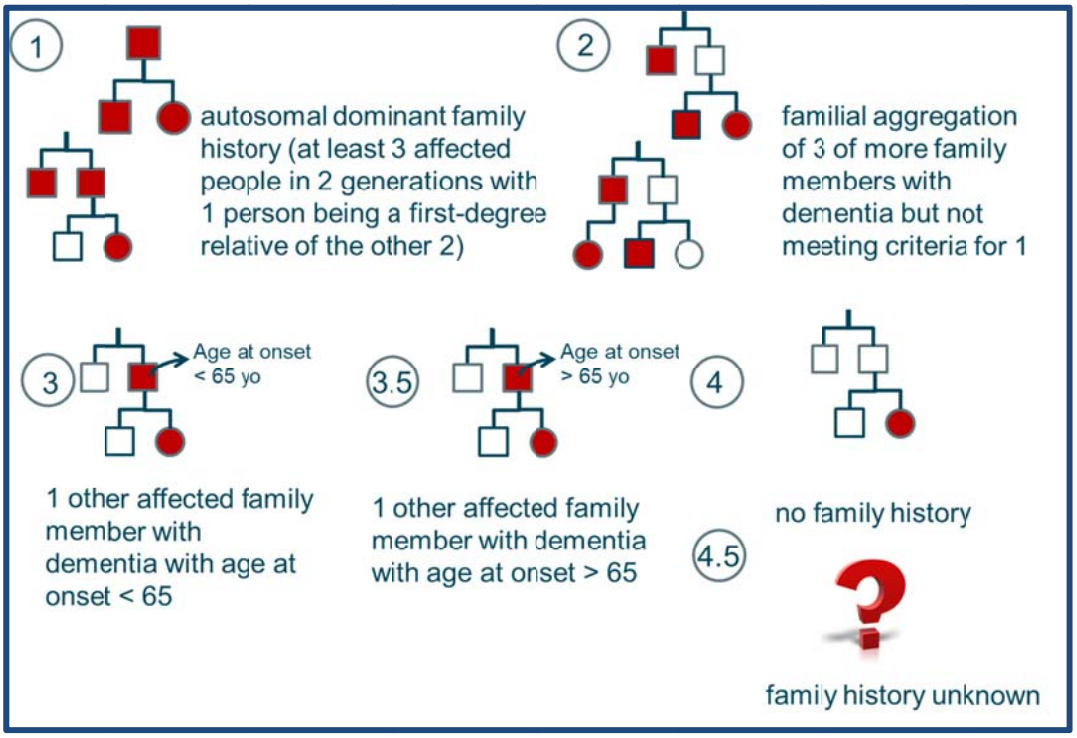

Figure II-1: The Goldman score as a measure of the strength of family history.

A GS of 1 corresponds to an autosomal dominant family history of symptoms consistent with the clinical syndrome, with at least 3 affected people in two generations linked by a first-degree relative. GS2 relates to the familial aggregation of three of more affected family members but not meeting criteria for 1; 3 denotes the presence of one other affected family member (3 if AAO is under 65, and 3.5 if onset over 65); 4 signifies no family history and 4.5 indicates a censored family history or no available information.

B. The HD Phenocopy Score

HD phenocopy patients, as described in detail in Results Chapter 3, are defined using an HD Phenocopy (HDPC) Score which ranks patients by the degree to which they are similar to the typical clinical presentation of HD. One point each is awarded for the presence of motor disorders, cognitive decline, and psychiatric features.

2 Sample DNA-extraction and control

A. Blood

First, for each sample, the sample ID was checked against the sample register and one 50ml Nunc tube and two 15ml Nunc tubes clearly labelled with the sample ID. Using the Nucleon^{TM} BACC3 Kit and following the manufacturer's instructions, 10ml Reagent A and 30ml nuclease-free, de-ionised water were added to the 50ml Nunc tube for

each sample. 3-4ml of patient blood were added to the tube and mixed with the diluted Reagent A for 4 minutes and the mixture then centrifuged at 1300G for 5 minutes at room temperature. Discarding the supernatant, but safeguarding the pellet, 2ml of Reagent B was then added to the pellet, shaken and transferred to a new 15ml Nunc tube. 500µl of sodium perchlorate was added to the Reagent B/pellet mixture and shaken. 2ml chloroform were added to the tube and combined with the mixture by shaking, then 300ul of Resin beads (Gilson) were added to the tube without shaking and the mix then centrifuged for 3mins at 1300rpm at room temperature. The top phase was removed to a fresh 15ml Nunc tube and the lower phase discarded. 2ml ethanol was added to the supernatant in the new 15ml Nunc tube and shaken to combine, causing the DNA to fall out and become visible. With a fire-sealed, sterilized glass pipette, the DNA was spooled out of the liquid, air-dried and added to 500ul TE buffer in a DNA tube. The extracted DNA was left to rehydrate in Tris-EDTA (TE) buffer and waste was disposed of in a bin filled with sodium hydroxide. I personally performed this technique for approx. 12 samples.

B. Saliva

Saliva was extracted using the prepIT-L2P kit (5ml, DNA Genotek, Ottawa, Canada). First, for each sample, the sample ID was checked against the sample register and two 15ml Nunc tubes, one DNA tube and one glass rod clearly labelled with the sample ID. Sometime prior to extraction, the saliva samples were incubated at 65°C in the oven for one to two hours. The saliva samples were then transferred to a 15ml Nunc tube, 40µl PT-L2P per millilitre of saliva were added to the tube, which was vortexed briefly to mix. After letting the tube rest for 10 minutes on ice or at room temperature, it was centrifuged at 11,400 x *g* for 10 minutes, then the supernatant was carefully removed to a new 15ml Nunc tube and the pellet discarded. After adding 2ml of 100% ethanol and inverting several times, the tube was left to stand for 10 minutes at room temperature. If the DNA pellet was visible at that point, it was spooled out of the liquid, air-dried and added to 500ul TE buffer in a DNA tube with a fire-sealed, sterilized glass pipette. If the DNA pellet wasn't visible after adding ethanol, the tube was centrifuged at 11,400 x *g* for 10 minutes, then the supernatant was removed and discarded and 1 millilitre 70% ethanol added without disturbing the pellet and the tube

left to stand for one minute. After gently swirling the mixture around in the tube, the ethanol was removed and the DNA rehydrated in TE and briefly vortexed. Finally, the rehydrated DNA was transferred to a DNA tube. I personally performed this technique for approx. 4 samples.

C. DNA quality control on the Agilent 4200 TapeStation System ®

For a small subset of samples, it was necessary to assess the concentration and integrity of the DNA used for next-generation sequencing. This was achieved using the 4200 TapeStation system® and genomic DNA screen tape analysis (all Agilent Technologies, Santa Clara, California, USA), which permits sizing of DNA fragments from 200 to >60,000bp.

After letting the Genomic DNA Reagents come to room temperature for 30 minutes, the TapeStation was switched on, the Controller software initiated and the DNA ScreenTape and loading tips placed in to the 4200 TapeStation instrument. After gently vortexing the reagents and spinning them down, 10µl of Genomic DNA Sample Buffer was pipetted into the wells of a tube strip; 1µl of Genomic DNA Ladder was pipetted into position A1 and the test samples pipetted into their respective wells on the strip (or 96-well plate). The tube strip was then sealed with caps (or the 96-well plate with a foil seal), the reagents and DNA mixed using the IKA vortex at 2000rpm for 1min, and the samples and reagents spun down. After loading the samples into the 4200 TapeStation system® with the ladder in position A1, the caps were removed from the tube strips (the Agilent foil seal for 96-well plates can remain), the samples were then allocated a specific location on the testing strip (or 96-well plate) and the run initiated by clicking start. After the run, the Agilent Tapestation Analysis Software displayed the results automatically. I personally performed this technique for approx. 20 samples.

3 Gene-panel sequencing of dementia syndromes

A. Selection of samples for gene-panel sequencing

a. Patient samples for gene-panel sequencing

The 2784 patient samples comprised 1052 Alzheimer's disease (AD) cases, 794 frontotemporal dementia (FTD) cases, 299 samples collected from patients seen

through the National Prion Clinic, and 639 samples from patients with a dementia syndrome not consistent with other categories and associated with motor symptoms (DemMot). Diagnoses were based on the clinical assessment of the referring clinician, rather than pathological data, in order to reflect clinical reality.

2352 patient samples were chosen retrospectively from the MRC Prion Unit DNA sample collection. These samples were collected over a 20 year period as they were sent in for single gene testing for *PRNP*, *PSEN1*, *PSEN2*, *APP*, *GRN*, and *C9orf72* by clinicians working mostly at the National Hospital for Neurology and Neurosurgery as well as other London and southeast England hospitals. These also included 120 FTD samples from Cambridge sent in for earlier genetic testing in the study.

Selection of samples was based on the availability of research consent and of the clinical data to be used in the predictive modelling and was blind to any mutations detected in earlier tests. One of the cases with a previously detected deleterious mutation was also included in the chip sequencing (and the mutation confirmed); for 47 cases later chosen for chip sequencing, an increased risk for AD due to their ApoE genotype was already known. In addition to these retrospectively analysed samples, 432 patient samples were referred prospectively for the study; 165 patients were recruited prospectively from Memory clinics in Bristol, Cardiff, Southampton and UCLH for gene-panel testing research, and 267 patients were referred to the Division of Neurogenetics at the National Hospital for Neurology and Neurosurgery (NHNN) for clinical gene-panel testing, these comprised samples from all cohorts except prion patients and were also included in this analysis. The study was approved by the local research ethics committee.

As only high order clinical categories were used, any change in diagnostic criteria over the period of sample collection should not have affected the cohort classification. I examined available clinical data to personally select approx. 30-40% of all samples sequenced, others were selected based on quality and abundance of the DNA sample. AD samples were mostly chosen from the MRC Prion Unit DNA sample collection; FTD

patients comprised the UCL and Cambridge University FTD Cohorts. Prion disease patients were referred to the National Prion Clinic for *PRNP* gene testing based on a suspicion of inherited prion disease. DemMot was a category we defined to assemble a variety of clinical syndromes that comprise a cognitive disorder and pyramidal or extrapyramidal features, not fitting any of the other diagnostic categories, in order to explore the usefulness of a dementia panel in cryptic movement disorder cases, which feature dementia as part of a complex syndrome. Most of the patients included in the DemMot cohort were cryptic cases without a formal clinical diagnosis at the time of referral. A majority had previously been tested for Huntington's disease (which was negative in all cases), which typically reflected a combination of cognitive symptoms and either a hyper- or hypokinetic movement disorder, often in conjunction with psychiatric and/or frontal symptoms, and a positive or censored family history. 31% of patients presented with chorea, in over 10% of patients CBS was considered, 9% presented with an atypical parkinsonian syndrome, and 3% were identified as consistent with a PSP phenotype. In addition, three patients with dementia with Lewy bodies, and three cases with clinical features of motor neuron disease were included in the cohort. The clinical diagnosis was confirmed in 91.0% of the 78 cases from AD and FTD cohorts where cerebrospinal fluid (CSF) results were available. Neuropathological data was available for 122 patients (4.5% of all patients) and confirmed the clinical diagnosis in 102 (83.6%) of cases. Only one case with an unexpected mutation went to post-mortem examination (from our AD cohort with an unexpected *C9orf72* expansion); neuropathological data confirmed the expected TDP pathology.

Inevitably at one site, certain mutations will be overrepresented because of the tendency for particular variants to cluster in ethno-geographical groups or extended pedigrees. In order to exclude the possibility that the inclusion of known relatives with the same mutation might confer bias and influence our main findings, all family members were identified, based on this definition: that the second case from a family reported the diagnosis of the proband in their clinical family history. On this basis I identified 73 family cases. Since variants identified in these individuals did not lead to

bias of the overall findings as confirmed by statistical re-analysis, I included these cases in the present results.

b. Control samples for gene-panel sequencing

457 control samples were obtained from relatives of patients attending the memory clinic at Cardiff and Vale NHS hospital. An additional 10 controls were samples processed in the MRC Prion Unit Human Genetics group as predictive testing in families with a known genetic mutation causing disease, but who were not mutation carriers. The local ethics committees approved the inclusion of these samples in the study.

c. Clinical predictors

Age at onset (AAO), sex, site of sample origin and family history were documented from the clinical notes and referral cards and used to assess their validity as clinical predictors. The strength of a patient's family history was stratified using a modified Goldman score (GS) [78,151] (see "Cohort classification scores – The Goldman Score"). Information on ethnicity was assumed to be incomplete and supplemented by comparing genotypes at 133 sequenced SNPs from our study participants to those from individuals from British and continental outgroup populations genotyped by the 1000 Genomes study[152]. While the number of SNPs is insufficient for the inference of ancestry, it facilitated the identification of population specific clusters using principal components analysis (PCA) implemented with PLINK[153]. 105 study participants (3.2%) who were outliers from the British cluster could thereby be inferred to be of non-white British ancestry, carrying 14 out of 354 deleterious variants. Since variants identified in these individuals did not lead to bias of the overall findings, these cases were included in the results and analyses.

B. Next-generation gene panel sequencing

a. The IonTorrent® PGM platform

Next-generation sequencing allows for cost-effective, high-throughput, fast and accurate amplification, sequencing and analysis of DNA samples compared to the older Sanger sequencing. 2974 samples (2517 patients, 457 controls) were run on an IonTorrent PGM sequencer (Thermo Fisher Scientific) using a previously validated

dementia gene panel[113]. Custom primers designate the DNA of interest, which were amplified in a first PCR step. The generated amplicons were partially digested and barcoded to identify each sample individually. For quality control and according to the protocol, target amplification was assessed via qPCR and the DNA concentration equalized; barcoded samples were then pooled into a library before being linked to beads. After a further, emulsion-based PCR amplification step and enrichment of the templated beads, the enriched template-positive ion-sphere particles were measured on a Qubit® 2.0 Fluorometer. The amplified, enriched sample library was then loaded onto a sequencing chip and placed on the IonTorrent® PGM machine, which in turn flooded the chip with the four different nucleotides. The IonTorrent® technology relies on the fact that, during sequencing, every time a nucleotide is incorporated into the DNA-strand being sequenced, the DNA polymerase releases an H+ ion which is measured directly. After each run, chip loading and the number of aligned reads were evaluated. I personally performed sample library preparations, quality control steps and sequencing runs for approx. 25% of the samples included.

i. The MRC dementia gene panel

The MRC dementia gene panel has been developed to cover the commonest genes linked to neurodegeneration in previous studies. It comprised the open reading frame and intron/exon boundaries of 17 dementia genes – *APP, CHMP2B, CSF1R, FUS, GRN, ITM2B, MAPT, NOTCH3, PRNP, PSEN1, PSEN2, SERPINI1, SQSTM1, TARDBP, TREM2, TYROBP* and *VCP* –on the IonTorrent® PGM sequencer. The panel and method have been previously examined and validated with excellent results[113], but necessarily excluded other genes such as *TBK1*, which were discovered later[154]. The development and validation of the panel pre-date my involvement in the project. *APP, PSEN1* and *PSEN2* are typically associated with Alzheimer's disease as Mendelian causes, and *TREM2* variants as risk factors (when identified in the heterozygous state, or Nasu-Hakola disease in the homozygous state), while *GRN* and *MAPT* are known causes of fronto-temporal dementia*,* with *FUS, SQSTM1, TARDBP,* and *VCP* more commonly linked to amyotrophic lateral sclerosis but with some overlap on the spectrum with fronto-temporal dementia. *PRNP* is linked to rapidly progressive dementia, both in its

familial forms and sporadic cases where the misfolded prion protein can be found. Other genes like *CHMP2B, CSF1R, ITM2B, NOTCH3, SERPINI1,* and *TYROBP* are associated with much rarer neurodegenerative syndromes.

Because of technical limitations inherent in the methodology, next-generation sequencing does not lend itself easily to the testing for repeat disorders. For this reason, the MRC dementia gene panel was supplemented by testing for one of the more common causes of dementia, *C9orf72* expansions[131], as well as *PRNP* octapeptide insertions (OPRI)[155] separately (as explained below), and *APOE* genotype by minor groove binding probe.

ii. Primer Design for IonTorrent ® sequencing

Covering 17 genes including exon flanking regions and some untranslated sequence, 21581kB of reference sequence were submitted to Ion AmpliSeq™ Designer to design custom primers using the 200bp length option. 214 amplicons divided into 2 pools of similar numbers were returned under the design name IAD36795. A detailed list of the included genes and genome sections can be found in Table V-1**.**

iii. Library Preparation, Template Preparation and Sequencing

Following the manufacturer's instructions (Ion AmpliSeq™ DNA and RNA Library Preparation) and using the Ion AmpliSeq™ Library Kit 2.0, genomic DNA was first amplified with both IAD36795 primer pools in a 96 well plate, using DNA straight from stock without quantification or dilution and half-size reactions as detailed in the user manual; 18 amplification cycles were used with a 4 minute extension time. After amplification, the libraries based on the two IAD36795 primer pools were pooled and the primer sequences partially digested using FuPa reagent; Ion Xpress™ barcode adapters were then ligated to the amplicons using DNA ligase. Following a cleaning step using Agencourt® AMPure® XP Reagent, the pooled libraries were then diluted 1:100 on a separate 96 well plate and their concentration determined via PCR using the Kapa Ion Torrent Library Quantification kit (ROX Low qPCR Mix, Kapa Biosystems, Wilmington, MA 01887, USA). Based on the sample concentration measured in the PCR step, the libraries were diluted and equalized to a 30pM concentration. Equal amounts

of library, typically 5µl, were then pooled to make up an AmpliSeq library. Typically, 32 samples were processed at a time and pooled in a sample library to be run on a 318 chip.

Following the manufacturer's instructions (Ion PGM Template OT2 200 Kit User Guide) and using the Ion PGM™ Template OT2 200 Kit and Ion OneTouch™ 2 System, the combined AmpliSeq library consisting of typically 32 30pM pooled libraries was added to the OT2 200 amplification solution and processed on the Ion OneTouch™ 2 System. In this step, PCR amplification of amplicons linked to Ion PGM™ Template OT2 200 Ion Sphere™ Particles (ISPs) occurred in emulsion, the ISPs were collected at the end. In the Ion Sphere™ Quality Control assay, ISPs were then assessed to determine the percentage of ISPs templated with library, which should be between 10-30%. Subsequently, ISPs were enriched for templated ISPs using the Ion OneTouch™ ES; during this process, ISPs were bound to Dynabeads® MyOne™ Streptavidin C1 Beads, washed of untemplated ISPs and then eluted from the beads.

Following the manufacturer's user guide (Ion PGM™ Sequencing 200 Kit User guide) and using the Ion PGM™ Sequencing 200 Kit v2 and IonTorrent Personal Genome Machine (PGM), the first PGM needed to be cleaned and initialized, a process during which wash buffer was prepared and pH-adjusted; dNTP solutions were then prepared and reagents checked. While the template-positive ISPs were annealed to the sequencing primer, the sequencing chip was checked for faults. After Ion PGM™ Sequencing 200 v2 Polymerase was bound to the sample, it was loaded into the chip and the chip placed on the PGM to start the sequencing run.

iv. Analysis using NextGENe® and GeneticistAssistant®

The run data was removed from the Ion Torrent PGM server and aligned to the preloaded human hg37 build (Human_v37_3_dbsnp135_dna) in NextGENe (© Softgenetics). Coverage curve data from NextGENe was used to determine whether at least 95% of a sample had been sequenced a minimum of 10 times; if this was not the case, the sample failed quality control and was not analysed further. Samples had an average of 157,350 mapped sequencing reads, 90% of which were on target; average

mean depth of coverage was 676 with a uniformity of 94.7%. On average, 99.5% of the target sequence was covered at least tenfold. BAM and VCF files were then imported into GeneticistAssistant (©Softgenetics) for further analysis. The GeneticistAssistant output included information on *in silico* predictions, the number of times a variant had been observed in the dataset and whether it had been analysed before. In GeneticistAssistant, coverage of the variant (>10x), zygosity and variant (allele) frequency were assessed for each variant, with allele frequency expected to lie between 0.25 and 0.75 for heterozygous variants, and between 0.8 and 1 for homozygous variants. In uncertain cases, forward and backward reads (similar reads in both directions) were evaluated in NextGENe and novel DVs were confirmed by Sanger-sequencing. I personally performed the quality control tests for approximately 40% of the samples, and analysed and classified all variants in all of the samples.

b. Additional Tests

i. Fragment analysis for repeat expansions

C9orf72 expansion

The test for the *C9orf72* expansion is PCR-based. Primers, one of which is fluorescently tagged, bind outside and within the repeat region which is amplified[156] and the product then run out on the 3730XL Sequencer. A master mix of 1µl Dimethylsulfoxide (DMSO, Sigma-Aldrich, St Louis, MO, USA), 2.5µl Q solution (or 5M betaine, Sigma-Aldrich, St Louis, MO, USA)), 0.25µl deaza GTP (10mM), 0.25µl water, 0.5µl MgCl2 (25mM, Quiagen, Hilden, Germany), 7µl Fast start PCR mix (Roche Molecular Systems, Indianapolis, IN, United States), 0.4µl Primer F (100uM, 6-FAM-AGTCGCTAGAGGCGAAAGC, Eurofins Genomics, Ebersberg, Germany), 0.2µl Primer R (100uM, TACGCATCCCAGTTTGAGACGGGGGCCGGGGCCGGGGCCGGGG, Eurofins Genomics, Ebersberg, Germany) and 12.5µl Primer Tail (100uM, TACGCATCCCAGTTTGAGACG, Eurofins Genomics, Ebersberg, Germany) was made up per well and aliquoted into a 96-well plate (Thermo Fisher Scientific, Waltham, MA, USA). After adding 1.5µl of genomic DNA, a repeat prime PCR was performed on a Veriti® Thermal Cycler (Applied Biosystems, Thermo Fisher Scientific, Foster City, CA, United States) with conditions set out in Table II-1Table II-1. The fragment analysis

master mix was prepared with 10µl HiDi formamide (Life Technologies Invitrogen, Carlsbad, CA, United States) and 0.5µl GeneScan LIZ 500 (ladder, Life Technologies Invitrogen, Carlsbad, CA, United States) per sample. After combining 10µl of fragment analysis master mix and 2µl of PCR product in a non-skirted 96-well plate, the mix was covered, spun down and denatured at 95˚C for 2 minutes, then transferred to a pre-cooled block at 4˚C and incubated for 2 minutes. The plate was then run on the 3739XL sequencer as a fragment analysis with a 50cm capillary array length and GeneMapper50® Pop7 settings for G5RCT modified to a slower injection time of 45 seconds. The data was analysed using the PeakScanner software (©LifeTechnologies) with GS500LIZ as the size standard and sizing default –pp as the analysis method. Each sample's amplicon peaks were then assessed individually. I personally performed this technique for approx. 30 samples.

Table II-1: Cycling conditions for the *C9orf72* repeat prime PCR

Temperature ˚C	Time (min)
95	15 min
94	1 min
70	1min
72	3min, 2 cycles total
94	1min
68	1min
72	3min, 3 cycles total
94	1min
66	1min
72	3m, 4 cycles total
94	1 min
64	1 min
72	3 min, 6 cycles total
94	1 min
60	1 min
72	3 min, 7 cycles total
94	1 min
58	1 min
72	3 min, 8 cycles total
94	1 min
56	1 min
72	3 min, 5 cycles total
72	10 min
10	hold

PRNP octapeptide repeat insertions

The test for *PRNP* octapeptide repeat insertions or deletions is PCR-based. Primers, one of which is fluorescently tagged, bind around the octapeptide repeat region which is amplified and the product then run out on the 3730XL Sequencer. Two positive controls, one for the 1-octapeptide repeat deletion (1OPRD) polymorphism and one for a 6-octapeptide repeat insertion (6OPRI) were included in each run, as well a negative control without added DNA.

A premix of 12.5µl MegaMix-Royal (MMR, MegaMix-Royal - Hot Start PCR Master Mix, Microzone, Haywards Heath, UK), 0.125 µl Primer 5-FAM (100uM, GACCTGGGCCTGTGCAAGAAGCGC, Eurofins Genomics, Ebersberg, Germany), 0.125 µl Primer 6 (100uM, GGCACTTCCCAGCATGTAGCCG, Eurofins Genomics, Ebersberg, Germany), 11.25 µl water was prepared per well, aliquoted and combined with 1 µl DNA in each well in a 96-well plate. The plate was then covered, spun down and run on a Veriti® Thermal Cycler (Applied Biosystems, Thermo Fisher Scientific, Foster City, CA, United States) for 34 cycles using the parameters set out in Table II-2.

Table II-2: Cycling conditions for the *PRNP* repeat prime PCR

Temperature ˚C	Time (min)
95	5 min (MMR), 3min (MMD)
95	30sec
65	40sec
72	45sec

The fragment analysis master mix was prepared with 10µl Hi-Di formamide (Life Technologies Invitrogen, Carlsbad, CA, United States) and 0.5µl GeneScan LIZ 500 (ladder, Life Technologies Invitrogen, Carlsbad, CA, United States) per sample. After combining 10µl of fragment analysis master mix and 2µl of PCR product in a non-skirted 96-well plate, the mix was covered, spun down and denatured at 95˚C for 2 minutes, then transferred to a pre-cooled block at 4˚C and incubated for 2 minutes. The plate was then run on the 3739XL sequencer as a fragment analysis with a 50cm capillary array length and GeneMapper50® Pop7 settings for G5RCT and the data

analysed using the PeakScanner software (© LifeTechnologies) with GS500LIZ as the size standard and sizing default –pp as the analysis method. Each sample's amplicon peaks were then assessed individually. I personally performed this technique for approx. 20 samples.

ii. ApoE testing

Probe 34 (rs429358, Life Technologies Invitrogen, Carlsbad, CA, United States) and probe 36 (rs7412, Life Technologies Invitrogen, Carlsbad, CA, United States) were defrosted and diluted 1:2 from 40x to 20x in deionised water; two master PCR mixes was then prepared using 2.5ul Taqman GTXpress Master Mix (Applied Biosystems), 0.25ul primer (20x) and 1.25ul water per sample and for each of the two primers. 4ul/sample of each PCR master mix was aliquoted into a 96-well plate and mixed with 1ul genomic DNA; positive controls were included for each probe. After sealing with a thin plate seal, vortexing and a spinning down the plate at 30000rpm for 1 minute, the plate was then loaded onto the Applied Biosystems QuantStudio 12K Flex using the Quantstudio® software (genotyping experiment, Taqman reagents and Fast run were selected) and run 34 cycles using the parameters set out in Table II-3Table II-3. Data was analysed using the Quantstudio® software. I personally performed this technique for approx. 50 samples.

Table II-3: Cycling conditions for ApoE genotyping

Temperature ˚C	Time (min)
60	30sec
95	20sec
95	1sec, 40 cycles
60	20sec
60	30sec

c. Sanger sequencing - confirmatory testing

Novel variants and variants detected under difficult sequencing conditions, such as close to homopolymer runs, were Sanger-sequenced to confirm their validity.

iii. Primer design

In the first instance, primers for confirmatory Sanger sequencing were designed using PrimerZ[157] querying the human genome by promoter regions and exons (NCBI) via the input of gene names. SNPs were masked, the promoter region set to 0, and the maximum exon length set to 500bp. The exon flanking sequencing length was defined as 250bp and product size was allowed to range from 300 to 700bp. 30bp from each end of the exon were defined as the excluded regions to avoid overlap with the exon or its splicing region. Optimum prime size was defined as 20bp (min. 18bp, max. 25bp), optimum primer melting temperature (Tm) was defined as 60˚C (min. 55 ˚C, max. 65 ˚C), and the optimum percentage of guanine and cytosine in each primer was defined as 50% (min. 40%, max. 60%). The list of generated primers for each gene was then checked for the *in-silico* PCR result and only primers that passed this test were used. If PrimerZ was unable to suggest suitable primers due to the location of the variant in question, for example in the case of intronic variants or variants in untranslated regions (UTR), primers were designed using Primer3[158]. In this case, the genomic sequence (GRCh37 Homo sapiens assembly) was first downloaded from Ensembl[159], the variant location identified and an exclusion zone of 25bp defined via square brackets on either side of the variant; the sequence was then archived for later referral. The sequence surrounding the variant in question was then inputted into Primer3 using the same conditions as detailed above for PrimerZ. The suggested primers were then pasted into the *in silico* PCR tool of the UCSC genome browser[160,161]; primers were only deemed acceptable if they defined a unique section of the genomic sequence. I personally designed the primers for most of the confirmatory Sanger sequencing reactions, approx. 18, while primers were already in use in the laboratory for another 2 genetic locations.

iv. Sanger Sequencing

To prepare the first PCR master mix, 12.5µl MegaMix Royal®, 1.25µl forward primer (10mM), 1.25µl reverse primer (10mM) and 9µl water were mixed per sample, aliquoted into a 96-well plate and combined with 1µl genomic DNA template per well. The samples were then amplified by PCR under the conditions set out in Table II-4;

subsequently, the PCR products were run on a gel to determine their size and to ensure a single band only became apparent.

Table II-4: Cycling conditions for 1st PCR step in Sanger Sequencing

Temperature ˚C	Time (min)
95	5 min
95	30 seconds
60 (decrease 1˚C/cycle)	30 seconds, 10 cycles
72	1min
95	30 seconds
50	30 seconds, 25 cycles
72	1min
72	10min, 1 cycle
10	hold

If the samples passed this quality control, 20µl of each PCR product was then cleaned by adding 20µl Microcleaner® (Microzone, Haywards Heath, UK) to each well, vortexing and incubating the mixture for 5 minutes at room temperature. The plates were then centrifuged at 3000G for 40 minutes at room temperature, and the supernatant subsequently removed by spinning the plates upside down on a tissue paper at 40G for 30 seconds. The pellets were then re-suspended in 160µl water by pipetting the mixture up and down and vortexing the mixture before incubating them for 5 minutes at room temperature, vortexing again and finally spinning the plates down briefly.

For the second PCR step, separate master mixes were prepared for the forward and reverse primers, combining 1 µl Big Dye® (Applied Biosystems, Thermo Fisher Scientific, Foster City, CA, United States), 5 µl Better Buffer (Applied Biosystems, Thermo Fisher Scientific, Foster City, CA, United States), 0.75 µl Primer (forward or reverse, 5mM) and 7.25 µl water per well; these were aliquoted into two separate wells per sample and each combined with 1µl PCR cleaned and diluted PCR product. A PCR was performed using the conditions set out in Table II-5Table II-5. Leftover product from this PCR step may be stored at -20˚C.

Table II-5: Cycling conditions for 2nd PCR step in Sanger Sequencing

Temperature ˚C	Time (min)
96	1 min, 1 cycle
96	10 seconds
50 (decrease 1˚C/cycle)	5 seconds, 25 cycles
60	4min
10	hold

Following the second PCR step, 3.75µl EDTA (0.125M) was added to each well, then 45µl 100% ethanol was added to each well and mixed my pipetting up and down. After incubating the mixture for 15 minutes at room temperature, the plates were centrifuged at 3000G for 30minutes at 4˚C and the supernatant subsequently spun off onto a tissue at 190G on pulse. 60µl of 70% ethanol was then added to each sample without mixing and the plates centrifuged at 1650G for 15minutes at 4˚C. After removing the supernatant again by spinning down the plates onto a tissue at 190G for 1 minute, the plates were left to incubate uncovered on a PCR block for 5 minutes at 37˚C. If necessary, plates could be stored at this point at -20˚C, otherwise 10µl Hi-Di formamide was added to each well; the plate was then vortexed, spun down and incubated at 95˚C for 2 minutes and then at 4˚C for 2 minutes. The plates were then run on the 3739XL sequencer using the BD1 FastSeq50 Pop7 protocol on a 50cm capillary array. The data analysed using the SequenceScanner software (© AppliedBiosystems). I personally performed this technique for approx. 5 samples.

C. The Illumina platform at NHNN Neurogenetics Laboratory

In addition to the samples run on the IonTorrent® platform, 267 patient samples referred for testing to the Neurogenetics Laboratory at the NHNN were included in the statistical analyses. They were sequenced on an Illumina MiSeq or HiSeq platform using the Neurogenetics Laboratory Dementia Panel, which included 17 genes: *APP, CHMP2B, CSF1R, DNMT1, FUS, GRN, HTRA1, ITM2B, MAPT, NOTCH3, PRNP, PSEN1, PSEN2, TARDBP, TREM2, TYROBP,* and *VCP*. All RefSeq transcripts of these genes were sequenced, covering all coding exons and 15 base pairs of flanking intronic sequences; for *MAPT*, this was extended to 15 base pairs of flanking intronic sequences to cover known intronic splicing mutations. Only variants in genes overlapping the MRC

Dementia Gene Panel were included in this study. The Nextera Rapid Capture Custom Enrichment Kit (Illumina) was used according to manufacturer's protocols for the library preparation and enrichment. Samples were 99% covered at 30x, with an average read depth of 500X consistently obtained. Sequencing regions with coverage lower than 10X were manually inspected. Deleterious variants in these clinically sequenced samples were confirmed by bi-directional Sanger sequencing. These patient samples were not tested for *APOE*, *C9orf72* or *PRNP* insertional mutations. I personally re-classified the variants identified in these samples based on the same criteria used to classify all other variants.

D. Previous exome sequencing

To evaluate whether the MRC Dementia Gene Panel covered most of the relevant genes, 715 patients (AD n=509, FTD n=83, DemMot=31, Prion=92, no controls) who were tested on the panel and had also previously been exome sequenced at Source Bioscience (Nottingham, UK) were assessed for any additional causal variants not covered by the panel. After Agilent-based exome capture (Agilent, Santa Clara, US), samples were paired-end sequenced on the HiSeq2000 sequencer (Illumina, San Diego, US); sequencing reads were aligned to GRCh37 using Novoalign. Genome Analysis Toolkit (GATK) was used for QC and variant calling, followed by annotation with ANNOVAR. Mean coverage across the cohorts was 64x, and 81.5% of targeted bases were covered >10x. Two deleterious variants were detected in genes not included in the panel (see results), but no other known pathogenic variants were returned by Ensembl's Variant Effect Predictor[162] in the exome-sequencing data.

E. Variant classification

All variants detected on the IonTorrent® chip were assessed for their likely pathogenicity. Variant classification was based on the ACMG-AMP guidelines[132], we introduced modifications and clarifications specific to our disease area and circumstances and removed criteria unsuited to our setting. To allow for a more detailed classification reflecting the application of our criteria to research data, we introduced a "Potentially deleterious" category, which, while not useful for patient feedback in a clinical setting, comprises variants that may warrant further research.

The algorithm used for classification is based on the level of evidence available for each variant as set out in Table II-6 which is combined for a final classification as explained in Table II-7. Intronic variants were assessed using Human Splicing Finder HSF V3.0 and classified according to our criteria[163]. Only variants with a population frequency <5% were manually classified. Variants detected at the Neurogenetics Laboratory at NHNN were also classified according to these criteria, but for these samples I did not report likely benign, benign or synonymous variants. I personally classified all variants identified reported in this study.

Table II-6: Evidence used to classify variants according to their pathogenicity level

Variants identified in a sample were classified according to the information available about them. This included the type of mutation in question, its position in the gene and/or protein, its frequency in online population databases, in silico predictions of effects on proteins, and whether it had previously been reported in families, single cases or controls.

Evidence level	Criteria
Pathogenic Strong	1) Coding amino-acid change previously published as deleterious with evidence of segregation in more than one pedigree or in multiple unrelated patients with the same phenotype 2) Null variant in a gene where loss of function (LOF) is a known disease mechanism (caveat LOF variants at extreme 3' end) 3) Variant in a gene associated with an expected very rare pathology (e.g. *PRNP* mutation and prion pathology) 4) Explained mechanism of pathophysiology of variant using *in vitro* or *in vivo* studies 5) Found in a mutational hotspot i.e. a domain where many other pathogenic mutations are seen, generally with additionally support from *in silico* prediction software
Pathogenic Moderate	1) Coding amino-acid change previously and justifiably published as deleterious but without evidence of segregation or in a single pedigree/patient 2) Novel missense change at an amino acid residue where a different pathogenic missense change has been seen 3) A very different amino-acid change at the same site or next to one with a less dramatic amino-acid change but deleterious 4) In a gene the mechanism of which is understood and the effect of the variant is in keeping with that mechanism; 5) Protein length changes as a result of in-frame deletions/insertions in a nonrepeat region or stop-loss variants 6) Mutation in a gene associated with a rare pathology in a case with a compatible clinical syndrome 7) Intronic variant affecting splicing or protein length

Evidence level	Criteria
Pathogenic Supporting	1) Variant with a major amino-acid change near or in a functional domain (e.g. active site of an enzyme) but not in a mutational hotspot 2) Multiple lines of computational evidence support a deleterious effect on the gene or gene product (conservation, evolutionary, splicing impact, etc.) Caveat: Because many in silico algorithms use the same or very similar input for their predictions, each algorithm should not be counted as an independent criterion 3) Reported in both cases and controls, but more cases than controls (statistically significant in a study)
Pathogenic Risk factor	1) Previously reported as risk factor, either variant itself or clear established pattern in gene 2) >1 in 10000 in *GnomAD*; 3) The prevalence of the variant in affected individuals is significantly increased compared with the prevalence in controls
Benign Independent	Allele frequency >5% on *GnomAD*, or 1000 genomes project
Benign Strong	1) Allele frequency >1% on *GnomAD*; 2) Reported benign in multiple pedigrees or with insight into gene/protein mechanism 3) Allele frequency is greater than expected for disorder 4) Lack of segregation in affected members of a family, caveat: phenocopies and penetrance 5) Seen in equal or greater frequencies in controls than cases
Benign Moderate	1) Allele frequency over 0.1% on *GnomAD* 2) Reported benign in one case or pedigree 3) Genetic mechanism inconsistent with pathological phenotype, or known mutation spectrum
Benign Supporting	1) Missense variant in a gene for which primarily truncating variants are known to cause disease or the mechanism is very specific and known 2) Multiple lines of computational evidence suggest no impact on gene or gene product (conservation, evolutionary, splicing impact, etc.) 3) A synonymous (silent) variant for which splicing prediction algorithms predict no impact to the splice consensus sequence

Table II-7: Criteria for variant classification

The evidence available about each variant was combined to determine its likely effect and likelihood of causing disease

Pathogenicity	Algorithm
Deleterious	Found in patient(s) and not controls OR in significant excess in patients AND seen on *GnomAD* at less than 1 in 50,000; AND Pathogenic Strong evidence 1) OR 2), PLUS one additional Pathogenic Strong or two Pathogenic moderate or one Pathogenic moderate and one Pathogenic Supporting criterion
Likely deleterious	The prevalence of the variant in affected individuals is significantly increased compared with the prevalence in controls, or only seen on *GnomAD* at less than 1 in 10,000; AND Pathogenic Moderate evidence 1) OR 2) OR 3) AND one additional Pathogenic Strong or Moderate or Supporting criteria.
Possibly deleterious	Found on *GnomAD* at less than 1 in 5000 and at least one Supporting criterion
Uncertain	Insufficient or conflicting evidence Missense mutation not nearby other missense mutations thought to be pathogenic
Likely benign	One Benign Strong criteria OR one Benign Moderate AND one Benign Supporting criteria OR two Benign Supporting criteria
Benign	Benign Independent OR one Benign Strong evidence criterion AND two further Benign Moderate or Benign Supporting criteria
Risk factor	Previously reported as risk factor, either variant itself or clear established pattern in gene, AND >1 in 10000 in *GnomAD*; AND the prevalence of the variant in affected individuals is significantly increased compared with the prevalence in controls

F. Statistical Analysis of the gene-panel sequencing data

a. SPSS

Statistical analysis for associations, predictors and relative risks were performed using the SPSS software (© IBM, Version 24). Statistical tests employed included logistical regression for the value of clinical predictors and the frequency of at least likely

deleterious variants, univariate analysis of variance (ANOVA) with Bonferroni post-hoc test for the age at onset and healthy survival analysis, as well as for the analysis of age at onset per gene. Fisher's exact test was used to assess the frequency of concurrent mutations. Chi-Square-Tests were used to assess the frequency of novel variants, ApoE genotype status, *GRN* missense and *TREM2* variant frequencies. Risk factors were assessed using a Chi-Square-test for the overall analysis and logistical regression to assess cohorts individually. Prior to the analysis, five key independent hypotheses were defined as stated in the introduction; subsequently, statistical analyses were carried out with a pre-defined statistical threshold of $p<0.01$. Any subsequent secondary, exploratory tests were carried out without further corrections for multiple testing. SPSS version 25 was used to assess the influence of ApoE4 on FTD patients with tau pathology using t-tests, unilateral ANOVA, Chi-Square-tests and linear regression. I personally performed all statistical tests in SPSS.

b. Copy Number Variant (CNV) Analysis

An analysis for larger insertion or deletion, i.e. a "copy number variant analysis" (CNV) of the data was carried out using DECoN[48]. Using the coverage data for each sample based on the .bam files, this software determines which areas of the sequenced genome have a significantly higher or lower coverage that expected, suggesting a duplication or deletion of at least one of the alleles, respectively. While I did not perform the bioinformatic coding and running of the analysis software, I personally analysed the output from the software for potential CNVs.

c. Penetrance calculations

v. Prevalence and lifetime risk of AD and FTD

As a basis for penetrance calculations, a Boolean literature search of pubmed.com was performed for "dementia" AND "epidemiology" from 2008 to January 2017. The retrieved articles were further screened to assess whether they addressed the prevalence of early-onset Alzheimer's disease (AD) and/or early-onset fronto-temporal dementia.

vi. Penetrance calculations

Penetrance calculations were performed in Microsoft Excel® based on the methodology laid out by Minikel *et al.*[28], both for variants identified in this dataset and reportedly pathogenic variants described in the literature. First a Boolean search of PubMed[164] was performed to estimate the prevalence and hence the baseline risk for early-onset Alzheimer's disease and early-onset fronto-temporal dementia. Based on the baseline risk (BR), the allele frequency in cases (AF_cases) in this study and the background population allele frequency on *GnomAD*[123] (AF_pop), the likely penetrance of any variant observed both in cases and controls could then be calculated as AF_cases*BR/AF_pop. Wilson's method for binary distributions was used to calculate the 95% confidence interval[28,165]. I personally performed all prevalence research, life-time risk estimates and penetrance calculations for all applicable mutations reported in this study.

4 Whole-genome sequencing of the HD phenocopy cohort

A. Defining HD phenocopy syndromes

a. Conventional definition of HD phenocopy syndromes

Huntington's disease (HD) is one of the commonest autosomal dominant adult onset neurodegenerative conditions; it is caused by a CAG repeat expansion in the first exon of the huntingtin gene on chromosome 4; it is the commonest genetically determined neurodegenerative disease, with a prevalence of at least 12.4 per 100,000 people[94]. HD is typically defined by a triad of movement, cognitive, and psychiatric symptoms; however, while chorea is the most frequent motor symptom and usually accompanied by cognitive decline, patients can also suffer from akinetic-rigid syndromes, dystonia or ataxia instead, as well as presenting with purely cognitive or psychiatric symptoms[166] (see Figure II-2). A family history may not always be obvious[13] and AAO may range from childhood to old age, with most cases developing symptoms in mid-life[166]. Due to clinical heterogeneity, the clinical diagnosis can therefore be difficult; when HD testing was first introduced in the 1990s, about 1% of those in whom HD is suspected did not carry the requisite CAG expansion[102,167]. Previous research has defined HD phenocopy syndromes as patients who were referred for HD testing by an experienced neurologist

or neurogeneticist who considered it a potential diagnosis, but who receive a negative test result[13,103]. Recent publications have identified *C9orf72* expansions as the most common identified genetic variant at the base of HD-like syndromes with a prevalence of 1.95%[13], other differential diagnoses include the HD-like syndromes (HDL) 1, 2, 3 and 4 – the latter also called spino-cerebellar atrophy 17- and dentatorubral-pallidoluysian atrophy[13,103].

Figure II-2: Symptomatic spectrum of HD and HDPC syndromes

Huntington's disease is defined by a triad of potentially very diverse symptoms affecting the motor, cognitive and psychiatric domains. While onset of the disease is marked by the onset of motor symptoms, cognitive and psychiatric symptoms can precede or follow this milestone. HDPC syndromes can therefore present a wide range of symptoms and symptom combinations that would be compatible with Huntington's disease.

To define the cohort for genetic testing, I followed the established practice, as per above. In addition, and in order to assess the goodness of fit of the precedent, I personally designed and analysed a survey on HD testing from Movement Disorder specialists and Neurogeneticists, and personally examined clinical data from patients from two clinics for Neurogenetics at NHNN with positive and negative HD tests, supplemented with patients that are part of the UCL HDPC cohort.

b. Clinical comparison between HD and HDPC patients

i. The Expert survey on HD testing

With the intention of gauging which symptoms most typically prompt experienced expert clinicians to order an HD test and which symptoms would stop them from doing so – a 10-question survey was created online[168] and sent out by email to 130 experienced neurologists and neurogeneticists, most of who are listed as the clinical leads of their respective local HD clinics on the website of the European Huntington's disease network (EHDN). Of the solicited clinicians, 51 responded and replied to the survey questions, which can be found in Chapter 2. The answers to the survey were anonymised by the website, extracted to a spreadsheet and analysed further using Microsoft Office Excel© and SPSS24.

ii. Patients from Neurogenetics clinics

Patients who presented to two separate Neurogenetics clinics at NHNN and Great Ormond Street Hospital (GOSH) between 2016 and 2018 were listed and their clinical records analysed after the exclusion of duplicate presentations. In order to attain data comparable to the survey data, patient records were checked to determine the time point when it was decided that a patient should be referred for HD testing, the clinical symptoms that were present in the patient at that point in time were then transferred to a spreadsheet. In the case of pre-symptomatic HD testing of patients with a known positive genetic HD test result in the family, the symptoms reported at the appointment following the onset of motor symptoms (which defines the official clinical onset of HD) were recorded instead.

In addition to comparing the relative frequency of HD and HDPC diagnoses in these well-established Neurogenetics clinics, prevalences of symptoms in both the HD and HDPC cohorts were calculated and compared using a Fisher's exact test in SPSS26 and logistical regression in R. In addition, conditional probabilities in R were used to assess the likelihood of two symptoms occurring together in the same patient. Symptoms that occurred at a statistically significant different rate in each cohort, as well as common symptoms, were selected for graphical representation and analysis. To even

out the numbers, the cohorts from the Neurogenetics clinic was supplemented with the most recently included HDPC patients from the genetic cohort. For details see Results Chapter 3.

c. *Stratification of the HPC cohort and sample selection for WGS*

The UCL HDPC cohort was collected based on the previously used, heuristic definition of HDPC patients. All patients had been previously suspected of HD by a specialist physician, undergone HD gene expansion testing at the Division of Neurogenetics UCLH, and received a negative HD gene test result. Previous research established *C9orf72* expansion as the most common genetic cause of a HD phenocopy syndrome[13], while *PRNP* insertional mutations, spinocerebellar ataxia-17 and dentatorubral-pallidoluysian atrophy are also known to be mistaken for HD[103]. These discoveries have broadened our understanding of HD phenocopy patients, but a genetic diagnosis is still only available to a small fraction of these patients. In the first instance, the UCL HDPC cohort was tested using the MRC Dementia Gene Panel with the added *C9orf72* expansion and *PRNP* insertion tests; patients with a causal mutations were excluded from further study (see Results Chapter 1: Gene-panel sequencing). A subcohort of carefully selected patients from the cohort was then carefully chosen for whole-genome sequencing (WGS) in order to improve our knowledge of potential causes of HDPC syndromes. The most suitable patients should bear a strong resemblance to the classical triad of HD and have a strong chance of a genetic cause of their disease; a number of factors and scores were therefore combined into a weighted ranking score to reflect the selection criteria. I personally created and weighted the score for the selection of patients for WGS. The weighting of the score is described in Table II-8; elements included the HDPC score, the modified Goldman score, the age at onset, the availability (or future availability) of neuropathological diagnostic confirmation, and how recently the patient was last seen in clinic.

A family history of neurodegenerative disease substantially increases the chances of identifying a deleterious genetic variant[24], this was represented and stratified using the Goldman score. Importantly, while the Goldman score normally rates the family history of a case according to the previous or simultaneous occurrence of other

relatives with the same or similar syndrome, in the case of HDPC patients, any relatives with a neurodegenerative syndrome was included. This was done to reflect that many of the known HDPC genes (and not least the HD *HTT* expansion itself) can cause a wide spectrum of symptoms. These may not always be recognizable as one syndrome, for example in the case of *C9orf72* expansions causing both FTD and motor neuron disease. Importantly, relatives from previous generations, even with typical HD symptoms may have been diagnosed with a multitude of different diagnostic labels and familial disorders should per se be weighted as being more HD-like. However, since recessively inherited diseases, such as Friedreich's ataxia and pantothenate-kinase-associated neurodegeneration (PKAN), have been shown to mimic Huntington's disease, a lack of other affected family members does not preclude a genetic diagnosis; adoptions, indeterminate paternity and *de novo* mutations can also explain a lack of family history. The prevalence of sporadic neurodegenerative diseases increases with age[2], and while some genetic diseases do strike later in life, a young age of onset of disease is a strong indicator that there may be an underlying genetic cause. Some patients donate their brains for post-mortem study, and while this does not make a genetic diagnosis more likely, it does render any novel variants of uncertain pathogenicity more easily interpretable, hence the allocation of points for the availability of present or future neuropathological investigations. Samples from patients who were seen in clinic more recently were also allocated extra points as it is likely to make re-establishing contact easier; the patient is more likely to be still alive and to have surviving family members who could be contacted for further investigations or segregations analyses.

Each category of the criteria for each sample was awarded a number of points; these were then multiplied by the respective criterion's weighting factor and added up to generate a total score for each sample with a maximum achievable score of 100 points. For simplicity, any patient with a familial syndrome and all the classical clinical features of HD, or a patient from a family suspected to have HD with a very strong, clear-cut autosomal dominant family history (GS 1) was selected for sequencing. Other patients, who partially resemble clinical HD, or for whom there was incomplete evidence of a

familial disorder were ranked according to the criteria set out above. Points were assigned to these criteria according to their weighted importance (see Table II-8) and then added up to create a total ranking score

Table II-8: Ranking and weighting factors for the selection of HD phenocopy samples for whole-genome sequencing

Each category of the criteria for each sample is awarded a number of points; these are then multiplied by the respective criterion's weighting factor and added up to generate a total score for each sample. The maximum achievable score in this ranking is 100 points.

Criteria	Category	Points awarded
HD phenocopy score	3	30
	2	20
	1	5
Goldman score	1	20
	2	10
	3	4
	3.5	2
	4	0
Age at onset	≤35	30
	≤45	20
	≤55	10
	≤65	5
Neuropathology	Yes	15
	No	0
Years since seen in clinic	Last 2 years	5
	Last 3 years	2
	Last 5 years	1
Maximum Total		100

DNA concentration, quantity and quality were assessed for all selected samples using a TapeStation and Genomic DNA Screen Tape (both Agilent). If DNA stocks were insufficient for WGS and could not be replenished from the Laboratory for Neurogenetics at NHNN, the next sample down in the ranking was chosen as a replacement; this was the case for 3 samples. The scores for the selected top 50 samples are listed in Table II-9.

Table II-9: Details of patient samples selected for whole-genome sequencing

Based on the selection criteria described in Table II-8, 50 samples were selected for whole-genome sequencing; their details are described in this table. Sex, age at onset (AAO) and the strength of the family history (Goldman score, see Figure II-1) were noted, and the clinical notes scoured for symptoms the patients developed in their lifetime, as well as the last time they presented in clinic at the National Hospital for Neurology and Neurosurgery (NHNN). Based on this information and their respective score, patient samples were shortlisted and selected.

HDPC Sample	Sex	AAO	Goldman Score	HDPS Movement Disorder	HDPS Cognitive Decline	HDPS Psychiatric Disturbance	HDPS Total	Neuro-pathology available	Chorea present in patient	Last Clinic within (years)	Scoring Total
1	Female	43	1	1	1	1	3	no	yes	2	85
2	Female	55	1	1	1	1	3	no	yes	2	85
3	Male	59	1	1	1	1	3	no	no	5	81
4	Male	61	1	1	1	0	2	yes	no	2	80
5	Female	69	2	1	1	1	3	no	yes	2	75
6	Female	45	2	1	1	1	3	no	yes	2	75
7	Male	50	3.5	1	1	1	3	no	yes	2	67
8	Female	49	3.5	1	1	1	3	no	yes	2	67
9	Male	59	4	1	1	1	3	no	yes	2	65
10	Male	53	4	1	1	1	3	no	no	2	65
11	Male	70	4	1	1	1	3	no	yes	2	65

HDPC Sample	Sex	AAO	Goldman Score	HDPS Movement Disorder	HDPS Cognitive Decline	HDPS Psychiatric Disturbance	HDPS Total	Neuro-pathology available	Chorea present in patient	Last Clinic within (years)	Scoring Total
12	Female	39	1	1	1	0	2	no	yes	2	65
13	Female	49	4.5	1	1	1	3	no	no	2	65
14	Male	43	4	1	1	1	3	no	yes	2	65
15	Male	64	4	1	1	1	3	no	yes	2	65
16	Female	11	1	1	0	1	2	no	yes	2	65
17	Male	43	3.5	1	1	1	3	no	no	3	64
18	Male	53	3.5	1	1	1	3	no	no	53	63
19	Male	39	3.5	1	1	1	3	no	no	-	62
20	Male	47	4	1	1	1	3	no	no	3	62
21	Female	43	1	0	1	1	2	no	no	-	60
22	Female	69	4	1	1	1	3	no	yes	-	60
23	Male	33	4	1	1	1	3	no	yes	-	60
24	Female	34	2	1	0	1	2	no	no	2	55
25	Female	55	2	1	1	0	2	no	no	2	55
26	Male	75	3.5	1	1	0	2	no	yes	2	47

HDPC Sample	Sex	AAO	Goldman Score	HDPS Movement Disorder	HDPS Cognitive Decline	HDPS Psychiatric Disturbance	HDPS Total	Neuro-pathology available	Chorea present in patient	Last Clinic within (years)	Scoring Total
27	Female	63	3.5	1	1	0	2	no	yes	2	47
28	Female	66	3.5	1	0	1	2	no	yes	2	47
29	Male	44	3.5	1	1	0	2	no	no	2	47
30	Female	62	3.5	1	1	0	2	no	no	2	47
31	Female	34	3.5	1	0	1	2	no	yes	2	47
32	Female	70	3.5	1	0	1	2	no	yes	2	47
33	Female	65	1	1	1	0	2	no	no	5	46
34	Male	16	4	0	1	1	2	no	no	2	45
35	Female	68	4	1	0	1	2	no	yes	2	45
36	Female	46	3	1	1	0	2	no	no	5	45
37	Male	77	4	1	1	0	2	no	yes	2	45
38	Male	51	4	1	1	0	2	no	yes	2	45
39	Male	64	4	1	1	0	2	no	yes	2	45
40	Female	53	4	1	1	0	2	no	yes	2	45
41	Female	46	4	1	1	0	2	no	no	2	45

HDPC Sample	Sex	AAO	Goldman Score	HDPS Movement Disorder	HDPS Cognitive Decline	HDPS Psychiatric Disturbance	HDPS Total	Neuro-pathology available	Chorea present in patient	Last Clinic within (years)	Scoring Total
42	Female	53	4	1	1	0	2	no	yes	2	45
43	Female	58	4	1	1	0	2	no	no	2	45
44	Female	22	1	1	0	0	1	no	yes	2	35
45	Female	52	1	1	0	0	1	no	yes	2	35
46	Female	52	1	1	0	0	1	no	yes	2	35
47	Male	40	1	1	0	0	1	no	no	2	35
48	Male	67	1	0	1	0	1	no	no	2	35
49	Female	35	1	1	0	0	1	no	no	-	30
50	Female	45	1	0	0	0	0	0	no	5	21

B. Sequencing, pre-processing and annotation pipeline

Samples were sequenced on a commercial basis at Edinburgh Genomics: Clinical Genomics using Illumina SeqLab, which integrates Illumina TruSeq library preparation, Illumina cBot2 cluster generation, Illumina HiSeqX sequencing, Hamilton Microlab STAR integrative automation, and Genologics Clarity LIMS X Edition, which is a high-throughput, amplicon-based Illumina sequencing platform.

a. Quality control

Samples were assessed for integrity and concentration using the Agilent 4200 TapeStation© (Agilent Technologies, Santa Clara, CA, USA) and diluted to a concentration of 50ng/ml. Upon receipt at Edinburgh Genomics (Edinburgh, UK), samples were again tested for quality and quantity using and AATI Fragment Analyzer and the DNF-487 Standard Sensitivity Genomic DNA Analysis Kit (both Agilent Technologies, Santa Clara, CA, USA); quality control (QC) results were then assessed using the AATI ProSize 2.0 software (Agilent Technologies, Santa Clara, CA, USA), which scores the sample integrity and only samples with a quality score <7 were taken forward for sequencing after normalisation.

b. Library preparation, Library QC and Sequencing

At Edinburgh Genomics, using the Illumina SeqLab specific TruSeq Nano High Throughput library preparation kits (Illumina, San Diego, CA, USA) in conjunction with the Hamilton MicroLab STAR (Hamilton®, Reno, NV, USA) and Clarity LIMS X Edition (Illumina, San Diego, CA, USA), libraries WGS were prepared, normalised to the concentration and volume required and sheared to a 450bp mean insert size using a Covaris LE220 focused-ultrasonicator (Covaris, Woburn, MA, USA). After ligation to blunt ended, Atailed, size selected, TruSeq adapters (Illumina, San Diego, CA, USA), the sequences were enriched using 8 cycles of PCR before the libraries were assessed for mean peak size and quantity using the Caliper GX Touch with a HT DNA 1k/12K/HI SENS LabChip and HT DNA HI SENS Reagent Kit (all PerkinElmer, Waltham, MA, USA). Based on this data, the libraries were normalised to 5nM and the actual concentration was measured using a Roche LightCycler 480 and a Kapa Illumina Library Quantification kit and Standards (all Roche, Basel, Switzerland).

For sequencing, the libraries were denatured and pooled in eights for clustering and sequencing using a Hamilton MicroLab STAR with Genologics Clarity LIMS X Edition (Hamilton®, Reno, NV, USA); they were clustered onto HiSeqX Flow cell v2.5 on cBot2s (Illumina, San Diego, CA, USA), before being transferred with the clustered flow cell to a HiSeqX for sequencing using a HiSeqX Ten Reagent kit v2.5 (both Illumina, San Diego, CA, USA).

c. Bioinformatics analysis and pre-processing

The generated sequence was demultiplexed using bcl2fastq (2.17.1.14, Illumina, San Diego, CA, USA), allowing 1 mismatch when assigning reads to barcodes adapters (Read1: AGATCGGAAGAGCACACGTCTGAACTCCAGTCA, Read2: AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT) were trimmed during this process. Alignment, bam file preparation and variant detection were performed using BCBio-Nextgen (0.9.7) with bwa mem (0.7.13)[169] to align the raw reads to the Genome Reference Consortium Human Build 38 (GrCh38) (with alt, decoy and HLA sequences); samblaster (0.1.22)[170] was used to mark the duplicated fragments, and the Genome Analysis ToolKit (GATK, 3.4-0-g7e26428) for the indel realignment, base recalibration, and genotype likelihoods[171]. The final gvcf file was created using HaplotypeCaller [171].

C. Whole-genome sequencing data analysis

a. Analysis using ExpansionHunter®

Many of known HD phenocopy syndromes are other expansion disorders. While patients had typically been investigated clinically for at least some of the alternative causes of their symptoms, this was necessarily heterogeneous and incomplete. All samples selected for whole-genome sequencing had however been screened for C9orf72 expansions during the gene-panel sequencing part of the project.

In order to exclude other known expansion-related causes of neurodegenerative disease, we used a programme called ExpansionHunter®, which uses data from typical whole-genome sequencing to investigate the potential presence of disease-causing expansions[130]. It is typically applied to PCR-free WGS data, but after consultations with

the creators of the programme and additional quality control steps, our data was assessed as suitable for this programme. ExpansionHunter® typically screens the genes *AR*, *ATN1*, *ATXN1*, *ATXN2*, *ATXN3*, *ATXN7*, *ATXN10*, *C9ORF72*, *CACNA1A*, *CBL*, *CSTB*, *DMPK*, *DMPK*, *FMR1*, *FXN*, *HTT*, *JPH3* and *PPP2R2B* for known disease-causing expansions.

b. Ingenuity Variant Analysis Platform

The variant annotation and interpretation analysis were generated through the use of Ingenuity® Variant Analysis™ software (IVA™, www.qiagenbioinformatics.com) from Ingenuity Systems. IVA™ is an online analysis platform, which offers in internal knowledge base using published research and links to external databases. In this study, we used Variant Analysis™ version 5.4.20190308; this version references CADD v1.3, Allele Frequency Community 2018-09-06, EVS ESP6500SI-V2, Refseq Gene Model 2018-07-10, JASPAR 2013-11, Ingenuity Knowledge Base Snapshot Timestamp 2019-03-01 11:17:42.0, Vista Enhancer 2012-07, Clinical Trials Stepford 190301.000, PolyPhen-2 v2.2.2, 1000 Genome Frequency phase3v5b, ExAC 0.3.1, iva Oct 4 11:04 iva-1.0.736.jar, PhyloP 2009-11, DbSNP 151, TargetScan 6.2, GENCODE Release 28, CentoMD 5.0, Ingenuity Knowledge Base Stepford 190301.000, OMIM May 26, 2017, *GnomAD*[123] 2.0.1, BSIFT 2016-02-23, TCGA 2013-09-05, Clinvar 2018-08-01, DGV 2016-05-15, COSMIC v86, HGMD 2018.3, and SIFT4G 2016-02-23.

Samples were uploaded to IVA™ in a standard variant call format (VCF) using the high-volume data stream uploader and combined and filtered for further analysis. I personally performed all IVA™ analyses and designed the analysis steps.

i. In silico prediction tools and scores

PolyPhen-2

The PolyPhen-2 score expresses the probability that an amino acid substitution will be damaging to the structure and function of a human protein based on physical and comparative considerations. Based on its probability calculations, estimates false and true positive rates and outputs a qualitative appraisal as benign, possibly damaging or probably damaging[172].

SIFT

SIFT is an acronym of "Sorting Intolerant from Tolerant" and is a programme predicting whether an amino acid change will affect protein function, relying heavily on phylogenetic conservation. It ranges from 0 to 1, with scores predicting a damaging substitution <0.05 and >0.05 predicting a tolerated substitution[173].

BSIFT

This is a bi-directional form of the SIFT score, which compares the mutant and the wild-type allele in order to assess relative functional activity; it ranges from -1 to 1 with higher values representing activating variants[174].

CADD score

The CADD score (combines annotation dependent depletion) is a scoring tool that combines multiple sources of information and annotations. The generated raw scores are transformed into scaled, phred-like scores ranging from 1 to 99, based on its rank relative to all possible 8.6 billion variants in the human reference genome. A CADD score of ≥20 refers to the top 1% of potentially deleterious variants, while a CADD score of ≥30 refers to the top 0.1% of potentially deleterious variants, and etc.[175].

PhyloP score

The phyloP score measures the evolutionary conservation at an individual base-pair level and expresses the evolutionary constraints that may apply to functionally significant genetic variation[176]. Positive score equal slower than expected evolution at "conserved" sites, while negative scores equal faster than expected evolution at "fast-evolving" sites.

MaxENTScan

MaxEntScan is based on modelling the sequences of short sequence motifs based on maximum entropy non-adjacent as well as adjacent dependencies between positions; it can be used to score both 5' and 3' splice sites. When applied to a splice site variant, this score will assess any changes in entropy and whether there is disruption to the splice site[177,178].

ENCODE TFBS

IVA™ uses ENCODE TFBS (transcription factor binding site) to qualify the calling of a loss of a promoter site. The ENCODE project aims to identify and list all functional elements in the human genome and make them accessible to the research community using both databases and algorithms[179].

RVIS gene score

The Residual Variation Intolerance Score (RVIS) gene score is not part of the IVA™ platform but was used to manually further restrict the number of genes of interest. The RVIS is a gene-based measure of genetic intolerance. It is based on allele frequency data in the online population database ExAC and ranks genes according to whether they have more or less common functional variation than would be expected based on the genome wide expectation and apparently neutral variation in the gene. A negative score is a measure of an intolerant gene, while a positive score expresses more tolerance for genetic variation. In addition, all genes are ranked from most intolerant to tolerant and given a percentile of how intolerant of variation it is[180]. Genes excluded on this basis are listed in Table V-2.

ii. IVA™ filters

IVA™ offers a number of filters and tools that can be customized using many different options to narrow down the number of variants and genes that may be linked to the syndrome in question. Even pre-analysis, one is asked to choose whether to include only variants in exonic regions, exclude common variants observed with high allele frequency in public databases (1000 Genomes Project, *ExAC*, *GnomAD*, NHLBI ESP exomes and the allele frequency community) or present in dbSNP, and to keep only variants above minimum confidence standards. The analysis setting then offers more filters and options to focus the results. More detailed descriptions of the filter options are described below, and an example of how such filters may be applied is shown in Figure II-3.

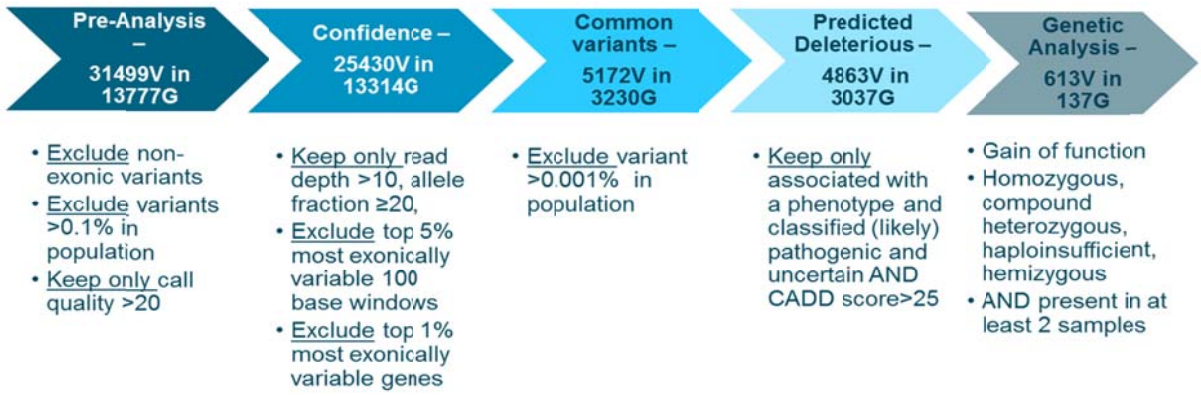

Figure II-3: Filtering pipeline for exonic variants

In this example, only exonic variants with a frequency in the population of less than 0.1% and of a sufficient call quality (>20) were included in the analysis and processed further. These were then further filtered using additional, more stringent confidence filter, excluding all but the rarest variants observed on population databases (<0.001%) and keeping only variants meeting certain criteria for a likely deleterious effect on. Finally, variants were restricted to gain of function or loss of function variants identified in at least 2 samples. Further filters could then be applied to the remaining variants. In the arrows, "V" stands for "variants" and "G" stands for "genes".

Confidence filter

The confidence in the call quality filter is based on call quality, the variant passing upstream pipeline filtering, read depth, genotype quality, allele fraction, and the variant being located outside the top x% most exonically variable 100base windows in healthy public genomes and outside the top x% most exonically variable genes in healthy public genomes (1000 Genomes). The call quality score transforms the QUAL column in a standard VCF file using a Phred scale and expresses the likelihood of a non-reference variant actually being present at the reported site. For the purposes of this analysis, these settings were kept constant at a call quality of at least 20 in any case, a read depth of at least 10, and an allele fraction of at least 20; variants were located outside the top 5% most exonically variable 100base windows in healthy public genomes and outside the top 1% most exonically variable genes in healthy public genomes (1000 Genomes).

Common variants filter

The aim of the common variants filter is to exclude variants from the analysis that are common in the normal population, or at least more common than the disease of phenotype of interest. It offers the possibility to exclude variants found in online population databases at an allele frequency of at least 0.001% (lowest possible setting)

in all samples in the Allele Frequency Community, the 1000 Genomes Project, and the NHLBI ESP exomes, as well as either all or the highest subpopulation frequency on ExAC and *GnomAD*[123], or are present in dbSNP, or unless they are an established pathogenic common variant. In this analysis, all variants observed in online population databases at an allele frequency of at least 0.001% in all samples in all the population databases listed above were excluded, as were variants present on dbSNP, without exceptions for established Pathogenic common variants.

Predicted deleterious filter

Selections made here allow for the selection of variants predicted or observed to affect gene function or expression, with the option of keeping only variants no more than a certain number of bases into an intron. They can have previously been associated with a phenotype and be classified according ACMG guidelines, or listed on the HGMD®, ClinVar or CentoMED™ databases. Or they can be associated with gain of function of a gene as per the literature, as a gene fusion, inferred activation mutation by Ingenuity, predicted gain of function by BSIFT or as a microRNA binding site. Or variants can be associated with loss of function of a gene as a frameshift, in frame indel, or start/stop codon change, a missense unless predicted tolerated by SIFT or PolyPhen-2, predicted deleterious by having a CADD score over a certain cut-off (typically 20 or 25), being nullizygous, a splice site loss up to a number of bases, typically 2, into an intron or as predicted by MaxEntScan, deleterious to a microRNA, copy number loss, promoter loss (with or without ENCODE TFBS), enhancer, evolutionarily conserved with a phyloP p-value of less than or equal to 0.0, or falling into an untranslated region (UTR). For the purposes of this analysis, variants that were classified likely benign or benign according to the ACMG guidelines were excluded; the remainder of the options was adapted to individual analysis requirements, as explained in the "Gene- and expression based analysis" and "By-variant analysis" sections.

Genetic analysis filter

In principle, IVA™'s genetic analysis filter permits the selection of inferred gain- or loss-of-function variants, tumour specific variants, recessive variants or dominant variants.

As further options, gain-of-function heterozygous, compound heterozygous, homozygous, haplozygous and haploinsufficient variants can be selected for

Biological Context

The biological context filter highlights variants that have published links to diseases or drug response based on the input of biological or clinical terms, including diseases, symptoms, genes, domains, processes and pathways. For the phenotype-driven ranking analyses, the terms “chorea”, “dystonia”, “dementia”, “cognitive decline”, “psychiatric disturbance”, “hereditary disease”, “neurodegeneration” and “Huntington disease” were included.

Phenotype-driven ranking

This filter allows for the input of Human Phenotype Ontology IDs, phenotypes, and diseases, which are then computed to rank candidate genes by their known associations with the search terms, prioritizing combinations of several search terms. This filter was only used once in the phenotype-driven ranking section of the variant-driven analysis. The symptoms and phenotypes supplied were “chorea”, “dystonia”, “dementia”, “cognitive decline”, “psychiatric disturbance”, “Huntington disease”, and “striatal degeneration”.

IVA™ filtering option not applicable to this study

Cancer driven variants

Within an analysed dataset, this filter highlights variants which are predicted or established to be associated with tumorigenesis or metastasis.

Custom Annotation

These settings enable the generation of filters based on a variant, regions, or gene annotations imported into IVA™.

Pharmacogenetics

Based on published evidence in the literature, this filter permits the identification of variants inferred or observed to impact drug response, metabolism or toxicity.

Physical Location

This filter permits the identification of variants located on a particular chromosome or part thereof.

Statistical Association

This filter requires control samples to be uploaded to Ingenuity; it uses association tests to compare allele frequencies between case and control samples.

User-Defined Variants

Using this filter, the user can define and save a set of variants for reuse in other analysis.

iii. Gene- and expression based analysis

Ingenuity gene filter pipeline

For this analysis, the dataset was limited to exonic variants only in the pre-analysis filters. Variants and genes were further restricted using the confidence and common variant filters, as explained above and as illustrated in Figure II-3Figure II-3. The Predicted deleterious filter kept only variants associated with a phenotype and classified as pathogenic, likely pathogenic or of uncertain significance, or associated with a gain-of-function or with a loss of function except for falling into an untranslated region. Then a Genetic analysis filter would be applied, limiting variants to genes in which at least two variants had been found.

Manual gene list curation

Subsequently, the data was exported from IVA™ and manually curated further using the RVIS score, numbers of cases with variants in a given gene, and Genotype-Tissue Expression (GTEx) data. Genes that were among the 20% most tolerant of genetic variation were excluded from further analysis, as were genes in which variants meeting the above criteria were found in more than 15 cases, as we were looking for rare causes of disease.. For the remaining 256 genes, gene expression levels were checked on GTeX[181]; the highest median gene expression in the central nervous system (CNS) was compared to the highest non-neural expression level of the gene anywhere in the body to gauge whether the gene in question is relatively highly expressed in the brain. The gene encoding the Huntington's disease protein, *HTT*, which had a ratio of 0.591, was chosen as the cut-off and anything with a lower ratio (range 0 - 0.589) was excluded from further analysis. The excluded genes were added as additional exclusion filters to IVA™ (see Table V-3) and the remaining variants analysed using the platform.

Genes were shortlisted based on their intolerance to genetic variation as per the RVIS and exported variants arranged in the same order.

iv. Exploratory variant analysis

In addition to the gene-based, streamlined approach described above, we also attempted to identify variants of interest based on criteria more focused on gene effects and phenotypic correlations; this analysis was applied to the entire genomes, subject only to the filters described below. Details for filter options are described in "IVA™ filters".

The pre-analysis filter excluded any variant observed at least 1% in the online population databases; the confidence filter used our standard settings. These variants were further filtered to exclude variants observed at least 0.001% of all samples in the Ingenuity® Allele Frequency Community, the 1000 Genomes Project, the NHLBI ESP exomes, and the subpopulation with the highest allele frequency in ExAC and in *GnomAD*[123]; pathogenic common variants were also excluded as we were looking for rare Mendelian causes of disease. Subsequently, only variants no more than 20 bases into an intron, and not classified benign or likely benign – or listed in HGMD®, ClinVar or CentoMD™ were included. Added to these were variants associated with a gain of function of a gene established in the literature, due to gene fusion, predicted to cause gain of function by BSIFT, or were associated with loss of function of a gene due to a frameshift, in-frame indel, or start/stop codon change, predicted deleterious by having a CADD score >20, a splice site loss up to 2 bases into an intron or as predicted by MaxEntScan, or predicted to lead to a copy number loss.

Subsequently, separate analyses were carried out using additional filters:

- novel deleterious variants, assessing only variants associated with a phenotype and classified either likely pathogenic or pathogenic for their possible links to neurodegenerative disease
- disruption to start and stop codons, as well as breakpoint variants, assessed for their phenotypic links and functional implications
- Phenotype-driven ranking analysis, using the "Biological context filter" and the "Phenotype-driven ranking" described above.

- Haploinsufficient variants with a CADD score ≥25, with at least two hits in the same gene
- Variants in 270 genes known to cause neurodegenerative disease (see Table V-4 in the appendix.

The results are summarized under Variant-based analysis in Results Chapter 4 and are organized according on their purported effect.

III. Results

1 Chapter 1: MRC Dementia Gene Panel sequencing

A. Descriptive statistics of the panel sequencing dataset

a. Patient samples

The total sample count of 3241 samples was made up of 2784 patient and 457 control samples, including four patient cohorts (AD, FTD, Prion and Dementia with motor symptoms). Patient samples comprised 1052 Alzheimer's disease (AD) cases, 794 frontotemporal dementia (FTD) cases, 299 samples collected from patients seen through the National Prion Clinic, and 639 samples from patients with a dementia syndrome with motor symptoms. 2352 patient samples were chosen retrospectively from the MRC Prion Unit DNA sample collection, if consent for research use was available. Over the years, these samples had been sent in for single gene testing by clinicians working mostly at the National Hospital for Neurology and Neurosurgery as well as other London hospitals. These samples also included 120 FTD samples from Cambridge sent in for earlier genetic testing. In addition to these retrospectively analysed samples, 165 patients were recruited prospectively from Memory clinics in Bristol, Cardiff, Southampton and UCLH to be tested specifically in this study. Furthermore, 267 patient samples had been referred for testing to the Neurogenetics Laboratory at the National Hospital for Neurology and Neurosurgery (NHNN); these samples were tested using a panel similar to the MRC Dementia Gene panel and included on the results level (non-synonymous variants only). In total, 432 patients (15.5% of all patients) were therefore included prospectively in the study. Baseline characteristics of patients are shown in Table III-1.

Cohorts were analysed regarding their gender balance, the rate of early-onset disease before the age of 65 and the strength of their family history (stratified by Goldman score, see Figure II-1). The site of the sample origin, i.e. whether it was a retrospective sample chosen from the collection of a cognitive centre, a prospectively recruited sample from a Memory clinic at a primary referral centre, a sample from the National

Prion Clinic or a control sample was also noted and rates of variant detection compared between the samples chosen from a centre collection or a Memory clinic.

Table III-1: Baseline characteristics of the included patient cohorts and controls

The 3241 samples were made up by 1052 AD patients, 794 FTD patients, 299 prion patients and 639 patients with dementia with motor symptoms, as well as 457 elderly controls. This table shows the distribution of each cohort in numbers and percentages for their baseline characteristics sex, age at onset (AAO), early- or late onset disease (over or under 65 AAO), Goldman score[150] and the site of the sample origin, i.e. whether it was a retrospective sample chosen from the collection of a cognitive centre or the National Prion Clinic, or, a prospectively recruited sample from a Memory Clinic at a primary referral centre, or the NHNN Division of Clinical Neurogenetics, or a control sample. AAO in controls refers to age at testing.

					Goldman Score						Site of sample origin				
Cohort	Total N	Male (%)	Mean AAO	Early Onset (%)	1	2	3	3.5	4	4.5	NHNN Cognitive Centre	NHS Cognitive Disorder Clinics	National Prion Clinic	NHNN Clinical Neurogenetics	Cardiff University
AD	1052	400 (38.0%)	57.7	78	73 (6.9%)	30 (2.9%)	66 (6.3%)	103 (9.8%)	245 (23.3%)	535 (50.9%)	873 (83.0%)	103 (9.8%)	1 (0.1%)	75 (7.1%)	0
FTD	794	419 (52.8%)	58.4	61	72 (9.1%)	43 (5.4%)	50 (6.3%)	64 (8.1%)	298 (37.5%)	267 (33.6%)	645 (81.2%)	45 (5.7%)	0	104 (13.1%)	0
Prion	299	121 (40.5%)	57.6	55	34 (11.4%)	6 (2.0%)	11 (3.7%)	15 (5.0%)	207 (69.2%)	26 (8.7%)	2 (0.7%)	0	296 (99.0%)	1 (0.3%)	0
DemMot	639	280 (43.8%)	55.8	70	24 (3.8%)	12 (1.9%)	18 (2.8%)	84 (13.2%)	215 (33.7%)	286 (44.8%)	535 (83.7%)	17 (2.7%)	0	87 (13.6%)	0
Controls	457	237 (51.9%)	76.6	-	4 (0.9%)	0	0	2 (0.4%)	0	451 (98.7%)	6 (1.3%)	0	3 (0.7%)	0	448 (98.0%)
All	3241	1457 (45.0%)	60.3	59	207 (6.4%)	91 (2.8%)	145 (4.5%)	268 (8.3%)	965 (29.8%)	1565 (48.3%)	2061 (63.6%)	165 (5.1%)	300 (9.3%)	267 (8.2%)	448 (13.8%)

b. *Control samples*

448 control samples were obtained from relatives of patients attending the memory clinic at Cardiff and Vale NHS Hospital. An additional 9 controls were samples processed in the MRC Prion Unit Human Genetics group as predictive testing in families with a known genetic mutation causing disease, but who were not mutation carriers. Baseline characteristics of controls are set out in Table III-1.

c. *Sequencing statistics*

On the MRC Dementia Gene Panel, samples had an average of 157,350 mapped sequencing reads, 90% of which were on target; average mean depth of coverage was 676 with a uniformity of 94.7%. Uniformity of sequencing expresses the distribution of overall coverage to ensure all regions have been sufficiently sequenced for interpretation. On average, 99.5% of the target sequence was covered at least tenfold.

B. Classification of variants

a. *Deleterious and likely deleterious variants*

There is no computational or experimental tool to perfectly classify individual variants by their pathogenicity, the current start-of-the-art clinical method is decision making that considers multiple factors and is adaptable to multiple potential genes/disease mechanisms (see Material and Methods, Table II-6 and Table II-7). All the analyses reported here were repeated following exclusion of known family members (n=73), on the basis of the proband being identified in the family history of the second case, with no significant change to any findings for interpretation. Following the principles set out in Material and Methods, I classified 13,152 variants in 3,241 individuals (Figure III-1 and Table III-2) and identified 352 DVs (deleterious or likely deleterious variants) in 341 patients (12.2% of patients, $p=2.8 \times 10^{-14}$, OR: 31.8, 95% CI); 7.88, 127.94; of the deleterious variants, 68 were *C9orf72* expansions and 28 were *PRNP* octapeptide repeat insertions (OPRIs). Exploratory Chi-Square tests revealed that a DV was most likely to be identified in prion disease ($p=2.196 \times 10^{-34}$; Odds Ratio: 90.91, 95% CI (21.28, 333.33)), followed by FTD ($p= 5.1126 \times 10^{-32}$; Odds Ratio: 58.82, 95% CI (14.29, 250)), AD ($p=3.0345 \times 10^{-9}$; Odds Ratio: 16.13, 95% CI (3.95, 66.67)) and patients with DemMot (p=0.000061; Odds Ratio: 9.61, 95% CI (2.28, 41.67)). Details of the

distribution of DVs in the cohorts can be found in Table III-2.Two additional DVs were seen in two male controls of 457 controls (0.4%, Figure III-1, Table III-2); one carried a missense mutation in *PRNP* (Gln212Pro) and the other a frame shift mutation in *GRN* (c.708+5_708+8delGTGA). The *PRNP* Gln212Pro variant was found in a healthy elderly control who was 67 years old at sampling. This variant is associated with Gerstmann-Straussler-Scheinker disease, which only rarely is associated with clinical onsets older than 67. The frameshift mutation in *GRN* affects a splice-site and has been previously reported in two FTD patients[182]. The control participant in whom this GRN c.708+5_708+8delGTGA variant was found in was 77 years old at sampling. Average age at onset for the GRN cases in our study was 60 years, ranging from 25 to 86 years of age at onset; the FTD patients reported previously had an AAO of 73 and 65, which also puts them in the higher age range of GRN DVs. Despite the two cases being elderly, the possibility of a clinical onset later in life can therefore not be excluded in either case.

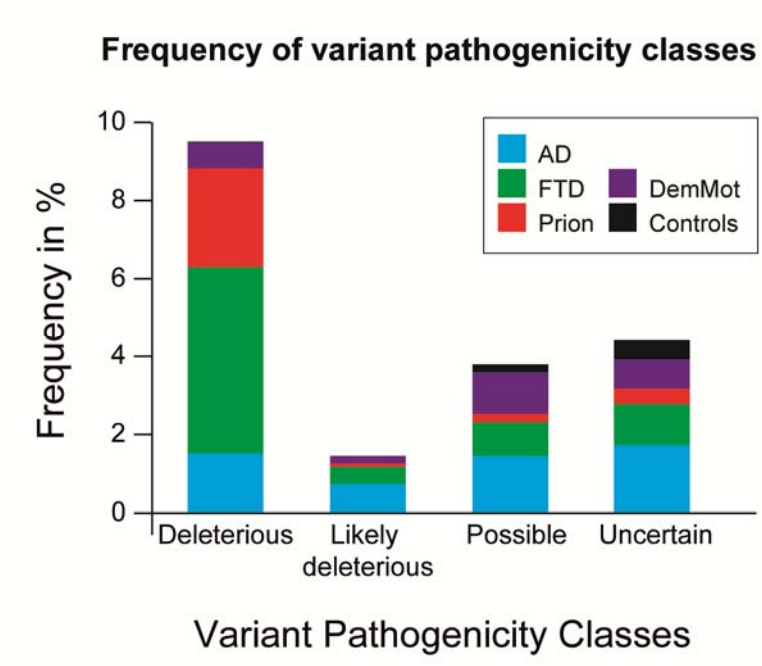

Figure III-1: Frequency of variant pathogenicity classes in the dataset

This figure shows the frequency of the various variant pathogenicity classes in the total dataset broken down by individual phenotypes

In addition to these 343 individuals carrying DVs, 121 possibly deleterious variants were found in 3.5% of all samples (4.1% patients, 1.3% of controls), in excess in cases vs. controls (p=0.005, OR: 3.06, 95% CI (1.34, 7.01)). Whilst the available data is not sufficient to classify variants in the potentially deleterious category as disease-causing, this suggests that two thirds of variants in this category might be reclassifiable as DVs if sufficient data were available. Potentially deleterious variants were most likely to be identified in patients with dementia with motor symptoms (p=0.023; Odds Ratio: 3.16, 95% CI (1.18, 8.47)), followed by AD patients (p=0.036; Odds ratio 2.74, 95% CI (1.06, 7.04), while the association did not reach significance in FTD and prion patients. This indicates that less established variants could play a significant role especially in the less well defined syndromes with atypical symptoms, which may also sometimes be given an AD diagnosis by default. These less established potentially deleterious variants warrant further research specific to each case including the potential for segregation in families.

Table III-2: Number of variants in each pathogenicity class observed in this dataset

The table shows the number of variants in each variant pathogenicity class per cohort and the percentage of variants per cohort. 15 novel DVs were identified in AD, 24 in FTD, 24 in prion patients, and 3 DemMot patients. AAO could not be established for 243 patient and control samples, thus the sum of Early and Late-Onset does not equal All Ages.

N Variants		Deleterious	Likely deleterious	Novel DVs	Possible	Uncertain	Likely benign	Benign	Risk Factor	Synonymous	Total N Cohort
Early-Onset	**AD**	38 (4.6%)	19 (2.3%)	12 (1.5%)	34 (4.2%)	70 (8.6%)	39 (4.8%)	545 (66.6%)	147 (18%)	1957 (239.2%)	818
	FTD	112 (23%)	10 (2.1%)	14 (2.9%)	13 (2.7%)	65 (13.3%)	24 (4.9%)	314 (64.5%)	81 (16.6%)	1364 (280.1%)	487
	Prion	57 (35%)	1 (0.6%)	0 (0%)	4 (2.5%)	2 (1.2%)	5 (3.1%)	93 (57.1%)	29 (17.8%)	563 (345.4%)	163
	DemMot	19 (4.3%)	4 (0.9%)	2 (0.4%)	14 (3.1%)	20 (4.5%)	24 (5.4%)	278 (62.2%)	82 (18.3%)	1221 (273.2%)	447
	Controls	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	47 (1175%)	5 (125%)	2 (50%)	17 (425%)	4
	Total	226 (11.8%)	34 (1.8%)	29 (1.5%)	65 (3.4%)	157 (8.2%)	139 (7.2%)	1234 (64.3%)	341 (17.8%)	5122 (266.9%)	1919
Late-Onset	**AD**	9 (4%)	4 (1.8%)	3 (1.3%)	12 (5.3%)	9 (4%)	91 (40.3%)	166 (73.5%)	39 (17.3%)	698 (308.8%)	226
	FTD	16 (9.9%)	2 (1.2%)	3 (1.9%)	6 (3.7%)	20 (12.3%)	59 (36.4%)	84 (51.9%)	34 (21%)	451 (278.4%)	162
	Prion	5 (7.5%)	1 (1.5%)	0 (0%)	2 (3%)	1 (1.5%)	17 (25.4%)	29 (43.3%)	21 (31.3%)	204 (304.5%)	67
	DemMot	2 (1.1%)	1 (0.6%)	0 (0%)	8 (4.5%)	6 (3.4%)	85 (48.3%)	100 (56.8%)	17 (9.7%)	471 (267.6%)	176
	Controls	2 (0.4%)	0 (0%)	0 (0%)	5 (1.1%)	12 (2.7%)	3 (0.7%)	314 (70.1%)	49 (10.9%)	1415 (315.8%)	448
	Total	34 (3.2%)	8 (0.7%)	6 (0.6%)	33 (3.1%)	48 (4.4%)	254 (23.5%)	690 (63.9%)	160 (14.8%)	3239 (300.2%)	1079
All Ages	**AD**	48 (4.6%)	23 (2.2%)	15 (1.4%)	46 (4.4%)	79 (7.5%)	130 (12.4%)	717 (68.2%)	187 (17.8%)	2676 (254.4%)	1052
	FTD	155 (19.5%)	14 (1.8%)	20 (2.5%)	24 (3%)	107 (13.5%)	99 (12.5%)	500 (63%)	132 (16.6%)	2262 (284.9%)	794
	Prion	82 (27.4%)	3 (1%)	1 (0.3%)	6 (2%)	6 (2%)	35 (11.7%)	162 (54.2%)	67 (22.4%)	999 (334.1%)	299
	DemMot	22 (3.4%)	5 (0.8%)	3 (0.5%)	23 (3.6%)	32 (5%)	119 (18.6%)	387 (60.6%)	106 (16.6%)	1750 (273.9%)	639
	Controls	2 (0.4%)	0 (0%)	0 (0%)	5 (1.1%)	12 (2.6%)	51 (11.2%)	321 (70.2%)	54 (11.8%)	1447 (316.6%)	457
	Total (% of patients)	**309 (9.5%)**	**45 (1.4%)**	**39 (1.2%)**	**104 (3.2%)**	**236 (7.3%)**	**435 (13.4%)**	**2087 (64.4%)**	**546 (16.8%)**	**9134 (281.8%)**	**3241**

143 variants that could not be classified as benign or deleterious, termed uncertain, were seen in 4.4% samples (4.6% patients, 3.5% controls, p=0.37); synonymous variants in all tested genes were also overall not significantly more frequent in cases than controls (p=0.066). Due to reporting of variants in clinically sequenced cases, the analyses of uncertain and synonymous variants were restricted to cases run on the MRC Dementia Gene Panel.

Separately from this present study, 715 of the 2984 samples that were analysed with the MRC Dementia Gene Panel had also been exome-sequenced, which allowed discovery of only two additional mutations classified as DVs, in *TBK1*[183], and *DNMT1*[184], and no DVs in other neurology-relevant genes not included in the panel.

b. Novel deleterious variants

Novel variants were defined as variants that had not previously been reported in the literature nor found in the G*nomAD*[123] of exomes and genomes[124] at the time of analysis. Out of the 343 samples with 352 DVs detected in this dataset, 39 (11.3% patients with DVs, Chi-squared test p=0.004 were novel DVs, Figure III-2, Table V-5; 16 were identified in *GRN*, eight in *PSEN1*, six in *MAPT*, six in *CSF1R* and one each in *NOTCH3*, *PSEN2*, and *VCP*. Because of the well-known disease mechanism of *GRN* related to loss of function, novel variants in this gene were easier to classify as DVs than those in genes with less well understood pathomechanisms.

Overall, 1.4% of all included patients carried a novel DV, no *novel* DVs were found in controls. In a chi-square-test, cases were significantly more likely to carry a novel variant than controls albeit by a small margin. (p=0.04; Odds ratio: 1.14). In AD patients, 15 novel variants were identified (1.4% of the cohort), in FTD patients, 24 novel variants were identified (3.0% of the cohort), in prion patients two novel variants were identified (0.67% of the cohort), and in DemMot patients three novel variants were identified (0.66% of the cohort). Compared to controls, AD cases were significantly more likely to carry a novel variant (p=0.009; Odds ratio: 1.01, 95% CI (1.0, 1.02)), as were FTD patients (p=1.8 x 10-4; Odds ratio: 1.02, 95% CI (1.01, 1.03)), but for Prion patients (p=0.395) and patients with Dementia with motor symptoms

(p=0.145), the difference did not reach significance, possibly due to the low variant counts. Novel variants, and the underlying reasoning for their classification, are listed in the Appendix, Table V-5)

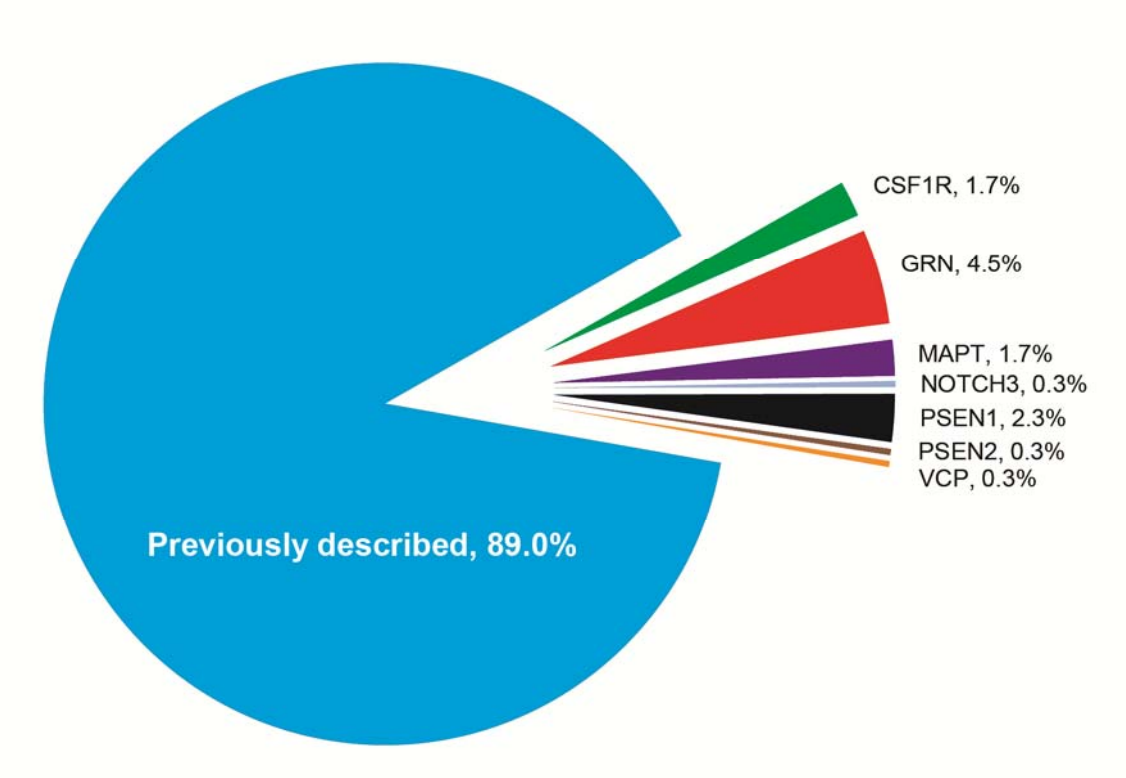

Figure III-2: Genes in which novel DVs were found.

11% of DVs were not previously described in the literature (n=39). Known mutations were found in PRNP (24.6%), C9orf72 (19.2%), MAPT (15.8%), GRN (9.9%), PSEN1 (9.6%), APP (3.4%), CSF1R (2.0%), VCP (1.7%), SQSTM1 (0.9%), TARDBP (0.6%), and CHMP2B, ITM2B, NOTCH3, PSEN2, and TREM2 (0.3% each).

c. Concurrent deleterious variants

There have been several reports of concurrent pathogenic mutations in patients with FTD[85,104,105]. Patients with DVs were encoded as two groups, those with one DV and those with two (none carried more than two variants fulfilling the criteria for a DV). 11 patients were found to carry two separate DVs in one or more genes and compared to those with one DV using a Binomial test[185]; the results showed an out of chance expectation, with $p=5.4 \times 10^{-14}$. These 11 patients with concurrent DVs represented 3.2% of patients with at least one DV, 0.4% single DV in controls; one AD patient, seven FTD patients, two patients with prion disease and one patient with DemMot). There

was a notable excess in FTD, as previously suggested by case reports (observed 7, expected 2.7, Chi-squared test, p=0.003)[85].

d. Non-UK ancestry

14 DVs were observed in 12 patient samples of non-UK ancestry; of these, 2 were *C9orf72* expansions and 2 cases carried a double DV. None of the DVs found in samples of non-UK ancestry were observed on *GnomAD*, except for *SQSTM1* Arg107Trp, the secondary variant of the concurrent *VCP* Arg155His mutation; at the time of analysis, *SQSTM1* Arg107Trp was described on *GnomAD* in 4 South Asian, 1 East Asian and 3 European samples, with a frequency of 0.013%, 0.0058% and 0.0027%, respectively.

e. Deleterious variants in genes not typically associated with the syndrome

DVs were often discovered in patients with clinical syndromes that would not normally prompt a request for sequencing of the implicated gene. In 58 patient samples, DVs were identified in genes that would not normally be screened in the clinical syndrome of the case it was discovered in; this equates to 16.9% of all patients with DVs, p=0.013, Figure III-3. This was calculated using a Chi-Square test comparing patients with DVs to controls; patients could have DVs in genes associated with their clinical syndrome or not, while all DVs in controls were treated as unexpected. In patients diagnosed with AD, three *C9orf72* expansions, nine DVs in *MAPT*, five in *CSF1R*, two in *GRN,* three in *PRNP*, and one each in *SQSTM1*, *TARDBP* and *VCP* were identified, as well as a homozygous *TREM2* DV normally associated with Nasu-Hakola disease. For FTD patients, five DVs were seen in *VCP*, three variants each were seen in *CSF1R* and *PSEN1*, two in *PRNP* and SQSTM1, and one each in *NOTCH3* and *CHMP2B*. While these latter two DVS were not found in genes completely unassociated with FTD, the gene tests would not normally be requested as a standalone test in clinical practice. Two DVs in *PSEN1* and one each in *GRN* and *VCP* were identified in patients referred with suspected prion disease, the latter as part of a concurrent mutation with a *PRNP* DV. In DemMot, four DVs were identified in *PSEN1*, three variants in *MAPT*, as well as one variant each in *ITM2B*, *PRNP*, *GRN* and *PSEN2*. However, based on an exploratory logistical regression, DVs in unexpected genes were not more likely to be identified in

one particular cohort compared to controls (Cohorts, p=0.235) or in patients with either a particularly weak or strong family history (Goldman score, p=0.836).

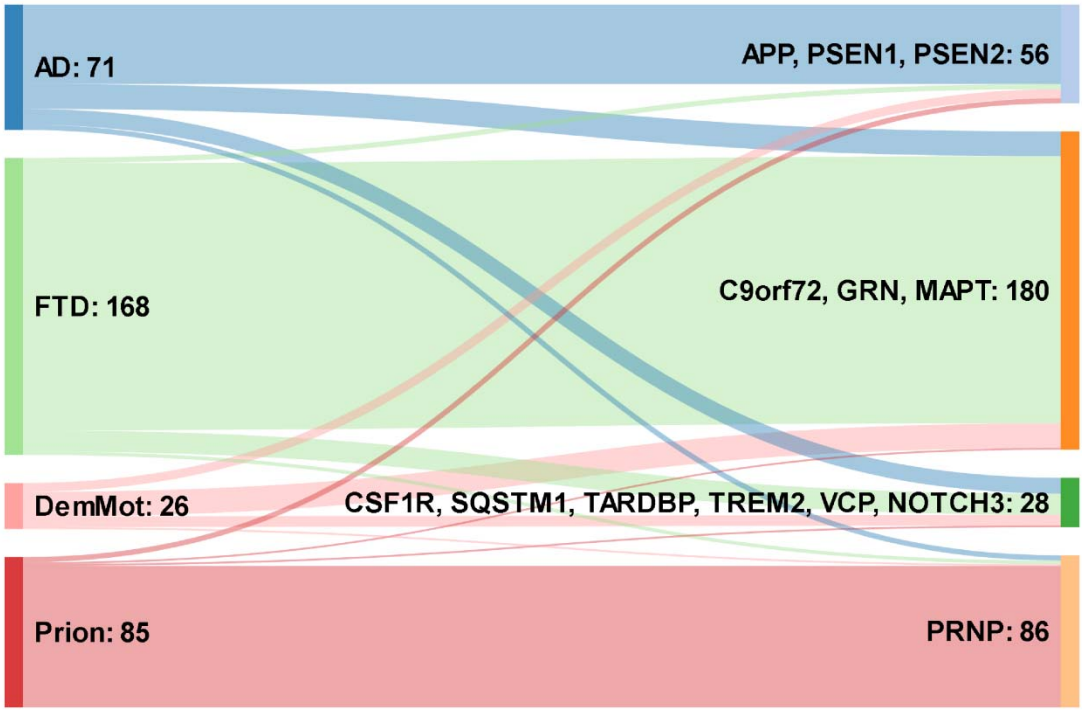

Figure III-3: Clinical syndrome is an imperfect predictor of the implicated gene

Numbers on the left refer to patients with clinical syndromes; numbers of the right refer to DVs in implicated genes.

CSF or post-mortem data was available for 198 patients to either confirm or contradict the clinical diagnosis; in 10% of cases the diagnosis was revised based on this additional information. More specifically, in 38 AD and 25 FTD cases, cerebrospinal fluid (CSF) was available to confirm the diagnosis; 37 AD cases (97.4%) and 19 FTD cases (76%) were found to have typical CSF results; an additional 15 FTD patients with logopenic aphasia all had CSF typical for AD. CSF results had been ordered by the treating clinician and taken into account, but may not always have been available when the referral diagnosis for genetic testing was made. 122 patients (4.5% of all patients; 33 AD, 40 FTD, 42 Prion and 7 DemMot cases) came to post-mortem examination; neuropathological data confirmed the clinical diagnosis in 29 AD, 27 FTD, 42 Prion and clarified four DemMot cases (87.9%, 67.5%, 100% and 57.1%, respectively). Neuropathological diagnoses not matching the clinical diagnosis included one case each of cerebellar degeneration, chronic traumatic encephalopathy, FTD and

leukodystrophy in AD; eight cases of AD, one case of globular glial tauopathy and one of progressive supranuclear palsy in FTD; as well as one case of AD, one case of FTD and one unclear case[186] in the DemMot cohort. Neuropathological data for a definitive diagnosis was available in only one case where an unexpected variant, a *C9orf72* expansion, was discovered; at post mortem examination, this patient who had been initially diagnosed with AD, was found to have *C9orf72* typical TDP deposition. These results demonstrate that overall the clinical diagnoses available when genetic testing was ordered were in the majority correct, although some unusual presentations will always be difficult to classify.

f. Age at onset associated with dementia genes

Neurodegenerative disease syndromes caused by DVs were found to have broadly similar ages at onset (AAO), independent of the gene in which these variants were identified (Figure III-4). An overall statistically significant difference in AAO by gene (ANOVA, p=0.003), was driven by the relatively late clinical onset in *GRN* and the relatively early onset of patients with DVs in *PSEN1*, *MAPT* and *PRNP*. Nevertheless, patients with *PRNP* DVs presented with a very wide range of AAO, stretching into old age (range 22 years to 79 years). DVs discovered in old age were not restricted to *PRNP*, and included *APP*, *PSEN1*, *C9orf72*, *GRN*, *MAPT*, *CHMP2B*, *CSF1R*, *TYROBP* and *VCP*.

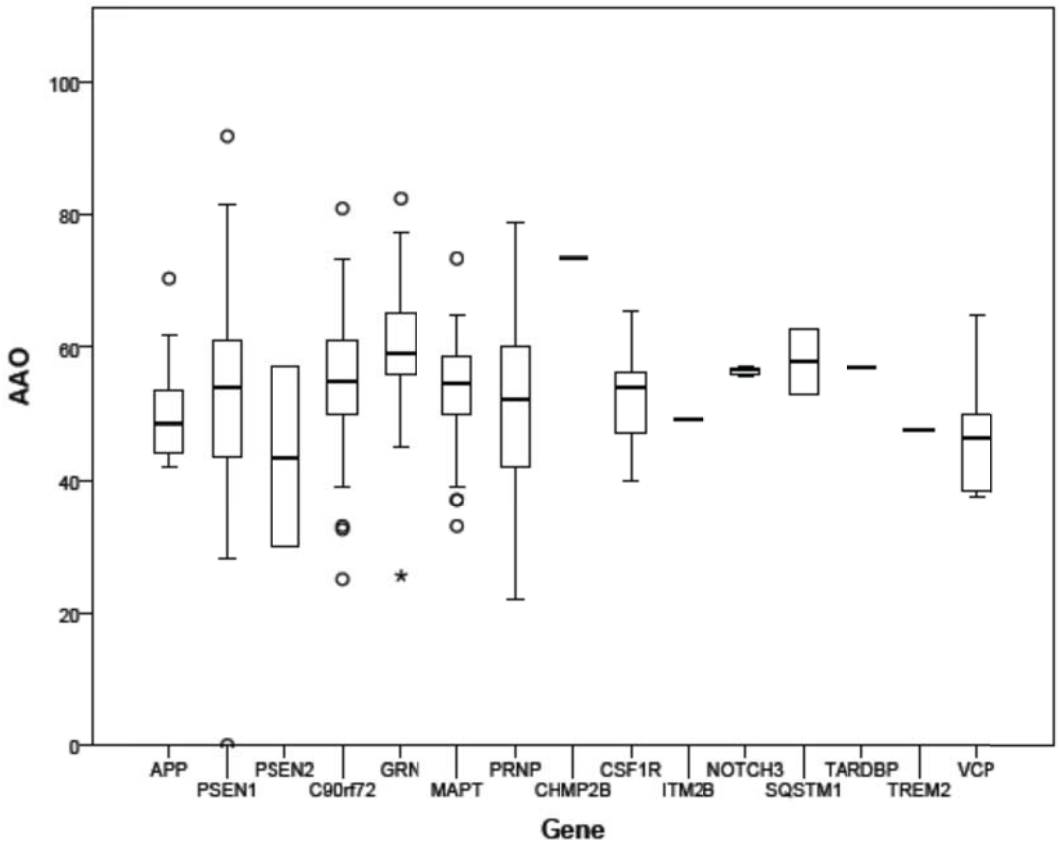

Figure III-4: Age at clinical onset (AAO) in patients with a DV, per gene

Box plots represent the interquartile range of age at onset (AAO) in each gene category with the median marked by a line; the whiskers represent the minimum and maximum AAO identified in each gene, except for values more than 1.5x the interquartile range from the median which are represented as outliers. Values between 1.5x and 3x the interquartile range a represented as points, while extreme values more than 3x the interquartile range from the median are represented as an asterix. In addition to the 294 cases from this dataset for whom age of onset was available, this graph contains data from an additional 54 patients with DVs who had been previously tested clinically at the MRC Prion Unit. These were added to increase sample size and more reliably report the range of AAO observed for DVs in each gene in the UK. We note that AAO for PSEN1 appears to be substantially earlier in the literature[60] which may relate to different methods of ascertainment of patients with a DV.

g. Risk factor analysis

i. ApoE

The distribution of ApoE genotypes across the cohorts is detailed in Table III-3; these data were only available for the cohorts run on the MRC Dementia Gene Panel. In the AD cohort, the ApoE genotypes 3 4 and 4 4 are significantly enriched (p=0.0003 and p=0.0011, respectively); for FTD patients, the ApoE genotype 4 4 (but no 3 4) was also

significantly enriched (p=0.00067). For three AD patients and five controls, the ApoE genotyping status could not be ascertained.

Table III-3: Distribution of ApoE genotypes across the patient and control cohorts

Cohort	ApoE Genotype						
	2 2	**2 3**	**3 3**	**2 4**	**3 4**	**4 4**	**Total**
AD	3 (0.31%)	68 (6.98%)	421 (43.22%)	25 (2.57%)	352 (36.14%)	105 (10.78%)	974
FTD	4 (0.58%)	70 (10.14%)	397 (57.54%)	20 (2.9%)	176 (25.51%)	23 (3.33%)	690
Prion	3 (1.01%)	38 (12.75%)	185 (62.08%)	5 (1.68%)	59 (19.8%)	8 (2.68%)	298
DemMot	7 (1.27%)	53 (9.6%)	326 (59.06%)	13 (2.36%)	124 (22.46%)	29 (5.25%)	552
Controls	3 (0.66%)	64 (14.16%)	280 (61.95%)	6 (1.33%)	86 (19.03%)	13 (2.88%)	452
Total	**20**	**293**	**1609**	**69**	**797**	**178**	**2966**

ii. Risk Factor Odds Ratios

In MRC Dementia Gene Panel dataset, variants that had previously been published as risk factors (either the variant itself or as a mutation pattern in a gene), and/or which appeared more frequent in cases than controls and observed on *GnomAD* more than 1 per 1000 alleles, were classified as risk factors. Because only DVs, or variants originally suspected of being a DV, from the clinically tested patients from the Department for Neurogenetics at NHNN were shared, it could not be included in the risk factor analysis.

Overall, in this sub-dataset, risk factor variants were more frequently found in cases than in controls (p=0.003). Variants classified as risk factors were most commonly identified in *TREM2* (5.9%), *SQSTM1* (5.61%), *MAPT* (3.9%), *TYROBP* (2.3%), *VCP* (1.4%), *PSEN2* (1.1%) and *GRN* (0.6%). These polymorphisms may be risk factors, but they require replication in order to confirm this effect, ideally in a large sample set of varied population in order to allow consideration of bias from population background and other effects.

iii. GRN missense and heterozygous TREM2 mutations

Rare *GRN* missense variants found in cases and controls were all classified as possibly pathogenic. Collectively, compared to controls (1.1%), *GRN* missense variants were

seen significantly in excess in patients (3.6%, p=0.004, OR 3.4 (1.4-8.4)) with AD (3.5%, p=0.006, OR=3.3) and FTD (3.7%, p=0.006, OR=3.4) but not DemMot (2.5%, p=0.1, OR=2.3) or prion disease (2.7%, p=0.2, OR=2.5). Compared to controls (3.7%), heterozygous *TREM2* missense variants were not significantly more common in dementia syndromes (4.9%; p=0.3), or in AD alone (5.9%, p=0.097; OR=1.6), albeit consistent with previously reported effect sizes[51]. While *GRN* haploinsufficiency is a well-established disease mechanism and any *GRN* missense mutations could interfere to a certain level with protein function, the effects of *TREM2* mutations are likely to be more complex.

h. Copy Number Variants

Using the DECoN[187] software, the NGS data from the MRC Dementia Gene Panel was assessed for any coverage anomalies. This highlighted copy number variants (CNVs) in two samples, one in *GRN* and one in *MAPT,* respectively. However, these variants could not be validated using Sanger sequencing and were classified as artefacts.

C. Predictive value of clinical patient information

a. Clinical syndrome

A logistical regression analysis with coverage, sex, ethnicity, healthcare setting, clinical syndrome, age at onset, and Goldman score as cofactors was carried out to assess the predictive value of clinical factors routinely used to select patients for genetic testing. Healthcare setting was included as a cofactor to determine whether DV is more or less likely to be identified in a sample originating from a dementia research centre or a primary referral centre Memory clinic; for this analysis, the control samples' origin site was excluded. Coverage, sex, ethnicity, healthcare setting, and prospective/retrospective recruitment did not influence the likelihood of a DV (p=0.97, p=0.33, 0.68, 0.61 and p=0.53, respectively, logistic regression); genders, ethnicities, sample coverage, sampling method and sample referral sites were therefore combined in further analyses. Compared to controls, AD (logistic regression, p=0.006; Odds Ratio (OR): 7.46, 95% Confidence Interval (CI) (1.77, 31.49)), FTD ($p=2.0x10^{-6}$; OR: 33.58, 95% CI (7.95, 141.81)), Prion patients ($p=2.24x10^{-9}$; OR: 92.54, 95% CI (20.98, 408.184)) and DemMot patients (p=0.042; Odds Ratio: 4.7, 95% CI (1.06, 20.87)) were significantly

more likely to carry a DV in order of declining frequency Prion>FTD>AD>DemMot (Table III-2).

b. *Age at onset*

AAO was a very strong predictor of finding a DV ($p=3.8x10^{-9}$, logistic regression, Figure III-5). Risk was high from early adulthood through to middle age and steadily declined into old age without clear change in risk at the traditional boundary of early and late-onset disease, age 65.

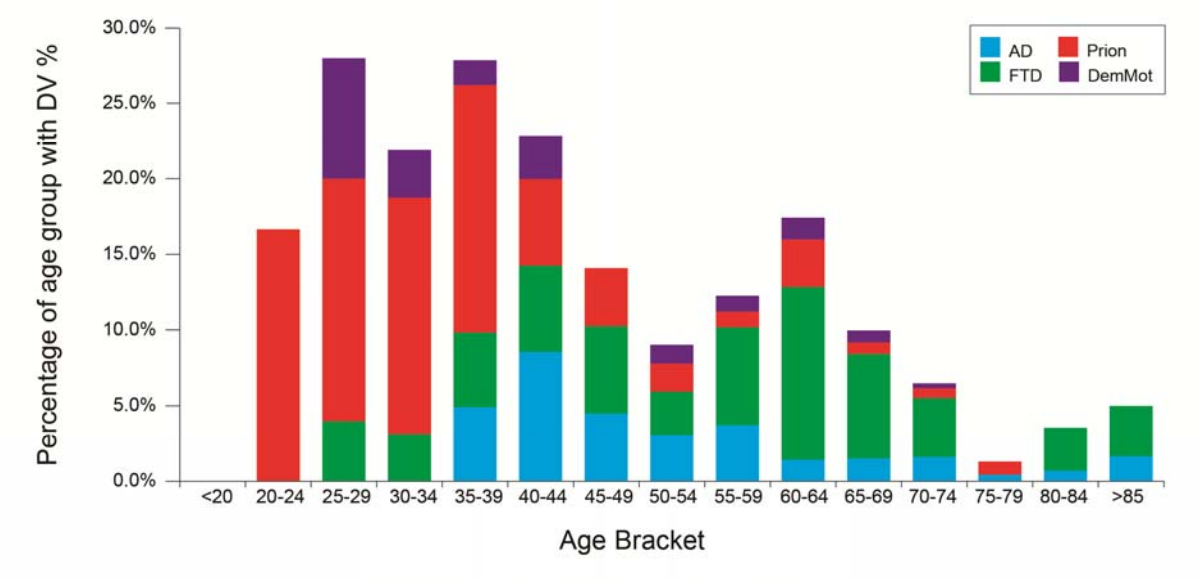

Figure III-5: Proportion of patients with a DV per age group (%)

The distribution is skewed towards the younger ages of onset, but we discovered many patients with DVs associated with elderly ages of onset, particularly in the presence of a family history.

c. *Family history stratified as the Goldman score*

Family history was also highly predictive of a DV ($p=4.6x10^{-38}$, logistic regression, Figure III-6, Figure III-1). To determine whether selection criteria for patients with late-onset dementia would obey a different set of rules, a separate logistical regression analysis was carried out including only patients who were 65 years or older when their symptoms first manifested themselves; again the control samples were excluded. The predictive value of a relevant family history was also strong in late-onset dementia, in which circumstance GS remained highly predictive of identifying a DV ($p=2.3x10^{-4}$, logistic regression), but age at onset no longer had a significant effect (p=0.452).

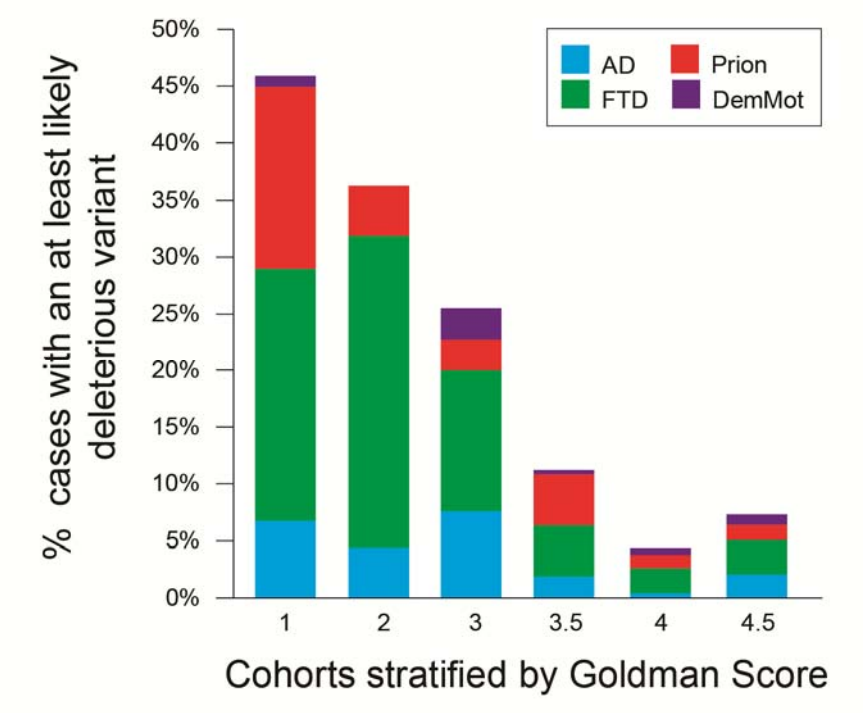

Figure III-6: Family history is a strong predictor for the identification of a DV

Stratifying cases by Goldman Score reveals its strong predictive value in identifying cases with a DV; however, deleterious or likely deleterious variants are found in clinically-relevant proportions of cases with no (GS4) or a censored (GS4.5) family history.

d. Summary and recommendations for testing

Based on these results, the combined effects of AAO, clinical syndrome and family history were combined to formulate recommendations for use of dementia gene panels (Figure III-7). Since in AD and FTD, it was difficult to predict which dementia gene a DV would be found in, gene panel diagnostics was particularly useful. Depending on clinical syndrome, AAO, and family history, yield of clinically relevant mutations was high (> 10%), medium (5–10%) or low (< 5%), although even the low (<5%) yield may be considered worthwhile by many clinicians depending on clinical scenarios and the wishes of patients and at-risk individuals. FTD subtype (behavioural variant, progressive non-fluent aphasia, semantic dementia, etc.) may also have an effect on yield and most appropriate testing, but was not the focus of this study and would require more research to stratify. Suspected prion disease patients should first be tested for *PRNP*; if this test is negative, patients should be tested as an AD

syndrome. The yield for dementia-motor syndromes was low yield on dementia panel testing (<5% all subgroups); possibly because too little information was available to classify novel and uncertain variants as DVs, or because they are caused by mutations in different genes. In these cases, investigations may follow the stepwise recommendations for HD-like syndromes, which are often caused by expansion disorders not well ascertained by gene panel diagnostics[13] and then proceed to WES or WGS.

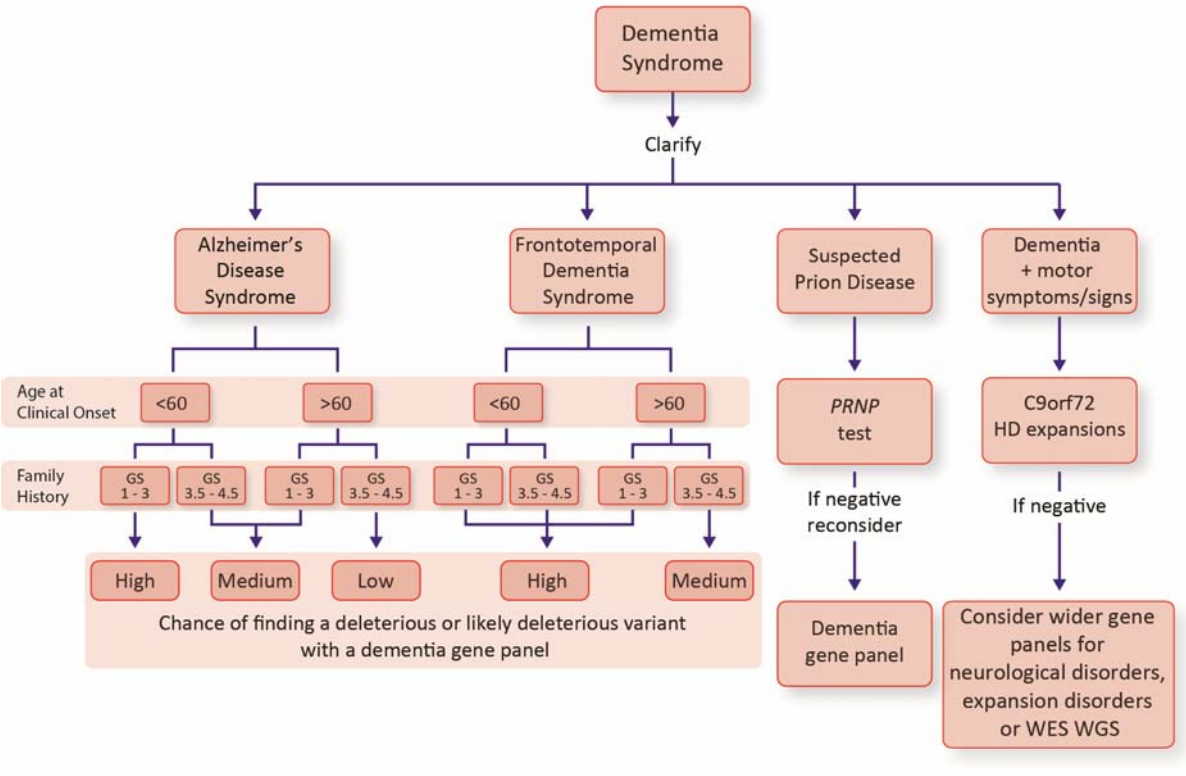

Figure III-7: Algorithm for patient selection for genetic testing.

AD and FTD cases benefitted most from use of a dementia panel since they had a high rate of DVs in genes not typically associated with the syndrome. Expected yields of clinically relevant mutations (high (> 10%), medium (5–10%) or low (< 5%)) can be stratified by clinical syndrome, AAO, and family history. Suspected prion disease patients should first be tested for *PRNP*; if negative, proceed as for an AD syndrome. The yield for dementia-motor syndromes was low on dementia panel testing (<5%); in these cases, investigations may follow the stepwise recommendations for HD-like syndromes[13] and then WES or WGS. *Figure previously published in Koriath et al., 2018*[24]

D. Penetrance

a. Prevalence calculations in EOAD and EOFTD

In their 2016 paper[28], Minikel *et al.* explain that, considering the low incidence and prevalence of prion disease, the number of variants that are reported pathogenic appears very high; the authors subsequently demonstrate that some reportedly pathogenic variants should be considered likely benign or incompletely penetrant based on their frequency in online population databases. Their methodology calculates the maximum expected cases in a population that is neither enriched nor depleted for the variants in question, multiplying the incidence, the proportion of genetic cases, the life expectancy of genetic cases and the number of individuals in the given population. Given the high number of published deleterious variants in early-onset AD (EOAD) and early-onset FTD (EOFTD), the same could be suspected for EOAD and EOFTD.

i. Prevalence of early-onset Alzheimer's disease

Lambert *et al.*[45] offer an overview of the prevalence of EOAD in the nine studies reviewed there, prevalences for EOAD range from 10.6 (age at onset 20-64 years) to 200 (age at onset 55-64 years) per 100,000. Several groups determine prevalences between 20 and 40 for the age at onset range 45-64 years. Some variation in estimates may be a reflection of different methodologies and age ranges in each of these studies as the prevalence of AD varies greatly with age. A prevalence estimate for EOAD of 30 per 100,000 reflects the ranges set out in these studies and the different numbers of patients included in the studies.

ii. Prevalence of early-onset FTD

Luukkainen et al. present a summary report of the prevalence of early-onset EOFTD as well as their own findings from northern Finland[188], while Coyle-Gilchrist et al. looked at the prevalence of FTD in two UK counties[73]. While methodology and analysed age groups varied between the different studies, the reported prevalence for EOFTD ranged from 2.7 to 35 per 100,000. The prevalence of 18 per 100,000 cases estimated here for EOFTD reflected of the ranges and numbers of patients included in the studies.

b. Lifetime risk of Alzheimer's disease and frontotemporal dementia

i. Alzheimer's disease

Based on the above estimates of prevalence, the lifetime risk of developing either EOAD or EOFTD (age of onset between 30 and 65 years of age) can be calculated. Given the prevalence EOAD of approx. 30/100,000 and the average disease duration for EOAD of approx. 13 years[87], new cases per year are estimated at 2.31/100,000. Based on the population aged 30 to 64 of England and Wales from the 2015 dataset (22,435,600 people), new cases per year are estimated at 518. New cases through a 35 year window for this population of England and Wales are expected at 18,121. For the population of England and Wales in 2015 (57,885,400 people), this would amount to a ratio of 0.00031 or 1 in 3,194 people who will develop EOAD in their lifetime.

ii. Frontotemporal dementia

For EOFTD, the prevalence is approx. 18/100,000 and disease duration is approx. 8 years on average[72]. New cases per year are therefore estimated to occur at a rate of 2.25/100000, or 504 cases for the population aged 30 to 64 of England and Wales from the 2015 dataset (22,435,600 people). Over 35 years, this would amount to 17,668 cases of EOFTD, a rate of 0.00034 for the total population of England and Wales (57,885,400 people) and mean that 1 in 3,276 people will develop EOFTD in their lifetime.

c. Summary and cross-check with national mortality statistics

The Office for National Statistics reports 706 deaths between the ages 30 to 69 years, which are attributed to AD, and 153 deaths between the ages 30 to 69 years, which are attributed to FTD. I chose the age range 30 to 69 years in order to exclude deaths most likely not caused by early-onset dementia. Given the average disease duration of late-onset Alzheimer's disease (LOAD) of 7 years[87], I excluded deaths over the age of 70. That said, deaths attributed to AD increase exponentially in the age-range 65-69 compared to younger ages, suggesting that some of them may be due to either rapidly-progressive LOAD or other causes expediting death in the presence of LOAD. This would explain the higher than expected number of reported deaths due to AD between the ages 30 to 69 years. While deaths attributed to AD are slightly higher

than expected, deaths attributed to FTD between the ages 30 to 69 years are slightly fewer than expected. Overall, these statistics are broadly consistent with the prevalence estimates calculated above.

d. Penetrance of genetic variants linked to EOAD and EOFTD

i. Published reportedly pathogenic variants

Based on their frequency in this dataset and on *GnomAD*, a number of published reportedly pathogenic variants in *APP*, *PSEN1* and *PSEN2* seem to be incompatible with being highly penetrant pathogenic mutations (Table III-4). At the time of analysis, on the Molgen database and on the mutation database of Alzforum (Alzgene), 302 variants in *APP*, *PSEN1* and *PSEN2* were listed as deleterious and a further 21 variants were reported as having unclear pathogenicity based on case reports. In *GRN*, *MAPT* and *VCP*, 153 variants were listed as deleterious and a further 64 variants were reported as having unclear pathogenicity based on case reports. However, as was to be expected, not all published reportedly pathogenic variants in *APP*, *PSEN1*, *PSEN2*, *GRN*, *MAPT* and *VCP*, were found in the present dataset. Indeed, of the reportedly pathogenic variants which are also present in the *GnomAD* dataset, only *PSEN1* Arg269His, *PSEN1* Pro264Leu, *GRN* Arg493Ter, *GRN* Leu469Phe, *GRN* Thr251Ser and *GRN* Glu287Asp were also observed in the present case series. To permit an assessment of the other reportedly pathogenic variants, I envisaged a scenario where they would have been observed just once in this dataset and performed calculations on this basis. While many of the reportedly deleterious variants are not observed at all or at very low frequencies in population databases, a small number of variants were detected repeatedly, calling into question the extent of their pathogenicity and penetrance. More specifically, *APP* Ala713Thr, *PSEN1* Ser170Phe, *PSEN1* Arg352Cys, *PSEN1* Arg358Gln, *PSEN1* Ser365Ala, *PSEN1* Val412Ile, *PSEN2* Thr430Met, *PSEN2* Ala237Val, and *PSEN2* Val214Leu are not enriched in this EOAD cohort, despite having been reported as Mendelian pathogenic variants. Penetrance for these variants was calculated to be 2% or less, with upper limits of the 95% confidence interval of less than 20%, frequently less than 10%. Indeed, penetrance for these variants is comparable to *PSEN1* Asn32Asn and *PSEN1* Asp40del, which are reported as uncertain

because they were reported in only one patient each. This discrepancy highlights the importance of checking population frequency of any variant intimated to be disease causing and to assess how likely it is to have a deleterious effect.

Table III-4: Reported pathogenic variants in AD genes APP, PSEN1 and PSEN2

Listed are reportedly pathogenic variants, which were also found on *GnomAD*, with their respective frequency in the population and likely penetrance. *PSEN1* Arg269His, *PSEN1* Pro264Leu and *APP* Ala713Thr were identified in our tested sample dataset; all three are reportedly pathogenic. While *PSEN1* Arg269His and *PSEN1* Pro264Leu are compatible with fully penetrant pathogenic mutations, *APP* Ala713Thr is likely to have a very limited penetrance based on its frequency in the general population. For all other variants, penetrance calculations were based on a hypothetical count of one in our sample (i.e. calculations are biased towards higher penetrance). 322 variants in *APP*, *PSEN1* and *PSEN2* were listed as pathogenic on the Molgen AD/FTD database) and/or Alzforum; variants listed as "pathogenic nature unclear" on Molgen were also included and marked as such in the column "Reported". The allele count and frequency of the protein changes induced by these variants on *GnomAD* is also included, except for copy number variants (CNVs).

Gene	Mutation	Exon	Domain	GnomAD Allele Count	GnomAD Allele Freq.	Source	Reported	Penetrance (95% CI)
Most likely reduced penetrance								
PSEN1	Val94Met	EX4	TM-I	2	0.0008%	Molgen	deleterious	5.2 (0.4%, 64.4%)
PSEN1	Thr354Ile	EX10	HL-VI b	2	0.0008%	Molgen	unclear	5.2 (0.4%, 64.4%)
PSEN2	Gln228Leu	EX7	TM-V	2	0.0008%	Molgen	deleterious	5.2 (0.4%, 64.4%)
PSEN1	Arg108Gln	EX4	HL-I	2	0.0008%	Molgen	deleterious	5.2 (0.4%, 64.3%)
PSEN2	Val148Ile	EX5	TM-II	2	0.0008%	Molgen	deleterious	5.2 (0.4%, 64%)
PSEN1	Glu123Lys	EX5	HL-I	3	0.0011%	Molgen	deleterious	3.9 (0.4%, 41.3%)
PSEN2	Lys161Arg	EX5		3	0.0011%	Alzgene	deleterious	3.8 (0.4%, 40.2%)
PSEN1	Gly206Ala	EX7	TM-IV	3	0.0012%	Molgen	deleterious	3.4 (0.3%, 34.8%)
PSEN2	Leu238Pro	EX7	TM-V	3	0.0012%	Molgen	deleterious	3.4 (0.3%, 34.8%)
PSEN1	Ala79Val	EX4	N-Term	4	0.0014%	Molgen	deleterious	2.9 (0.3%, 26.9%)
PSEN1	Arg352dup	EX10	HL-VI b	4	0.0014%	Molgen	unclear	2.9 (0.3%, 26.9%)
PSEN1	Ser365Ala	EX10	HL-VI b	5	0.0020%	Molgen	deleterious	2.1 (0.3%, 16.8%)
PSEN1	Ser170Phe	EX6	TM-III	8	0.0028%	Molgen	deleterious	1.4 (0.2%, 10.4%)
Most likely fully penetrant (95%CI >100% penetrance)								
PSEN1	Arg269His	EX8	HL-VI a	1	0.0004%	Molgen	deleterious	
PSEN1	Pro264Leu	EX8	HL-VI a	1	0.0004%	Molgen	deleterious	
PSEN1	Ile202Phe	EX7	TM-IV	1	0.0004%	Molgen	deleterious	
PSEN1	His214Tyr	EX7	HL-IV	1	0.0004%	Molgen	deleterious	
PSEN1	Leu219Phe	EX7	HL-IV	1	0.0004%	Molgen	deleterious	
PSEN1	Ile408Thr	EX11	TM-VIII	1	0.0004%	Molgen	deleterious	
PSEN2	Ala85Val	EX4	N-Term	1	0.0004%	Molgen	deleterious	
Most likely benign or only small increase in risk (95%CI <10% penetrance)								
APP	Ala713Thr	EX17	TM-I	26	0.0092%	Molgen	deleterious	
PSEN1	Val412Ile	EX11	TM-VIII	1	0.0033%	Molgen	deleterious	
PSEN2	Thr430Met	EX12		9	0.0036%	Alzgene	deleterious	
PSEN1	Arg352Cys	EX10	HL-VI b	12	0.0043%	Molgen	deleterious	
PSEN1	Arg358Gln	EX10	HL-VI b	11	0.0044%	Molgen	deleterious	
PSEN2	Ala237Val	EX7	TM-V	14	0.0056%	Molgen	deleterious	
PSEN2	Val214Leu	EX7		65	0.0230%	Alzgene	deleterious	
PSEN1	Asn32Asn	EX4	N-Term	12	0.0043%	Molgen	unclear	
PSEN1	Asp40del	EX4	N-Term	39	0.0138%	Molgen	unclear	

In the genes *GRN*, *MAPT* and *VCP* associated with FTD, most reportedly pathogenic variants appear to be very rare in the population (Table III-5); indeed *MAPT* Arg5His is

the only variant listed as pathogenic with an estimated penetrance of less than 2% (95% CI: 0.3%, 9.3%). However, a large number of variants in both *GRN* and *MAPT*, which are relatively frequent in the population have been reported as associated with FTD; however, for these, the level of evidence is classified as unclear because of the number or quality of the reports. Analysis of these variants and their frequency in the population (*GRN* Asp33Glu, *GRN* Pro34Pro, *GRN* Gly35Arg, *GRN* Gly70Ser, *GRN* Val77Ile, *GRN* Arg110Gln, *GRN* Thr138Thr, *GRN* Cys139Arg, *GRN* Arg212Trp, *GRN* Pro233Gln, *GRN* Glu287Asp, *GRN* Arg298His, *GRN* Pro392Pro, *GRN* Arg432Cys, *GRN* His447His, *GRN* Cys495Cys, *GRN* Val514Met, *GRN* Val519Met, *GRN* Cys521Tyr; *MAPT* Ala41Thr, *MAPT* Gly86Ser, *MAPT* Ala297Val, *MAPT* Ser318Leu, *MAPT* Ser427Phe, *MAPT* Arg448Ter) revealed that they are unlikely to be highly penetrant with a calculated penetrance of less than 3% and an upper limit of the confidence interval of less than 20% penetrance.

Table III-5: Reportedly pathogenic variants in FTD genes GRN, MAPT and VCP

Listed are reportedly pathogenic variants, which were also found on *GnomAD*, with their respective frequency in the population and likely penetrance, as well as reportedly unclear variants with an upper penetrance 95%CI >50%. *GRN* Arg493Ter, *GRN* Arg110Ter, *GRN* Leu469Phe and *GRN* Thr251Ser were identified in our dataset; the first two are reportedly pathogenic. *GRN* Arg493Ter, *GRN* Arg110Ter, *GRN* Leu469Phe and *GRN* Thr251Ser are all compatible with being fully penetrant pathogenic mutations based on their frequency in cases and in the population. For all other variants, penetrance calculations were based on a hypothetical count of one in our dataset. 217 variants in *GRN*, *MAPT* and *VCP* were listed as pathogenic on the Molgen AD/FTD database and/or Alzforum; variants listed as "unclear pathogenicity" and reported in cases were also included and marked as such in the column "Reported". The allele count and frequency of the protein changes induced by these variants on *GnomAD* is also included, except for copy number variants (CNVs).

Gene	Mutation	Exon	Domain	GnomAD Allele Count	GnomAD Allele Freq.	Source	Reported	Penetrance (95% CI)
MAPT	Gly55Arg	Ex2		3	0.001%	Alzgene	deleterious	7.6% (0.7%, 81.3%)
VCP	Arg191Gln	EX5	Linker 1	4	0.002%	Molgen	deleterious	5.1% (0.6%, 45.5%)
GRN	Cys222Tyr	EX7	GranB	3	0.001%	Molgen	unclear	6.8% (0.7%, 68.5%)
GRN	Arg535Ter	EX12	GranE	3	0.001%	Molgen	unclear	6.8% (0.7%, 68.5%)
MAPT	Arg5Cys	EX1		3	0.001%	Alzgene	unclear	6.8% (0.7%, 68.5%)
GRN	Arg564Cys	EX13	GranE	3	0.001%	Molgen	unclear	6.7% (0.7%, 67.7%)
VCP	Arg95His	EX3	CDC48	4	0.001%	Molgen	unclear	5.7% (0.6%, 53%)
Most likely fully penetrant (95%CI >100% penetrance)								
GRN	Arg493Ter	EX12	GranD	1	0.000%	Molgen	deleterious	
GRN	Arg110Ter	EX4	GranG	1	0.000%	Molgen	deleterious	
GRN	Leu469Phe	EX11	GranD	1	0.000%	Molgen	unclear	
GRN	Thr251Ser	EX8	GranB	3	0.001%	Molgen	unclear	
GRN	Pro127fs	EX5	GranF	1	0.000%	Molgen	deleterious	
VCP	Arg159Cys	EX5	CDC48	1	0.000%	Molgen	deleterious	
GRN	Cys253Ter	EX8	GranB	1	0.000%	Molgen	deleterious	
GRN	Thr382fs	EX10	GranC	1	0.000%	Molgen	deleterious	
GRN	Thr382fs	EX10	GranC	1	0.000%	Molgen	deleterious	
GRN	Gln130fs	EX5	GranF	2	0.001%	Molgen	deleterious	
GRN	Gln130fs	EX5	GranF	2	0.001%	Molgen	deleterious	
VCP	Arg159His	EX5	CDC48	2	0.001%	Molgen	deleterious	
VCP	Arg95Cys	EX3	CDC48	2	0.001%	Molgen	deleterious	
GRN	Cys105Arg	EX4	GranG	1	0.000%	Molgen	unclear	
GRN	Arg547Cys	EX12	GranE	1	0.000%	Molgen	unclear	
MAPT	Val75Ala	EX3	-	1	0.000%	Molgen	unclear	
GRN	Ala276Val	EX8	InterBA	2	0.001%	Molgen	unclear	
GRN	Pro451Leu	EX11	GranD	2	0.001%	Molgen	unclear	
Most likely benign or only small increase in risk (95%CI <10% penetrance)								
MAPT	Arg5His	EX1	N-Term	14	0.01%	Molgen	deleterious	

ii. Variants detected in the present dataset

Following the method set out by Minikel *et al.*[28] as explained above, all variants in *APP*, *PSEN1*, *PSEN2*, *GRN*, *MAPT* and *VCP* that were observed both in this dataset and on

GnomAD were assessed for their likely penetrance. For calculation purposes, the EOAD and the EOFTD cohorts included 768 and 447 cases, respectively. Analysis was limited to these subsets because it was not possible to reliably estimate the prevalence of genetically-determined dementia in old age. As set out above, the baseline lifetime risk for EOAD and EOFTD was assumed to be 0.031% and 0.034%, respectively, and compared against population data from the *GnomAD* database, which at the time of analysis contained information from 141,352 individuals[124]. These variants are listed in Table III-6.

Based on their relative frequency in cases and the general population, most of the variants classified as deleterious or likely deleterious in this dataset appear to be highly penetrant pathogenic mutations, with two exceptions: *APP* Ala713Thr was observed once and despite being reportedly pathogenic, its population frequency suggests that it may only have a marginal deleterious effect with a penetrance of 0.4% (95% CI 0.1%, 2.4%). Secondly, *PSEN2* Leu238Phe appears to be most likely partially penetrant with a likely penetrance 3.4% and an upper confidence interval of approx. 35%. On the basis of these data and calculations, it seems unlikely that a variant that exceeds a frequency of approx. 0.0016% in the general population, i.e. *GnomAD*, should be a fully penetrant cause of a rare disease like EOAD or EOFTD. Pending further evidence, it follows that of the "possibly deleterious" variants identified in this dataset, for which only limited evidence of pathogenicity is available, only *PSEN1* Arg41Serfs, *PSEN1* Arg41Thrfs, *PSEN1* Arg42Profs, *PSEN1* Asn39del, *PSEN1* Asn58Lys, *PSEN1* His46Tyr, *PSEN1* Asn39Tyr, *PSEN1* Gln15His, *PSEN1* Pro303Leu, *PSEN1* Ile148Val, *PSEN2* Arg62Cys, *PSEN2* Asp431Glu, *PSEN2* Tyr195Cys, *GRN* Leu469Phe, *GRN* Thr251Ser, *GRN* Arg579Cys and *MAPT* Arg194His should be suspected of potentially being fully penetrant disease-causing mutations. Based on their population frequency, the other "possibly deleterious" variants are more likely to be rare neutral polymorphisms or potentially slightly increase the risk of disease. The same reasoning can also be applied to the variants of uncertain significance, with the population frequency of *APP* Gly657Arg, *APP* Arg16Gln and *MAPT* Asn167Ser being all but incompatible with being highly penetrant. All variants classified as risk factors in this dataset appear very

slightly enriched in cases compared to the population (*GnomAD* [124]), but none could be fully penetrant, as would be expected.

Table III-6: Penetrance of DVs found in this dataset in APP, PSEN1, PSEN2, GRN, MAPT and VCP

The EOAD and the EOFTD cohorts included 757 and 421 cases, respectively, and the baseline lifetime risk for EOAD and EOFTD was assumed to be 0.031% and 0.034%, respectively. Variants were compared against population data from the *GnomAD* database, which contains information from 141,352 individuals.

Gene	Variant	Classification	Cohort	Case Allele Count	Case Allele Freq.	GnomAD Allele Count	GnomAD Allele Freq.	Penetrance and CI
MAPT	Gly389Arg	Deleterious	EOFTD	2	0.0048	4	0.000016	10.16% (1.6, 63.5%)
PSEN1	Ile227Val	Likely Deleterious	EOAD	1	0.0013	4	0.000014	2.88% (0.3, 26.9%)
GRN	Thr251Ser	Possible	EOAD + EOFTD	2	0.0016	3	0.000011	10.09% (1.3, 76.2%)
PSEN2	Tyr195Cys	Possible	EOFTD	1	0.0024	4	0.000016	5.08% (0.6, 45.1%)
PSEN1	Arg42Leu	Possible	EOAD	2	0.0026	5	0.00002	4.14% (0.7, 23.7%)
MAPT	Gly415Ser	Possible	EOFTD	1	0.0024	6	0.000021	3.81% (0.5, 30.2%)
PSEN2	Leu135Arg	Possible	EOFTD	1	0.0024	1	0.000033	2.44% (0.4, 16.6%)
GRN	Thr268Met	Possible	EOAD	1	0.0013	5	0.000018	2.31% (0.3, 19.7%)
FUS	Gly225Ser	Possible	EOAD	1	0.0013	5	0.000019	2.12% (0.3, 17.5%)
GRN	Pro458Leu	Possible	EOFTD	1	0.0024	12	0.000043	1.89% (0.3, 12%)
PSEN2	Leu225Pro	Risk factor	EOAD	1	0.0013	2	0.000008	5.17% (0.4, 64.3%)
MAPT	c.*16G>A	Uncertain	EOAD	1	0.0013	2	0.000008	5.15% (0.4, 64%)
MAPT	Asn167Ser	Uncertain	EOFTD	1	0.0024	9	0.000032	2.54% (0.4, 17.5%)
APP	Gly657Arg	Uncertain	EOFTD	1	0.0024	1	0.000033	2.44% (0.4, 16.6%)
APP	Arg16Gln	Uncertain	EOAD	1	0.0013	5	0.00003	1.36% (0.2, 9.5%)
MAPT	Gly201Ser	Uncertain	EOAD	1	0.0013	9	0.000033	1.25% (0.2, 8.6%)
Most likely fully penetrant (95%CI >100% penetrance)								
GRN	Arg493Ter	Deleterious	EOFTD	4	0.0095	1	0.000004	
PSEN1	Arg269His	Deleterious	EOAD	3	0.004	1	0.000004	
MAPT	Arg406Trp	Deleterious	EOFTD	6	0.0143	4	0.000016	
MAPT	Pro301Leu	Deleterious	EOAD + EOFTD	2	0.0017	1	0.000005	
GRN	Arg110Ter	Deleterious	EOFTD	1	0.0024	1	0.000004	
MAPT	Lys257Thr	Deleterious	EOFTD	1	0.0024	1	0.000004	
PSEN1	Pro264Leu	Deleterious	EOAD	1	0.0013	1	0.000004	
PSEN1	Ser132Ala	Likely Deleterious	EOAD + EOFTD	2	0.0017	1	0.000004	
PSEN1	Gln15His	Possible	EOFTD	1	0.0024	1	0.000004	

Gene	Variant	Classification	Cohort	Case Allele Count	Case Allele Freq.	GnomAD Allele Count	GnomAD Allele Freq.	Penetrance and CI
PSEN1	Pro303Leu	Possible	EOFTD	1	0.0024	1	0.000004	
PSEN2	Arg62Cys	Possible	EOFTD	1	0.0024	1	0.000004	
GRN	Leu469Phe	Possible	EOFTD	1	0.0024	1	0.000004	
PSEN1	Asn39Tyr	Possible	EOAD	1	0.0013	1	0.000004	
MAPT	Arg194His	Possible	EOFTD	1	0.0024	2	0.000008	
PSEN2	Arg163Cys	Risk factor	EOFTD	2	0.0048	1	0.000004	
VCP	Val133Ile	Uncertain	EOFTD	1	0.0024	1	0.000004	
FUS	Pro459Leu	Uncertain	EOAD	1	0.0013	1	0.000004	
MAPT	c.*19C>A	Uncertain	EOFTD	1	0.0024	2	0.000008	
Most likely benign or only small increase in risk (95%CI <10% penetrance)								
APP	Ala713Thr	Likely deleterious	EOAD	1	0.0013	26	0.000092	
GRN	Glu287Asp	Possible	EOAD	1	0.0013	10	0.000035	
GRN	Ala582Thr	Possible	EOFTD	1	0.0024	21	0.000075	
GRN	Arg110Gln	Possible	EOFTD	1	0.0024	22	0.000087	
GRN	Val514Met	Possible	EOAD	1	0.0013	13	0.000046	
FUS	Pro18Ser	Possible	EOFTD	1	0.0024	33	0.000117	
GRN	Arg535Gln	Possible	EOAD	1	0.0013	21	0.000083	
FUS	Pro431Leu	Possible	EOAD	1	0.0013	31	0.000111	
GRN	Asn39Tyr	Possible	EOAD	1	0.0013	47	0.000168	
GRN	His340Leu	Possible	EOAD	1	0.0013	24	0.000182	
GRN	c.-56T>G	Possible	EOAD	1	0.0013	24	0.000793	
PSEN2	Ser130Leu	Risk factor	EOAD + EOFTD	6	0.0051	177	0.000627	
VCP	Ile27Val	Risk factor	EOAD + EOFTD	4	0.0034	183	0.000647	
MAPT	Ala152Thr	Risk factor	EOAD + EOFTD	8	0.0068	403	0.00143	
PSEN2	Met174Val	Risk factor	EOAD	4	0.0053	166	0.000587	
MAPT	Ala152Thr	Risk factor	EOAD + EOFTD	2	0.0017	403	0.00143	
PSEN2	Arg62His	Risk factor	EOAD + EOFTD	7	0.0059	2642	0.00936	
APP	Arg16Gln	Uncertain	EOAD	1	0.0013	5	0.00003	
MAPT	Gly201Ser	Uncertain	EOAD	1	0.0013	9	0.000033	

2 Chapter 2: Apolipoprotein 4 and its effect in FTD

A. Background

Apolipoprotein E (ApoE) genotypes are the strongest known common risk factors for Alzheimer's disease (AD) [148,149]; more specifically, ApoE4 has been found to increase the risk of AD and lower the AAO, while ApoE2 was found to decrease the risk of AD and delay AAO[149,189]. Consensus about the mechanism(s) of risk conferred ApoE4 in AD remains elusive, but multiple lines of evidence suggest that ApoE4 risk in AD principally relates to enhanced and accelerated cerebral Aβ pathology[190]. More precisely, ApoE4 seems to accelerate the early seeding of amyloid pathology, most likely by decreasing Aβ clearance and enhancing Aβ aggregation[191].

In 2017, Shi *et al.* described how ApoE4 increases the burden of cerebral tau pathology, neuroinflammation and brain atrophy in a P301S mouse model of FTD and *in vitro*; they also demonstrated that in patients with a primary tauopathy, ApoE4 was associated with more severe regional neurodegeneration, and that ApoE4 AD patients with amyloid-β (Aβ) pathology showed faster disease progression[147]. The conclusion that ApoE4 worsens tauopathy independently of Aβ raises the question whether ApoE genotype also influences risk or clinical phenotype in patients with primary tauopathies. Similarly to what has been established for AD, ApoE2 has been reported to confer a protective effect on FTD patients, while ApoE4 has been shown to increase risk in several studies[192-194], although another study could not replicate this effect[195]. By contrast, a study in a tau transgenic mouse model recently demonstrated a deleterious effect of ApoE2[196]. However, association of ApoE with frontotemporal dementia (FTD) remains uncertain; it is complicated by two factors in clinically ascertained cohorts: the misdiagnosis of frontal variant AD for FTD, and the fact that FTD comprises tau, TDP-43 and other pathologies[27].

Prompted by Shi *et al.*[147], I therefore reanalysed the data gathered for the MRC Dementia Gene panel; I compared 704 FTD patients and 452 healthy elderly controls using SPSS 25 to test the hypothesis that ApoE genotypes have a modifying effect on

clinical phenotype in those with or expected to have tau pathology defined by a highly penetrant *MAPT* gene mutation or established by neuropathological examination.

B. The UCL FTD cohort

a. Clinical and genetic characteristics

704 patients from the UCL FTD cohort, for who *ApoE* genotype information was available, were included in this analysis and compared to the 456 healthy controls; all of them were tested for causative gene mutations and described in Results Chapter 1. Seven patients from the original MRC Dementia Gene Panel FTD cohort who were found to carry causative mutations in genes not typically associated with FTD were excluded from the present analyses.

For the FTD patients, average age at onset (AAO) was 58.1 years old, 351 (49.9%) were male; AAO was missing for 55 patients and for 25 patients sex information was not available. Dementia and personality change were the predominant symptoms in this series (54.7% - 33.3% "behavioural type FTD", 21.4% "FTD", 1.1% "Dementia"), followed by aphasia (31.8% - 16.9% progressive non-fluent aphasia, 13.1% semantic dementia, 1.8% logopenic progressive aphasia) and additional motor or muscular symptoms (11.5% - 6.0% corticobasal degeneration, 2.4% progressive supranuclear palsy, 3.7% FTD with motor neuron disease, 0.3% inclusion body myositis with Paget's disease and FTD).

Genetic diagnoses were available for 166 FTD patients included in this analysis. This included 59 patients with *C9orf72* expansions, 53 patients with *MAPT* mutations, 44 patients with *GRN* mutations, five cases with *VCP* mutations, two cases with *TBK1* mutations, as well as one case each with a mutation in *CHMP2B, TYROBP* and *TARDBP* mutation. Of these, 7 patients carried concurrent mutations; these included 5 patients with *C9orf72* expansions and concurrent *GRN* mutations, one patient with a *C9orf72* expansion and a concurrent *SQSTM1* mutation, as well as an additional *VCP* mutation in the patient with the *VCP* DV. These patients were only counted once, with patients with concurrent *C9orf72* expansions and *GRN* mutations listed as *C9orf72* expansions.

Patients with *MAPT* mutations were presumed to have tau pathology, while patients with *C9orf72* expansions, *GRN*, *TARDBP*, and *TBK1* mutations were presumed to have TBK-43 pathology. Patients with *VCP, CHMP2B,* and *TYROBP* mutations were excluded from the analyses comparing FTD-tau and FTD-TDP.

Of the 704 FTD patients included in the analysis, 220 were found to carry at least one ApoE4 allele; only three patients (all *MAPT* mutation carriers) were homozygous. AAO for ApoE4+ patients was 57.5 years old, and for ApoE- patients was 58.3 years old. As has been shown previously[192], ApoE4 was more common in clinically diagnosed FTD compared to healthy elderly controls; 221 patients (31.4%) and 105 controls (23.2%) were at least heterozygous for ApoE4 (Chi-squared-test, p=0.003), carrying either one or two ApoE4 alleles.

All details are listed in Table III-7; all statistical comparisons of AAO are laid out further on in “Effects on age at onset”.

Table III-7: Characteristics of patients with FTD and healthy elderly controls.

AAO = age at clinical onset, given in years of age; N = number of cases.

"All FTD" = all 704 FTD patients included in this study. "Clinical FTD" = 449 patients with a clinical diagnosis of FTD but no genetic or neuropathological diagnosis. "Confirmed FTD" = 255 patients with a genetic or neuropathological diagnosis, or both: (59 *C9orf72* expansions, 53 *MAPT* mutations, 44 *GRN* mutations, 5 *VCP* mutations, 2 *TBK1* mutations, 1 *CHMP2B 1, TYROBP* and 1 *TARDBP* mutation), 139 patients with neuro-pathologically confirmed disease: 70 patients with tau pathology, 65 patients with TDP-43 pathology, and 4 FUS cases. 19 patients had both neuro-pathologically confirmed tau pathology and a *MAPT* mutation, and 31 patients had both TDP-43 pathology and a DV (13 *C9orf72* expansion, 17 *GRN* mutations, and 1 *TBK1* mutation); these were only counted once, explaining disparities of sums in the table. "N Mutation – Tau" = cases with DVs in the *MAPT* gene. "N Mutation – TDP" = cases with DVs in *C9orf72* (59 cases), *GRN* (44 cases), *TARDBP* (one case) and *TBK1* (2 cases) genes, known to cause TBK-43 pathology. "N Mutation – Confirmed FTD " = all genetically confirmed cases as described above. Sub-cohorts referring to Aβ only include cases with available neuropathological data. AAO is given in years and in controls refers to age at testing.

	All FTD	Clinical FTD	Confirmed FTD	FTD - Tau	FTD - TDP	Controls
N	704	449	255	103	141	456
AAO (years)	58.1	58.9	56.4	55.7	58.1	76.6
N Mutation	-	-	166	53	107	-
N Neuropathology	-	-	139	70	65	-
N ApoE4+	220	140	81	32	44	105
ApoE4+ AAO (years)	57.5	59.1	54.9	53.0	58.0	75.2
ApoE4- AAO (years)	58.3	58.8	57.4	56.9	58.1	77.1
N Aβ+	-	-	44	26	18	-
AAO Aβ+	-	-	61	60.2	62.2	-
AAO Aβ-	-	-	57.1	56.9	58.4	-
AAO (N) ApoE4+ / Aβ -	-	-	51.4 (12)	49.3 (6)	56.2 (5)	-
AAO (N) ApoE4- / Aβ -	-	-	58.1 (71)	58.4 (30)	58.7 (37)	-
AAO (N) ApoE4+ / Aβ+	-	-	56.1 (23)	54.9 (15)	58 (8)	-
AAO (N) ApoE4- / Aβ+	-	-	66.3 (21)	67.0 (11)	65.6 (10)	-

b. *Neuropathological results*

The existing clinical and genetic data was supplemented with neuropathological information obtained from the Queen Square Brain Bank.

i. FTD pathology

Of the 704 FTD patients included in the analysis, 139 patients had neuro-pathologically confirmed disease: 70 patients with tau pathology and 65 patients with TDP-43 pathology, as well as 4 neuro-pathologically confirmed FUS cases. 19 patients with neuro-pathologically confirmed tau pathology also had a genetically confirmed deleterious MAPT mutation, while 31 patients with neuro-pathologically confirmed TDP-43 pathology were found to also carry a deleterious mutation (DV) (13 *C9orf72* expansions, 17 *GRN* mutations, 1 *TBK1* mutation). Patients with both genetically and neuro-pathologically confirmed disease were only counted once for the FTD-Tau and FTD-TDP cohorts, respectively, explaining disparities of sums in Table III-7.

FTD-tau comprised 103 patients with an average AAO of 55.7 years, including 32 ApoE4 allele carriers, while the FTD-TDP cohort was made up of 141 patients with an average AAO of 58.1 years, including 44 ApoE4 carriers. For one FTD-tau patient, information on AAO was not available.

ii. Amyloid-beta

For obvious reasons, information on amyloid-beta (Aβ) status was only available for the 139 patients with neuro-pathologically confirmed disease; 44 of these were Aβ positive. However, given that this was a retrospective study, records were not always complete and for 7 patients with tau pathology and 5 patients with TDP pathology information on Aβ status was not available.

c. *Effects on age at onset*

Patients were analysed in groups based on genetic and/or neuropathological data, patients in whom no causative mutation had been identified and for whom no neuropathological data were available either, were classified as “clinically diagnosed” and not included in analyses of genetic subgroups. “All FTD” comprised all 704 FTD patients included in this study; these were all patients from the UCL FTD cohort except

patients previously shown to carry mutations not typically associated with FTD[24]. "Clinical FTD" referred to the 449 patients with a clinical diagnosis of FTD but without a genetic diagnosis despite previous testing on the MRC Dementia Gene panel[24] (see Results Chapter 1) and no available neuropathological data.

"Confirmed FTD" comprised 255 patients with a genetic or neuro-pathological confirmation of their diagnosis or both, as described above. Based on genetic and neuro-pathological diagnoses, patients with confirmed FTD were grouped into two subcohorts: FTD- tau and FTD-TDP; 11 cases with mutations likely to cause neither tau nor TDP-43 pathology, or with other neuro-pathologically confirmed pathologies, such as FUS pathology, were excluded from these sub-cohorts (5 *VCP*, 1 *CHMP2B*, and 1 *TYROBP* mutations, and 4 cases with FUS pathology).

i. ApoE4

Based on Shi *et al.*[147], it could be expected that ApoE4 would have a modifying effect on AAO in FTD cases with proven or known or expected tau pathology. While AAO is known to be highly correlated among family members with a highly penetrant *MAPT* gene mutation[76], here, AAO in confirmed FTD patients who had *MAPT* gene mutations or FTLD-tau neuropathology was significantly lower in the presence of an ApoE4 allele (tau+ ApoE4+ compared to tau+ ApoE4- patients; 53.0 years vs. 56.9 years, t-test, p=0.043, Figure III-8 A). In a further exploratory analysis, we tested whether ApoE4 also affected non-tau cases in the same way. However, ApoE4 had no modifying effect on AAO in FTD patients with FTLD-TDP43 neuropathology or in FTD patients without an ascertained cause, nor did sex affect AAO in the different groups (TDP+ ApoE4+ compared to TDP+ ApoE4-: 57.6 years vs. 58.1 years; Clinical FTD ApoE4+ compared to Clinical FTD ApoE4-: 58.7 years vs. 58.9 years; ANOVA, all p-values > 0.05; all Figure III-8 A).

ii. Amyloid-beta

Given the accepted understanding that ApoE4 risk in AD is related to increased Aβ aggregation[191] and decreased Aβ clearance, it appeared imperative to test whether the modifying effect of ApoE in FTD-Tau might relate to the presence of Aβ co-pathology

without meeting criteria for a formal pathological diagnosis of AD in 139 cases with available neuropathological data. Post-mortem cases were neuro-pathologically staged according to Thal phases and Braak and Braak stages[197,198]. A number of cases (N=44) scored at Thal phases 0, 1, 2 or 3, but none reached Thal Phase 5 or Braak stage 6, needed for a definitive diagnosis of AD. ApoE4 carrier status increased the likelihood of Aβ pathology, both in cases with tau and TDP43 pathology (Chi-squared-test, p=0.001 and Chi-squared-test, p=0.006, respectively): for the tau+ ApoE4+ patients with FTD and available neuropathology data, 15 out of 21 (71.4%) had Aβ co-pathology compared to 11/42 (26.2%) ApoE4- tau+ cases (Chi-squared-test, p=0.001); for the TDP+ ApoE4+ FTD cases, 8/13 (61.5%) had Aβ co-pathology compared to 10/47 (21.3%) of ApoE4- TDP+ cases (Chi-squared-test, p=0.013). The prevalence of Aβ co-pathology was similar when tau+ cases were compared with TDP+ cases (Chi-squared-test, p=0.259) and was associated with ApoE4 carrier status, both in cases with tau pathology (Chi-squared-test, p=0.001) and without tau pathology (Chi-squared-test, p=0.006). Patients who were positive for Aβ pathology were significantly older at onset than patients who were found to have no Aβ pathology at post-mortem (60.9 vs. 57.1 years old, respectively, t-test, p=0.024). In patients with FTLD-tau, ApoE genotype and Aβ co-pathology had independent associations with age at clinical onset (linear regression, n=62, ApoE -10.7 years, $p=5.2 \times 10^{-5}$; Aβ +7.6 years, p=0.002). The effects of ApoE4+ carrier status and Aβ on AAO were independent of each other both in patients with and without tau (Univariate ANOVA, p=0.550 and p=0.545, respectively).

iii. Combined effects and interpretation

These analyses, carried out in a well-characterized cohort of FTD patients, provide evidence that ApoE4 genotype accelerates neurodegeneration in patients with *MAPT* mutations or FTLD-tau pathology independently of Aβ – notably AAO in FTD-tau cases was on average about 3.9 years earlier in those who carried an ApoE4 allele. Previously, Shi et al. proposed that ApoE4 genotype acts through a toxic gain of function mechanism to exacerbate or modify tau pathology, neuroinflammation, autophagy, and reactive astrocyte activation; they provided extensive evidence in a mouse model of FTD and in neuropathological cases of tauopathies, as well as clinical support in an AD cohort[147]. These data further strengthen this hypothesis providing

evidence in human that ApoE4 modifies the clinical phenotype of FTLD-tau, independently of Aβ co-pathology.

While ApoE4 status was associated with Aβ co-pathology at post-mortem examination in both tau+ and in TDP+ cases of FTD, its effect on AAO was specific to tau+ FTD cases (Figure III-8 A). Indeed, ApoE4 and Aβ co-pathology were associated with opposing effects on age at onset in this dataset in tau+ FTD cases (Figure III-8 B). The confounding impact of Aβ co-pathology, along with the heterogeneous nature of FTD pathology and the specificity of the effect of ApoE4 on tau pathology may explain why previous studies of ApoE4 in FTD[192-195] did not detect this influence. Because of the limits of this retrospective and observational dataset, further research in cell or animal models is needed to clarify whether Aβ has a protective effect.

In summary, the ApoE4 genotype was associated with lowered age at clinical onset in patients with frontotemporal dementia and tau pathology, and had a particularly strong effect once the confounding effects of amyloid beta pathology were taken into account. "Clinical FTD" patients were purposely included in this study to demonstrate that the effect of ApoE4 on age at onset is specific to the patients with proven or presumed tau pathology. The fact that there was no observable effect in the group overall, the clinically diagnosed cases and the TDP cases also goes some way towards explaining why this effect has not been observed before. Despite the robust statistical findings and the large cohort of FTD patients, analysis of pathological and gene mutation defined subgroups inevitably leads to relatively restricted numbers and therefore further replication efforts should be encouraged. Future studies should analyse the effects of ApoE genotype on FTD, accounting for differential effects on tau and Aβ co-pathology, and elucidate the cellular mechanisms for this effect *in vitro*.

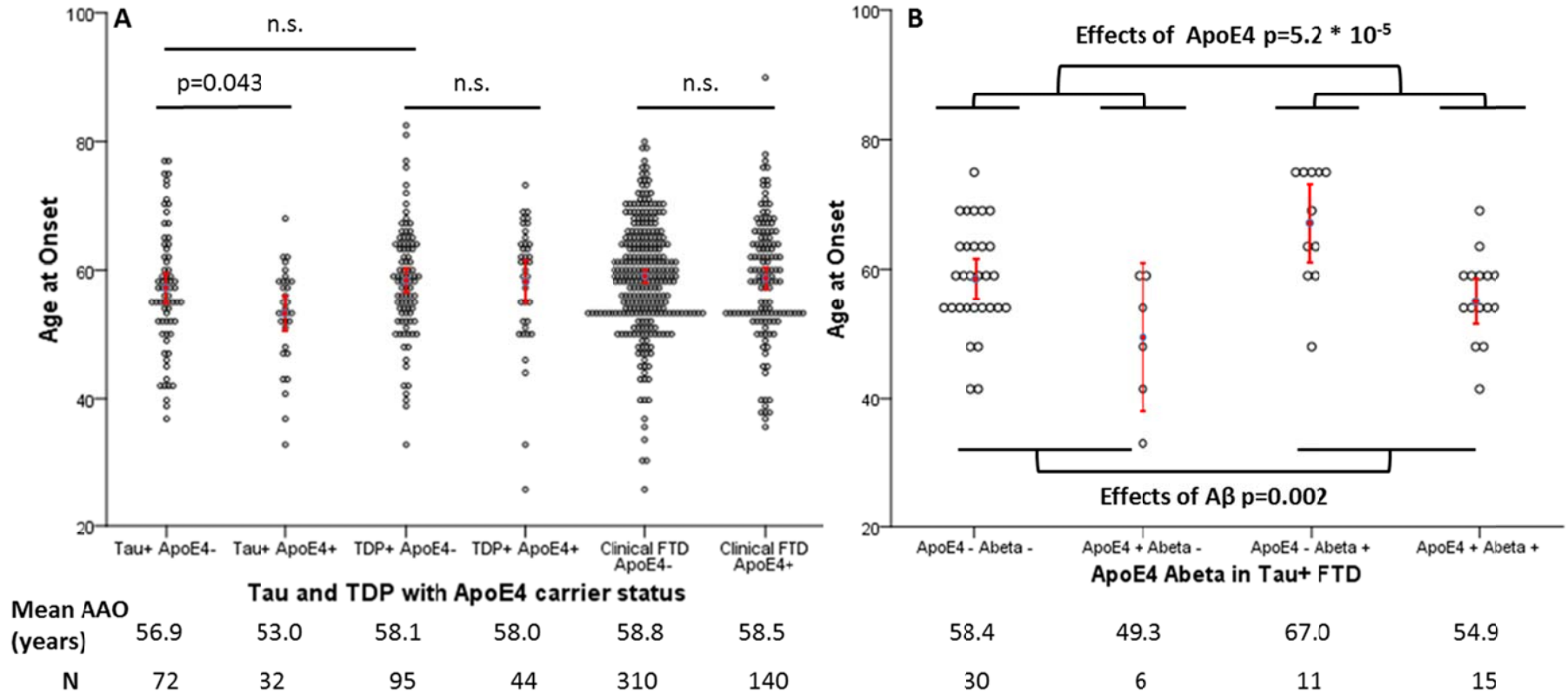

Figure III-8: ApoE4 lowers AAO in patients with FTD-tau

ApoE4+ status is associated with significantly lower age at onset (AAO) in FTD cases with *MAPT* mutations or FTLD-tau neuropathology, but not in those with TDP pathology or in those without an ascertained genetic or neuropathological diagnosis (Clinical FTD) (*Figure III-8 A*). ApoE4 carrier status is associated with lower AAO, while Aβ co-pathology was associated with later AAO (*Figure III-8 B*); Figure III-8 B displays the p values for the regression coefficients ApoE4 and Aβ. Each dot corresponds to the AAO of one case, with the mean and 95% confidence interval marked in red. The number of patients in each subcohort is given in line *N*. For Figure III-8 A, AAO was not available for 3 Tau+ApoE4- patients, 4 TDP+ApoE4- and 6 TDP+ApoE4+ cases, as well as 31 Clinical FD ApoE4- and 9 Clinical FTD ApoE4+ patients; for Figure III-8 B, 62 FTD Tau+ patients with available ApoE4 status and neuropathological data were included; AAO was not available for 1 ApoE4-Abeta- patient, and the Aβ status was unknown for 7 patients. These cases were excluded from the relevant calculations and from the figures.

3 Chapter 3: The HD phenocopy syndromes

A. Improving the definition of HD phenocopy syndromes

Huntington's disease (HD) is defined clinically by a movement disorder associated with neuropsychiatric/cognitive symptoms and is transmitted in families as an autosomal dominant trait. Patients in whom HD is suspected do not always carry the Huntington's gene (*HTT*) expansion and may be said to have a phenocopy syndrome. The prevalence is unclear and the cause of a large proportion of these cases with dementia and motor symptoms remains to be elucidated. Such patients range from those mimicking HD exactly, to those with partially overlapping clinical features.

The definition of HDPC syndromes as *patients in whom an experienced clinician deems HD likely enough to request a gene test* is, while simple, practical and taking into account real-world clinical uncertainties, is inherently subject to variation between clinicians based on judgement, as well as geography, time, and patient factors like detail of the symptom and family history and development of the clinical syndrome at the time of the examination. In order to improve our understanding of these complex syndromes and discover what might differentiate patients with HD from those with an HDPC syndrome, a survey on symptoms at presentation was sent out to clinicians with expertise in HD as a first step; additionally, patients who were tested for HD in two Neurogenetics clinics with a dementia with motor symptoms focus were analysed with regards to their symptoms at presentation.

a. Survey of Huntington's Disease expert clinicians on initial clinical presentation

In order to establish what distinguishes HD and HDPC clinically, 130 clinician experts on HD were invited to take part in a survey that considered the symptoms that lead the clinician to suspect HD or an HDPC syndrome when a patient first presents to clinic (for the survey questions and detailed responses see Table III-8. 52 experts replied to the survey; of these 36 (69.2%) were Neurology consultants, 11 (21.1%) were Neurogenetics consultants, 25 were movement disorders consultants, five were psychiatry consultants and one clinical genetics consultant, with some responders

selecting more than one option. Clinicians had been working with HD and HDPC patients for a long time (mean 18.4 years, range 3-36 years) and saw a large number of relevant patients per month (mean 17 patients, range 1-50). 12 respondents were located in the UK, 19 in continental Europe, and 21 responses came from the United States of America.

Symptoms that would make these experts most strongly consider an HD test (based on question 5) included chorea (100%), cognitive slowing (30.8%), a dysexecutive syndrome (30.8%), irritability (30.8%), gait abnormalities/falls (28.9%), and dystonia (25.0%), which is a frequent symptom in HD[95,96,199], as well as other disorders[200]. Symptoms that would most strongly suggest a different syndrome to them (based on question 7) included neuropathy (74.5%), limb weakness (54.9%), ataxia (35.3%), pain (33.3%), tremor (27.5%), and hallucinations (21.6%).

Table III-8: Survey questions sent to HD experts

Q1. What is your professional background? Please state your specialty.			
Neurology	36 / 70.6%	Movement Disorders	25 / 49%
Neurogenetics	11 / 21.6%	Psychiatry	5 / 9.8%
Clinical Genetics	1 / 2%		
Answered	51		
Q2. How many years have you been working with HD and HD phenocopy patients?			
3 years	1 / 1.9%	17 years	1 / 1.9%
4 years	1 / 1.9%	19 years	2 / 3.8%
6 years	2 / 3.8%	20 years	9 / 17.3%
7 years	1 / 1.9%	22 years	1 / 1.9%
8 years	2 / 3.8%	23 years	2 / 3.8%
10 years	4 / 7.7%	25 years	4 / 7.7%
11 years	3 / 5.8%	28 years	1 / 1.9%
12 years	1 / 1.9%	30 years	6 / 11.5%
13 years	1 / 1.9%	31 years	1 / 1.9%
14 years	2 / 3.8%	34 years	1 / 1.9%
15 years	5 / 9.6%	36 years	1 / 1.9%
Answered	52		
Q3. How many HD and/or HD phenocopy patients do you typically see in a month?			
1 patient	3 / 5.8%	18 patients	2 / 3.8%
2 patients	1 / 1.9%	20 patients	5 / 9.6%
4 patients	2 / 3.8%	25 patients	4 / 7.7%

5 patients	2 / 3.8%	30 patients	2 / 3.8%
6 patients	2 / 3.8%	35 patients	1 / 1.9%
8 patients	3 / 5.8%	40 patients	4 / 7.7%
10 patients	12 / 23.1%	45 patients	1 / 1.9%
12 patients	1 / 1.9%	50 patients	1 / 1.9%
15 patients	4 / 7.7%		
Answered	52		
Q4. In a patient with a compatible syndrome and a possibly positive family history (but without a genetic diagnosis in the family), what symptom or symptoms would make you most strongly consider an HD test?			
Chorea	52 / 100%	Agitation	12 / 23.1%
Dystonia	20 / 38.5%	Apathy	28 / 53.9%
Rigidity	9 / 17.3%	Anxiety	17 / 32.7%
Gait abnormalities / Falls	30 / 57.7%	Depression	20 / 38.5%
Ataxia	6 / 11.5%	Irritability	34 / 65.4%
Tremor	0 / 0%	Disinhibition	16 / 30.8%
Dysarthria	12 / 23.1%	Obsessive behaviour	12 / 23.1%
Dysphagia / Choking	11 / 21.2%	Paranoia	5 / 9.6%
Memory loss	9 / 17.3%	Delusions	4 / 7.7%
Disorientation / Navigational difficulties	5 / 9.6%	Hallucinations	2 / 3.9%
Cognitive slowing	25 / 48.1%	Change in dietary habits	0 / 0%
Dysexecutive syndrome	31 / 59.6%	Limb weakness	0 / 0%
Loss of empathy	13 / 25%	Weight loss	11 / 21.2%
Insomnia	4 / 7.7%	Pain	0 / 0%
Hypersomnia	0 / 0%	Neuropathy (sensory / motor)	0 / 0%
Answered	52		
Q5. Based on Q4, please tick only your top 3 symptom which would make you consider an HD test.			
Chorea	52 / 100%	Agitation	0 / 0%
Dystonia	13 / 25%	Apathy	6 / 11.5%
Rigidity	1 / 1.9%	Anxiety	2 / 3.9%
Gait abnormalities / Falls	15 / 28.9%	Depression	2 / 3.9%
Ataxia	2 / 3.9%	Irritability	16 / 30.8%
Tremor	0 / 0%	Disinhibition	6 / 11.5%
Dysarthria	1 / 1.9%	Obsessive behaviour	0 / 0%
Dysphagia / Choking	2 / 3.9%	Paranoia	1 / 1.9%
Memory loss	3 / 5.8%	Delusions	0 / 0%
Disorientation / Navigational difficulties	0 / 0%	Hallucinations	0 / 0%
Cognitive slowing	16 / 30.8%	Change in dietary habits	0 / 0%
Dysexecutive syndrome	16 / 30.8%	Limb weakness	0 / 0%
Loss of empathy	1 / 1.9%	Weight loss	1 / 1.9%
Insomnia	1 / 1.9%	Pain	0 / 0%

Hypersomnia	0 / 0%	Neuropathy (sensory / motor)	0 / 0%
Answered	52		
Q6. In such a patient, is there a clinical symptom (or symptoms) that, if present, would make you less inclined to order an HD test, possibly because that symptom is making another diagnosis more likely in your mind?			
Chorea	1 / 2%	Agitation	1 / 2%
Dystonia	1 / 2%	Apathy	0 / 0%
Rigidity	5 / 9.8%	Anxiety	2 / 3.9%
Gait abnormalities / Falls	2 / 3.9%	Depression	3 / 5.9%
Ataxia	22 / 43.1%	Irritability	0 / 0%
Tremor	25 / 49%	Disinhibition	4 / 7.8%
Dysarthria	2 / 3.9%	Obsessive behaviour	1 / 2%
Dysphagia / Choking	1 / 2%	Paranoia	0 / 0%
Memory loss	7 / 13.7%	Delusions	0 / 0%
Disorientation / Navigational difficulties	5 / 9.8%	Hallucinations	13 / 25.5%
Cognitive slowing	1 / 2%	Change in dietary habits	6 / 11.8%
Dysexecutive syndrome	0 / 0%	Limb weakness	30 / 58.8%
Loss of empathy	1 / 2%	Weight loss	0 / 0%
Insomnia	3 / 5.9%	Pain	28 / 54.9%
Hypersomnia	7 / 13.7%	Neuropathy (sensory / motor)	38 / 74.5%
Answered	51		
Q7. Based on Q6, please tick only your 3 top symptom choices that would make you less inclined to order an HD test.			
Chorea	0 / 0%	Agitation	0 / 0%
Dystonia	0 / 0%	Apathy	0 / 0%
Rigidity	1 / 2%	Anxiety	1 / 2%
Gait abnormalities / Falls	0 / 0%	Depression	1 / 2%
Ataxia	18 / 35.3%	Irritability	0 / 0%
Tremor	14 / 27.5%	Disinhibition	1 / 2%
Dysarthria	0 / 0%	Obsessive behaviour	0 / 0%
Dysphagia / Choking	1 / 2%	Paranoia	0 / 0%
Memory loss	4 / 7.8%	Delusions	0 / 0%
Disorientation / Navigational difficulties	3 / 5.9%	Hallucinations	11 / 21.6%
Cognitive slowing	0 / 0%	Change in dietary habits	6 / 11.8%
Dysexecutive syndrome	0 / 0%	Limb weakness	28 / 54.9%
Loss of empathy	0 / 0%	Weight loss	1 / 2%
Insomnia	1 / 2%	Pain	17 / 33.3%
Hypersomnia	3 / 5.9%	Neuropathy (sensory / motor)	38 / 74.5%
Answered	51		

Q8. What combination of symptoms would be sufficient for you to consider HD and test for it?	
Chorea - cognitive problems - anxiety - with family history (one affected parent or grandparent)	48 / 92.3%
Chorea - cognitive problems - anxiety - no family history	43 / 82.7%
Dystonia - cognitive problems - anxiety - with family history (one affected parent or grandparent)	36 / 69.2%
Dystonia - cognitive problems - anxiety - no family history	19 / 36.5%
Rigidity- cognitive problems - apathy- with family history (one affected parent or grandparent)	31 / 59.6%
Rigidity- cognitive problems - apathy- no family history	11 / 21.2%
Chorea - with family history (one affected parent or grandparent)	46 / 88.5%
Chorea - no family history	36 / 69.2%
Dystonia - with family history (one affected parent or grandparent)	27 / 51.9%
Dystonia - no family history	3 / 5.8%
Rigidity - with family history (one affected parent or grandparent)	25 / 48.1%
Rigidity - no family history	2 / 3.9%
Dysexecutive syndrome - with family history (one affected parent or grandparent)	29 / 55.8%
Dysexecutive syndrome - no family history	2 / 3.9%
Irritability - cognitive slowing - with family history (one affected parent or grandparent)	34 / 65.4%
Irritability - cognitive slowing - no family history	8 / 15.4%
Dysexecutive syndrome - depression - with family history (one affected parent or grandparent)	30 / 57.7%
Dysexecutive syndrome - depression - no family history	2 / 3.9%
Disinhibition - with family history (one affected parent or grandparent)	20 / 38.5%
Disinhibition - no family history	1 / 1.9%
Answered	52

Q9. Are there any symptoms that, if present, would make you expect that an HD test will come back negative, despite the obvious need to test for (and exclude) HD?			
Chorea	0 / 0%	Agitation	0 / 0%
Dystonia	1 / 1.9%	Apathy	0 / 0%
Rigidity	0 / 0%	Anxiety	1 / 1.9%
Gait abnormalities / Falls	0 / 0%	Depression	1 / 1.9%
Ataxia	11 / 21.2%	Irritability	0 / 0%
Tremor	13 / 25%	Disinhibition	0 / 0%
Dysarthria	1 / 1.9%	Obsessive behaviour	0 / 0%
Dysphagia / Choking	0 / 0%	Paranoia	1 / 1.9%
Memory loss	4 / 7.7%	Delusions	1 / 1.9%
Disorientation / Navigational difficulties	2 / 3.9%	Hallucinations	6 / 11.5%
Cognitive slowing	0 / 0%	Change in dietary habits	3 / 5.8%
Dysexecutive syndrome	1 / 1.9%	Limb weakness	24 / 46.2%

Loss of empathy	0 / 0%	Weight loss	1 / 1.9%
Insomnia	1 / 1.9%	Pain	21 / 40.4%
Hypersomnia	9 / 17.3%	Neuropathy (sensory / motor)	32 / 61.5%
None of these	12 / 23.1%		
Answered	52		

Q10. Based on Q9, which top 3 symptoms would make you expect that an HD test will come back negative, despite the obvious need to test for it?

Chorea	0 / 0%	Agitation	0 / 0%
Dystonia	1 / 1.9%	Apathy	0 / 0%
Rigidity	0 / 0%	Anxiety	1 / 1.9%
Gait abnormalities / Falls	0 / 0%	Depression	1 / 1.9%
Ataxia	8 / 15.4%	Irritability	0 / 0%
Tremor	10 / 19.2%	Disinhibition	0 / 0%
Dysarthria	0 / 0%	Obsessive behaviour	0 / 0%
Dysphagia / Choking	0 / 0%	Paranoia	1 / 1.9%
Memory loss	2 / 3.9%	Delusions	1 / 1.9%
Disorientation / Navigational difficulties	2 / 3.9%	Hallucinations	4 / 7.7%
Cognitive slowing	0 / 0%	Change in dietary habits	3 / 5.8%
Dysexecutive syndrome	1 / 1.9%	Limb weakness	24 / 46.2%
Loss of empathy	0 / 0%	Weight loss	2 / 3.9%
Insomnia	0 / 0%	Pain	18 / 34.6%
Hypersomnia	2 / 3.9%	Neuropathy (sensory / motor)	34 / 65.4%
None of these	10 / 19.2%		
Answered	52		

Symptoms that would make the clinician experts expect a negative HD test results despite the need to test due to other co-existing symptoms included (based on question 10) a sensory or motor neuropathy (65.4%), limb weakness (46.2%), pain (34.6%), tremor (19.2%), and ataxia (15.4%) ; however a substantial number of HD experts (19.2% - as many as those choosing rigidity and tremor) felt that none of the other symptoms would make them expect a negative HD test results. As for combinations of symptoms, most experts agreed that motor, cognitive, and psychiatric symptoms all warranted HD testing if the patient had a positive family history with an affected parent or grandparent; however only chorea, with or without additional symptoms, was deemed sufficiently suggestive to warrant HD testing in the absence of

a positive family history (see Table III-8, HD Expert Clinicians Survey and Results Table III-9, Question 8).

Ultimately, there was little overlap between what HD clinician experts considered typical of HD and of HDPC syndromes (Q5 and Q7). Few symptoms were expected at similar levels in both HD and HDPC syndromes, and all of them at low levels; these included rigidity, insomnia, and weightloss (all 1.9% vs. 2.0%, HD and HDPC syndromes respectively).

Given that these are survey results and not real life patients, and due to the way questions were worded in order to obtain the most objective answers, the results for symptoms that are expected in HD and in HDPC syndromes are difficult to compare. However, what can be gleaned from the results comparing the answers to more broadly worded questions 4 and 6 (see Table III-9) is that HD clinician experts expected patients presenting with ataxia, hallucinations, limb weakness, neuropathy, pain and tremor to be ultimately diagnosed with an HDPC syndrome rather than HD. 32 symptoms were compared using Chi-Square tests and corrected for multiple testing using the Bonferroni method. The largest differences between HD and HDPC (ΔHD - HDPC) were predicted to be found in chorea (+), neuropathy (-), irritability (+), dysexecutive syndrome (+), limb weakness (-), apathy (+), pain (-), falls (+), gait abnormality (+), tremor (-), cognitive slowing (+), dystonia (+), and depression (+).

It is important to note that Q4 and Q6, while differently phrased, both ask what is typical of HD, not what is typical for an HDPC syndrome. Question 4 asks what is typical for HD and Question 6 asks what symptoms would be suggestive of a different syndrome, rather than what symptoms could be expected in HDPC syndromes. This limits what can be inferred from Table III-9 to the actively made statements rather than potentially implied ones. It means that while the above statement is true, it cannot be inferred from the results that HD clinician expert would not expect HDPC syndromes to present with chorea, apathy or falls, for example.

Table III-9: Symptoms expected to be typical for HD or pointing towards HDPC syndromes

This table compares Q 4 and Q6 of the expert survey which aim to establish what HD expert clinicians expect to be typical and less typical symptoms in HD; p-values marked with an asterix remain significant at the 0.05 level after correcting for multiple testing using the Bonferroni method

Characteristic or Symptom	HD	HDPC	ΔHD-HDPC	p-value (Bonferroni corrected, Chi-Square Test))
Agitation	12 (23.1%)	1 (1.9%)	+ 21.2%	0.001*
Anxiety	17 (32.7%)	2 (3.8%)	+ 28.9%	0.0002*
Apathy	28 (53.8%)	0 (0%)	+ 53.8%	$4.7599*10^{-11}$*
Ataxia	6 (11.5%)	22 (42.3%)	- 30.8%	0.001
Chorea	52 (100%)	1 (1.9%)	+ 98.1%	$6.70*10^{-29}$*
Cognitive Slowing	25 (48.1%)	1 (1.9%)	+ 46.2%	$2.25*10^{-08}$*
Delusions	4 (7.7%)	1 (1.9%)	+ 5.8%	n.s.
Depression	20 (38.5%)	3 (5.8%)	32.7%	0.000089*
Disinhibition	16 (30.8%)	4 (7.7%)	+ 23.1%	0.005
Disorientation/Navigation	5 (9.6%)	5 (9.6%)	0.0%	n.s.
Dysarthria	12 (23.1%)	2 (3.8%)	+ 19.3%	0.008
Dysexecutive Syndrome	31 (59.6%)	0 (0%)	+ 59.6%	$1.37*10^{-12}$*
Dysphagia /Choking	11 (21.2%)	1 (1.9%)	+ 19.3%	0.004
Dystonia	20 (38.5%)	1 (1.9%)	+ 36.6%	0.000003*
Falls	30 (57.7%)	2 (3.8%)	+ 53.9%	$1.15*10^{-9}$*
Gait abnormality	30 (57.7%)	2 (3.8%)	+ 53.9%	$1.15*10^{-9}$*
Hallucinations	2 (3.8%)	13 (25%)	- 21.2%	0.004
Hypersomnia	0 (0%)	7 (13.5%)	- 13.5%	0.013
Insomnia	4 (7.7%)	3 (5.8%)	+ 1.9%	n.s.
Irritability	34 (65.4%)	0 (0%)	+ 65.4%	$2.93*10^{-14}$*
Limb weakness	0 (0%)	30 (57.7%)	- 57.7%	$4.6096*10^{-12}$*
Loss of Empathy	13 (25%)	1 (1.9%)	+ 23.1%	0.001*
Memory Loss	9 (17.3%)	7 (13.5%)	+ 3.8%	n.s.
Motor Neuropathy	0 (0%)	38 (73.1%)	- 73.1%	$9.78*10^{-17}$*
Sensory Neuropathy	0 (0%)	38 (73.1%)	- 73.1%	$9.78*10^{-17}$*
Obsessive Behaviour	12 (23.1%)	1 (1.9%)	+ 21.2%	0.002
Pain	0 (0%)	28 (53.8%)	- 53.8%	$4.76*10^{-11}$*
Paranoia	5 (9.6%)	0 (0%)	+ 9.6%%	0.057
Rigidity	9 (17.3%)	5 (9.6%)	+ 7.7%	n.s.
Sweet Tooth/Change in dietary habits	0 (0%)	6 (11.5%)	- 11.5%	0.027
Tremor	0 (0%)	27 (51.9%)	- 51.9%	$1.29*10^{-9}$*
Weight Loss	11 (21.2%)	0 (0%)	+ 21.2%	0.001*

b. *Clinical symptoms of HD and HDPC patients at HD testing*

In order to compare expert clinician perception of likely clinical indicators of a positive HD test result to symptoms documented at the point of HD testing, the electronic patient records of patients referred for HD testing were analysed. During the studied period from January 2016 until July 2018, 151 individual patients attended the two Neurogenetics clinics included in this investigation; 89 of them were tested for Huntington's disease (HD) and were included in this analysis, 13 patients were seen and had a positive presymptomatic test but did not develop symptoms during follow up, and 13 patients did not have the genetic test for HD and were excluded from the analysis. 59 patients had been tested for HD before attending the clinic between 1999 and 2015, 20 patients had their test ordered from this clinic and for eight patients the test date was not documented in the electronic patient records. 36 patients' electronic patient records were incomplete and their clinical picture at initial presentation could not be assessed. Out of the 89 patients tested for HD, 70 (78.7%) were found to have an expansion in the *HTT* gene and were diagnosed with HD, while a further 19 patients (21.3%) did not carry the expansion; the latter group are deemed to be patients with an HD Phenocopy (HDPC) syndrome (see Figure III-9).

Assuming these rates are comparable between centres, and given that the prevalence of HD is estimated at 10.6 to 13.7 individuals per 100,000 population[201], this would put the prevalence of HDPC syndromes, i.e. patients with a likely genetic, cryptic, undiagnosed movement disorder with cognitive and neuropsychiatric symptoms between 2.3 and 2.9 individuals per 100,000 population.

Of the 19 HDPC patients, two patients' symptoms were later suspected to be due to vascular causes (post-stroke and vascular dementia), two patients were thought to have drug-related chorea (one subsequently diagnosed as tardive dyskinesia following neuroleptic treatment and one amphetamine-related), and one patient was later thought to have reactivated Sydenham's chorea but later progressed although the diagnosis remained unclear; one patient had spinocerebellar ataxia 17 (SCA17) with an intermediate allele, one patient was later diagnosed with neurodegeneration with

brain iron accumulation without an identified causal mutation, one patient had a working diagnosis of Meige syndrome (oromandibular dystonia and blepharospasm) and one patient's childhood-onset chorea with dystonia was thought to be due to a lesion of the left caudate nucleus; a further 9 patients had an unclear progressive syndrome with chorea or dystonia and/or other symptoms. 52.8% of HD patients and 52.6% of HDPC patients were female. Details of these patients can be found in Table III-11. Running a logistical regression with GS4.5 (censored family history) as the constant, confirmed the overall significantly differentiating effect of the Goldman score (p= 0.021) between the HD and HDPC cohorts, but the various GS categories did not reach significance.

Of these patients in the HDPC cohort, only three patients received a certain diagnosis; these were the patient with Parkinson's disease confirmed by DaT scan, the patient with genetically confirmed SCA 17 (although with an intermediate allele count) and the patient with genetically confirmed SCA1.

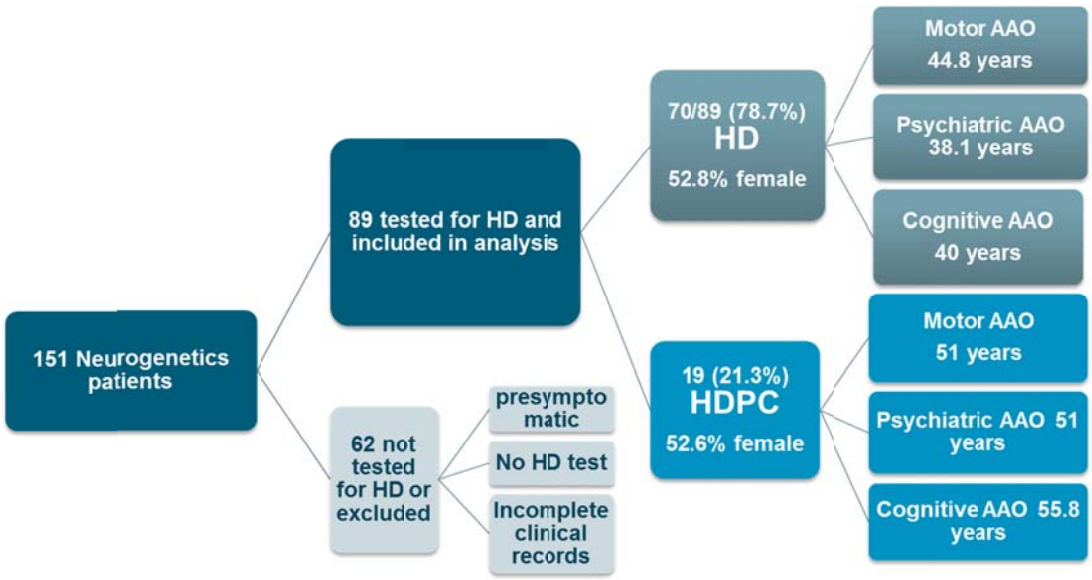

Figure III-9: Patients from two clinics for Neurogenetics counselling and HD testing

151 patients from two clinics for Neurogenetics were included in the analysis. 89 patients had an HD test; in 21.3% of cases tested for HD, the test result was negative. Ages at onset of motor symptoms were broadly similar in patients with both a positive and a negative HD test result, but ages at onset of psychiatric and cognitive symptoms varied more. This analysis did not distinguish on the basis of a known family history of HD, i.e. whether testing was or would have been carried out pre-symptomatically.

In order to improve statistical power, 50 additional HDPC cases were added to the calculations, bringing the total number of cases to 139 – 70 HD cases and 69 HDPC cases (see Table III-10). These cases were chosen based on the availability of sufficient clinical information from a subset of the UCL HD phenocopy cohort, whose HD test dates fall within the date range 2011 to 2015 when the majority of the cases seen in the Neurogenetics Clinic were tested. All electronic patient records were reviewed for clinical information, cases with incomplete patient records or patients included in the Neurogenetics clinic cohort were excluded. Of these additional 50 HDPC cases, seven had a working diagnosis of (atypical) Parkinson's disease, two had adult onset tics, 15 had chorea with or without cognitive or psychiatric features, seven had ataxia with sometimes additional dystonia or cognitive problems, five had psychiatric conditions (bipolar disorder, conversion disorder, obsessive compulsive disorder, Tourette syndrome and autistic spectrum disorder) three were diagnosed as tardive dyskinesias, three had frontotemporal dementia, two had progressive supranuclear palsy, two had young onset atypical dementia syndromes (one with adolescent onset), one had posterior cortical atrophy (PCA), and one patient each had jerky tremor, myoclonus, and restless legs.

The two patient cohorts were compared group-wise and examined for differences in 54 categories, including AAO, family history and clinical presentation; with the exception of the Goldman score, none of the results withstood Bonferroni's correction for multiple comparisons, but they remain interesting as an exploratory analysis. When they were diagnosed, the average age at motor onset (mAAO) of patients ultimately diagnosed with HD was 44.8 years old, while for HDPC patients it was 50.4 years old (mΔ=5.5; p=n.s., Independent t-test). Motor onset defines the official onset of HD, however, many HD patients experienced earlier psychiatric symptoms (average pAAO 38.1 years), while for HDPC patients, average pAAO was 47 years (pΔ=9.6; p=n.s., Independent t-test); for HD patients, cognitive symptoms average cAAO was 40 years, while for HDPC patients, it was 51.1years (cΔ=11.1; p=n.s., Independent t-test). At initial presentation, the average HDPC score for HD patients was 2.1/3, while for HDPC patients it was 2.2/3 (Δ=0; p=n.s.) or 2.9/4 and 2.6/4, respectively, when including a

positive family history (Δ=0.3; p=n.s.). A positive family history in this context is anyone with a relative with a diagnosis of adult-onset cognitive decline, psychiatric or both hyper- and hypokinetic motor symptoms (GS ≤ 3.5).

With regards to initial clinical presentation, most HD and HDPC patients presented with motor symptoms (HDPC motor score, 92.9% vs 92.8%, (Δ=0.001; p=n.s), as well as psychiatric (65.7% vs. 63.8%, respectively, HDPC score Δ=0.02, p=n.s., Chi-Square-test), and cognitive symptoms (54.3% vs. 60.9%, respectively, HDPC score Δ=0.07, p=n.s., Chi-Square-test). The commonest symptoms included chorea (68.6% of HD patients and 65.2% of HDPC patients, Δ=3.4%; p=n.s), depression (35.7% of HD patients and 27.5% of HDPC patients, Δ=8.2%; p=n.s), and memory loss (39.1% of HDPC patients, Δ=0.6%; p=n.s). Statistically significant differences were found for tremor (4.3% of HD patients and 26.1% of HDPC patients; Δ=21.8%; p=0.0003), dystonia (10% of HD patients and 33.3% of HDPC patients; Δ=23.3%; p=0.001), dysphagia/choking (31.4% of HD patients and 11.6% of HDPC patients; Δ=19.8%; p=0.007), dysarthria (31.4% of HD patients and 17.4% of HDPC patients; Δ=16.9%; p=0.033), disinhibition (0% of HD patients, 21.7% of HDPC patients; Δ=21.7%; p=0.00001), and insomnia (34.3% of HD patients and 11.6% of HDPC patients Δ=22.7%; p=0.002). Among statistically differing symptoms between the two cohorts, only insomnia, dysarthria and dysphagia were commoner in HD, while disinhibition, dystonia and tremor were more frequent in HDPC patients.

Weight loss was less specific than could have been expected at 4.2% difference (HD patients 11.4% and HDPC patients 7.2, p=n.s., Chi-Square test). Limb weakness, sensory neuropathy, spasticity, bladder and bowel dysfunction were likely to point towards an HDPC syndrome and were not described in HD in this cohort (all p=n.s.). In addition, some oculomotor problems, such as delayed saccade initiation and decreased range of movements, can be very sensitive for HD (although not necessarily specific), but are often not well documented; this is also the case for sustained tonic tongue movements. For none of the patients a United HD Rating Scale (UHDRS) score was recorded in the notes.

Goldman score (GS) as a measure of family history was significantly different between the cohorts (Chi-Square-test, $p=1.5058*10^{-11}$); GS was a differentiating factor between the two cohorts; HDPC patients had a GS1 in 20.3% of HDPC cases, GS2 in 5.8%, GS3 in 4.3%, GS3.5 in 17.4%, GS4 in 47.8% and GS4.5 in 4.3% of HDPC cases. Meanwhile, HD patients had a GS1 in 48.6% of HD cases, a GS2 in 4.3%, a GS3 in 24.3%, a GS3.5 in 4.3%, a GS4 in 14.3% and a GS4.5 in 4.3% of HD cases. Running a logistical regression test with GS4.5 (censored family history) as the constant, confirmed the overall significantly differentiating effect of the Goldman score ($p= 0.000009$) between the HD and HDPC cohorts, but the individual GS categories did not reach significance in isolation. Having a strong autosomal family history (GS1) and having no relevant family history (GS4) was therefore almost completely reversed between the two cohorts. This reflects that HD patients usually had a strong known family history of HD, with previously positive tests in the family (although in some cases family history was censored due to early deaths, loss of contact or adoption), while HDPC patients usually had a less clear family history and no definite genetic or neuropathological diagnosis had been established in their affected relatives.

Table III-10: Characteristics of the HD/HDPC cohort with additional HDPC cases

Symptom percentages are based on their being listed in the clinical notes and letters. Where possible a separate AAO for motor, cognitive and psychiatric symptoms was established from the notes (AAO Mot, AAO Cog, AAO Psy, respectively). In addition, the HDPC score was calculated for each patient based on whether they were displaying any symptoms in the cognitive, psychiatric and motor domains (cognitive, psychiatric, and motor HDPC score, respectively); this was used as a measure of how HD-like patients' clinical presentations were in terms of affecting different functional domains.

Characteristic or Symptom	HD (%HD)	HDPC (%HDPC)	ΔHD-HDPC	p-value (uncorrected Chi-square/t-test)
N	70	69		
Male	33 (47.1%)	29 (42.0%)	5.1%	n.s.
GS	2 (median)	4 (median)		n.s.
AAO Mot	44.9 years	50.4 years	5.5 years	n.s.
AAO Cog	40.0 years	51.1 years	11.1 years	n.s.
AAO Psy	37.4 years	47.0 years	9.6 years	n.s.
Agitation	6 (8.6%)	3 (4.3%)	+ 4.2%	n.s.

Characteristic or Symptom	HD (%HD)	HDPC (%HDPC)	ΔHD-HDPC	p-value (uncorrected Chi-square/t-test)
Anxiety	18 (25.7%)	15 (21.7%)	+ 4.0%	n.s.
Apathy	7 (10%)	4 (5.8%)	+ 4.2%	n.s.
Asymmetry	1 (1.4%)	2 (2.9%)	- 1.5%	n.s.
Ataxia	12 (17.1%)	13 (18.8%)	- 1.7%	n.s.
Babinski	1 (1.4%)	2 (2.9%)	- 1.5%	n.s.
Bladder/ Bowel Dysfunction	0 (0%)	1 (1.4%)	- 1.4%	n.s.
Bradykinesia	12 (17.1%)	11 (15.9%)	+ 1.2%	n.s.
Chorea	48 (68.6%)	45 (65.2%)	+ 3.4%	n.s.
Cognitive Slowing	17 (24.3%)	22 (31.9%)	- 7.6%	n.s.
Cortical Blindness	0 (0%)	1 (1.4%)	- 1.4%	n.s.
Delusions	2 (2.9%)	3 (4.3%)	- 1.5%	n.s.
Depression	25 (35.7%)	19 (27.5%)	+ 8.2%	n.s.
Disinhibition	0 (0%)	15 (21.7%)	- 21.7%	0.000012
Disorientation/Navigation	2 (2.9%)	4 (5.8%)	- 2.9%	n.s.
Dysarthria	24 (34.3%)	12 (17.4%)	+ 16.9%	0.033
Dysexecutive Syndrome	22 (31.4%)	26 (37.7%)	- 6.3%	n.s.
Dysphagia /Choking	22 (31.4%)	8 (11.6%)	+ 19.8%	0.007
Dystonia	7 (10%)	23 (33.3%)	- 23.3%	0.001
Falls	23 (32.9%)	14 (20.3%)	+ 12.6%	n.s.
Gait abnormality	38 (54.3%)	35 (50.7%)	+ 3.6%	n.s.
Hallucinations	1 (1.4%)	3 (4.3%)	- 2.9%	n.s.
Hypersomnia	1 (1.4%)	0 (0%)	+ 1.4%	n.s.
Insomnia	24 (34.3%)	8 (11.6%)	+ 22.7%	0.002
Irritability	23 (32.9%)	14 (20.3%)	+ 12.6%	n.s.
Limb weakness	0 (0%)	4 (5.8%)	- 5.8%	n.s.
Loss of Empathy	0 (0%)	3 (4.3%)	- 4.3%	n.s.
Lost Reflexes	0 (0%)	0 (0%)	0.0%	N/A - constant
Memory Loss	27 (38.6%)	27 (39.1%)	-0.6%	1
Myoclonus	3 (4.3%)	6 (8.7%)	- 4.4%	n.s.
Motor Neuropathy	0 (0%)	0 (0%)	0.0%	N/A - constant
Sensory Neuropathy	0 (0%)	3 (4.3%)	- 4.3%	n.s.
Nystagmus	0 (0%)	1 (1.4%)	- 1.4%	n.s.
Obsessive Behaviour	3 (4.3%)	5 (7.2%)	- 3.0%	n.s.
Ocular Range of Movement	3 (4.3%)	6 (8.7%)	- 4.4%	n.s.
Ocular Saccades	22 (31.4%)	16 (23.2%)	+ 8.2%	n.s.
Ocular Smooth Pursuit	17 (24.3%)	14 (20.3%)	+ 4.0%	n.s.
Pain	0 (0%)	3 (4.3%)	- 4.3%	n.s.
Paranoia	1 (1.4%)	4 (5.8%)	- 4.4%	n.s.

Characteristic or Symptom	HD (%HD)	HDPC (%HDPC)	ΔHD-HDPC	p-value (uncorrected Chi-square/t-test)
Rigidity	11 (15.7%)	8 (11.6%)	+ 4.1%	n.s.
Spasticity	0 (0%)	4 (5.8%)	- 5.8%	n.s.
Supranuclear Palsy	1 (1.4%)	5 (7.2%)	- 5.8%	n.s.
Sweet Tooth/Change in dietary habits	0 (0%)	3 (4.3%)	- 4.3%	n.s.
Tremor	3 (4.3%)	18 (26.1%)	- 21.8%	0.00031
Weight Loss	8 (11.4%)	5 (7.2%)	+ 4.2%	n.s.
Cognitive HDPC score	0.5	0.6	- 0.1	n.s.
Psychiatric HDPC score	0.7	0.6	0.0	n.s.
Motor HDPC score	0.9	0.9	0.0	n.s.
Total HDPC score (no FHx)	2.1	2.2	0.0	n.s.
Total HDPC+FH score (with FHx)	2.9	2.6	+ 0.3	n.s.

i. Differences and similarities between HD and HDPC syndromes

Based on this comparison between the HD and HDPC cohorts, no relevant family history, disinhibition, dystonia and especially tremor therefore seem to point towards an HDPC syndrome, while insomnia and early orolingual involvement with prominent dysarthria and dysphagia appear to be clues for an HD diagnosis. Rates of chorea were similar in both cohorts.

Given the variability in presentation, however, any single initial symptom does not offer much insight into the eventual diagnosis and future disease course. Comparing symptom patterns may add to the clinical picture, as can be seen from Figure III-10. Calculating the conditional probability of two given symptoms co-occurring in a given patient draws attention to commonalities and differences between the cohorts. Conditional probabilities express the likelihood of one event occurring given that another has already happened, i.e. that one symptoms occurs given the presence of another in any given patient. The likelihood must therefore be assessed both ways, with either symptom as the given condition (all statistical analyses were done in R version 3.5.1 (https://www.rproject.org/) using the base package and custom written scripts). While both HD and HDPC syndromes present with the typical symptom combination of a subcortical dementia, cognitive slowing and a dysexecutive

syndrome, the HD pattern of combining dysarthria with memory loss early on in the disease course (conditional probability 0.71/0.63) highlights its links between motor and cognitive symptoms; insomnia as an early symptom also appears suggestive of HD, especially when combined with patient reported memory loss. In addition HD patients appear to combine depression and irritability (conditional probability 0.56/0.61) more often than HDPC patients. HDPC patients typically show strong connections between related symptoms, such as disinhibition and irritability (conditional probability 0.64/0.6), cognitive slowing and executive dysfunction (conditional probability (0.77/0.65) and memory loss and a dysexecutive syndrome (conditional probability 0.69/0.67).

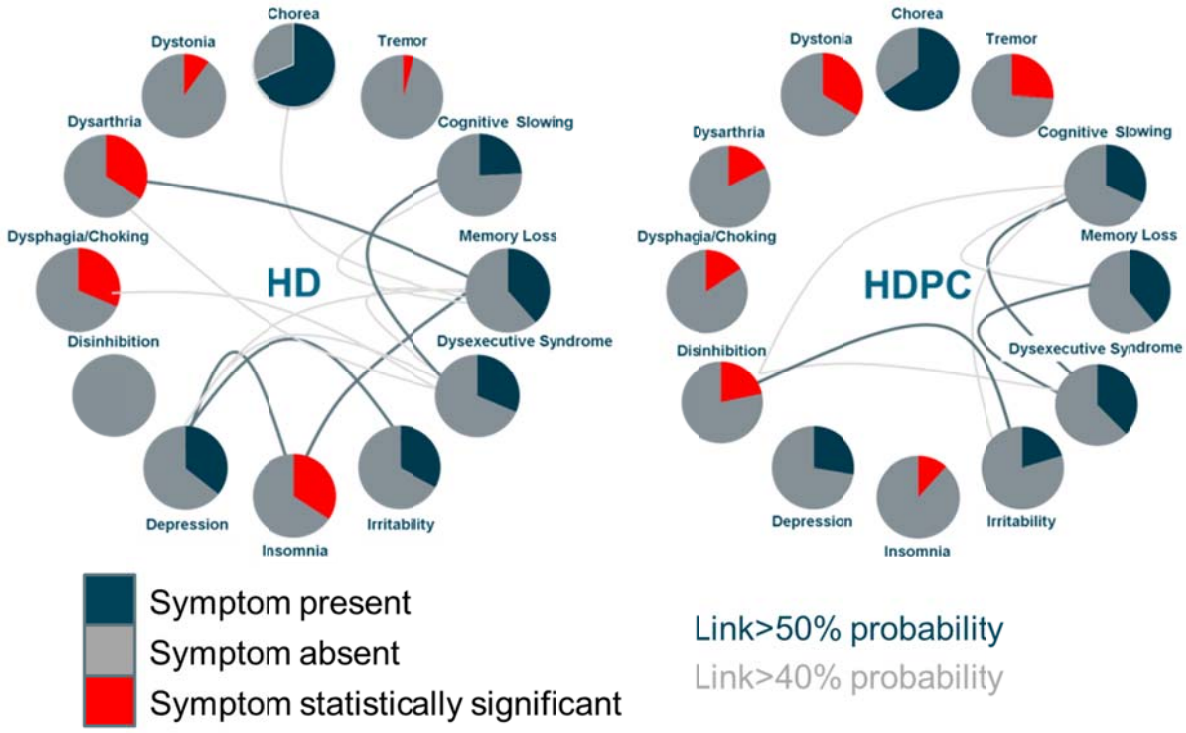

Figure III-10: HD and HDPC display different patterns of concurrent symptoms

Each symptom circle shows the percentage of patients in each group presenting with a given symptom; the circle is highlighted in red if the symptom differed significantly between the two groups ($p \leq 0.05$, logistical regression). Given the high number of included symptoms, none of the results withstood Bonferroni's correction for multiple comparisons, but they results are interesting from an exploratory vantage point. The likelihood of a patient presenting with any given symptom pair was calculated using conditional probabilities, which express the likelihood, that one symptom occurs given the presence of another in any given patient. If the likelihood is bilaterally equal or higher than 40%, a faint grey line connects the two symptoms, if it is equal or higher than 50%, the connecting line is emphasized in dark grey. All statistical analyses were done in R version 3.5.1 (https://www.rproject.org/) using the base package and custom written scripts.

ii. Analysis of HDPC cases from the Neurogenetics clinic only

If only the patients who presented to the Neurogenetics clinics were compared, the results remained similar. Average age at motor onset (mAAO) of patients ultimately diagnosed with HD was 44.8 years old, while for HDPC patients, it was 51 years old (Δ=6.2; p=n.s., Independent t-test), for HDPC patients psychiatric pAAO was 51 years, and cAAO was 55.8 years, compared to HD pAAO 38.1 years and cAA) 40 years ($p\Delta$=17.1; p=n.s and $c\Delta$=14.8; p=n.s., respectively). Significance levels for pAAO and

cAAO were severely affected by low availability of data for the HDPC cohort (4 cases for cAAO and 1 case for pAAO).
At initial presentation, the average HDPC score for all HDPC patients was 2.2/3 (Δ=0.82; p=n.s., Independent t-test) or 2.5/4 when including a positive family history (Δ=0.37; p=n.s., Independent t-test).

In this smaller cohort, all HDPC patients presented with motor symptoms compared to 92.9% of HD cases (HDPC motor score, 100% vs, 92.9% vs., respectively, HDPC score Δ=0.071; p=n.s., Chi-Square-test), as well as psychiatric (65.7% vs. 68.4%, respectively, HDPC score Δ=0.03; p=n.s., Chi-Square-test) and cognitive symptoms (54.3% vs. 52.6%, respectively, HDPC score Δ=0.02; p=n.s., Chi-Square-test). AAO and HDPC scores were compared as independent t-tests in SPSS to the same HD cohort described above. As examples of the commonest symptoms in each category, 68.6% of HD patients and 94.7% of HDPC patients had chorea (p=0.02, Chi-Square test), reflecting clinician's strong inclination to test for HD in patients with chorea even in more ambiguous cases (see survey results above). 35.7% of HD patients and 36.8% of HDPC patients had depression (p=n.s., Chi-Square test), and 31.4% of HD patients and 31.6% of HDPC patients had a dysexecutive syndrome (p=n.s., Chi-Square test). In the motor category, the most marked disparities were found in early dysarthria (29% commoner in HD, p=0.02., Chi-Square test) and dysphagia/choking (15.6% commoner in HD, p=n.s., Chi-Square test), dystonia (21.6% commoner in HDPC, p=0.03., Chi-Square test), bradykinesia (17.1% commoner in HD, p=n.s., Chi-Square test) and rigidity (15.7% commoner in HD, p=n.s., Chi-Square test), as well as tremor (11.5% commoner in HDPC, p=n.s., Chi-Square test).

In the clinic cohort, weight loss was commoner in HDPC than HD patients (HDPC 15.8%, Δ4.4%, p=n.s., Chi-Square test), but insomnia was more common in HD at 13.2% (p=n.s., Chi-Square test). Limb weakness, sensory neuropathy and spasticity were observed in HDPC but not in HD in this cohort. The largest difference in symptom prevalence for the cognitive category was loss of empathy at 5.3% (commoner in HDPC, p=n.s., Chi-Square test), followed by memory loss at 3.5% (also commoner in

HDPC patients, p=n.s., Chi-Square test). In the psychiatric category, the starkest differences were in disinhibition (15.8%, commoner in HDPC, p=0.009., Chi-Square test) and obsessive behaviour (11.5%, commoner in HDPC, p=n.s., Chi-Square test), followed by irritability (6.5%, commoner in HD, p=n.s., Chi-Square test), development of a “sweet tooth” (5.3%, commoner in HDPC, p=n.s., Chi-Square test), apathy and anxiety (both 4.7% and commoner in HD, both p=n.s., Chi-Square test).

The only significant differences in initial clinical presentation were disinhibition (0% in HD, 15.8% in HDPC, Δ=15.8%, p=0.009), chorea (68.6% in HD, 94.7% in HDPC, Δ=26.1%, p=0.02), dystonia (10.0% in HD, 31.6% in HDPC, Δ=21.6%, p=0.03), dysarthria (96.0% in HD, 4.0% in HDPC, Δ=92.0%, p=0.02), and spasticity (0% in HD, 10.5% in HDPC, Δ=10.5%, p=0.044); all calculated as Chi-Square tests.

Goldman score (GS) as a measure of family history was a distinguishing factor between the cohorts: HD patients had a GS1 in 48.6% of HD cases, a GS2 in 4.3%, a GS3 in 24.3%, a GS3.5 in 4.3%, a GS4 in 14.3%, and a GS4.5 in 4.3% of HD cases. Meanwhile, HDPC patients had a GS1 in 20.3% of HDPC cases, a GS2 in 5.8%, a GS3 in 4.3%, a GS3.5 in 17.4%, a GS4 in 47.8% and a GS4.5 in 4.3% of HDPC cases. Having a strong autosomal family history (GS1) and having no relevant family history was therefore almost completely reversed between the two cohorts. Running a logistical regression with GS4.5 (censored family history) as the constant, confirmed the overall significantly differentiating effect of the Goldman score (p= 0.021) between the HD and HDPC cohorts, but the various GS categories did not reach significance.

Table III-11: Characteristics of the Neurogenetics HD/HDPC cohort

Characteristic or Symptom	HD (%HD)	HDPC (%HDPC)	ΔHD-HDPC	P (uncorrected, Chi-Square test/ t-test)
N	70	19		
Male	33 (47.1%)	9 (47.4%)	0.2%	n.s.
Goldman Score (median)	2 (median)	4 (median)	2	0.02
Age at onset (Motor)	44.8 years	51 years	6.2 years	n.s.
Age at onset (Cognitive)	40 years	55.8 years	15.8 years	n.s.
Age at onset (Psychiatric)	38.1 years	51 years	12.9 years	0.08
Agitation	6 (8.6%)	1 (5.3%)	+ 3.3%	n.s.
Anxiety	18 (25.7%)	4 (21.1%)	+ 4.7%	n.s.
Apathy	7 (10%)	1 (5.3%)	+ 4.7%	n.s.
Asymmetry	1 (1.4%)	1 (5.3%)	- 3.8%	n.s.
Ataxia	12 (17.1%)	4 (21.1%)	- 3.9%	n.s.
Babinski	1 (1.4%)	0 (0%)	+ 1.4%	n.s.
Bladder / Bowel dysfunction	0 (0%)	0 (0%)	0.0%	n.s.
Bradykinesia	12 (17.1%)	0 (0%)	+ 17.1%	n.s.
Chorea	48 (68.6%)	18 (94.7%)	- 26.2%	0.02
Cognitive slowing	17 (24.3%)	5 (26.3%)	- 2.0%	n.s.
Cortical Blindness	0 (0%)	0 (0%)	0.0%	N/A constant
Delusions	2 (2.9%)	0 (0%)	+ 2.9%	n.s.
Depression	25 (35.7%)	7 (36.8%)	- 1.1%	n.s.
Disinhibition	0 (0%)	3 (15.8%)	- 15.8%	n.s.
Disorientation / Navigation	2 (2.9%)	1 (5.3%)	- 2.4%	n.s.
Dysarthria	24 (34.3%)	1 (5.3%)	+ 29.0%	0.02
Dysexecutive syndrome	22 (31.4%)	6 (31.6%)	- 0.2%	n.s.
Dysphagia / Choking	22 (31.4%)	3 (15.8%)	+ 15.6%	n.s.
Dystonia	7 (10%)	6 (31.6%)	- 21.6%	0.03
Falls	23 (32.9%)	5 (26.3%)	+ 6.5%	n.s.
Gait abnormality	38 (54.3%)	10 (52.6%)	+ 1.7%	n.s.
Hallucinations	1 (1.4%)	0 (0%)	+ 1.4%	n.s.
Hypersomnia	1 (1.4%)	0 (0%)	+ 1.4%	n.s.
Insomnia	24 (34.3%)	4 (21.1%)	+ 13.2%	n.s.
Irritability	23 (32.9%)	5 (26.3%)	+ 6.5%	0.01

Characteristic or Symptom	HD (%HD)	HDPC (%HDPC)	ΔHD-HDPC	P (uncorrected, Chi-Square test/ t-test)
Limb weakness	0 (0%)	1 (5.3%)	- 5.3%	n.s.
Loss of empathy	0 (0%)	1 (5.3%)	- 5.3%	n.s.
Lost reflexes	0 (0%)	0 (0%)	0.0%	N/A constant
Memory loss	27 (38.6%)	8 (42.1%)	- 3.5%	n.s.
Myoclonus	3 (4.3%)	1 (5.3%)	- 1.0%	n.s.
Neuropathy (motor)	0 (0%)	0 (0%)	0.0%	N/A constant
Neuropathy (sensory)	0 (0%)	1 (5.3%)	- 5.3%	n.s.
Nystagmus	0 (0%)	0 (0%)	0.0%	N/A constant
Obsessive behaviour	3 (4.3%)	3 (15.8%)	- 11.5%	n.s.
Pain	0 (0%)	1 (5.3%)	- 5.3%	n.s.
Paranoia	1 (1.4%)	0 (0%)	+ 1.4%	n.s.
Range of Movement	3 (4.3%)	3 (15.8%)	- 11.5%	n.s.
Rigidity	11 (15.7%)	0 (0%)	+ 15.7%	n.s.
Saccades	14 (22.6%)	0 (0%)	+ 22.6%	n.s.
Smooth pursuit	11 (17.2%)	4 (25%)	- 7.8%	n.s.
Spasticity	0 (0%)	1 (5.6%)	- 5.6%	0.04
Supranuclear palsy	1 (1.4%)	0 (0%)	+ 1.4%	n.s.
Sweet Tooth / Change in dietary habits	0 (0%)	1 (5.3%)	- 5.3%	n.s.
Tremor	3 (4.3%)	3 (15.8%)	- 11.5%	n.s.
Weight loss	8 (11.4%)	3 (15.8%)	- 4.4%	0.06
Cognitive HDPC score	0.5	0.5	0.0	n.s.
Psychiatric HDPC score	0.7	0.7	0.0	n.s.
Motor HDPC score	0.9	1.0	0.1	n.s.
Total HDPC score (no FHx)	2.1	2.2	0.1	n.s.
Total HDPC+FH score (with FHx)	2.9	2.5	0.4	n.s.

iii. Differences and similarities between HD and HDPC syndromes

Comparing only the patients who presented to the Neurogenetics clinics shifts the magnitudes of the differences between HD and HDPC cases and of the symptoms present at the same time (Figure III-11), reflecting the heterozygous nature of HDPC syndromes. However the overall pattern observed in the clinical presentation of each cohort remains the same.

While both HD and HDPC present with a subcortical dementia including cognitive slowing and a dysexecutive syndrome, and early orolingual involvement with dysarthria and dysphagia in combination with memory loss appears to point towards HD (see above). In contrast, and also in the smaller cohort HDPC patients appear to combine symptoms that are related to each other, such as irritability and disinhibition (conditional probability 0.60/0.64) and memory loss and a dysexecutive syndrome (conditional probability 0.83/0.63). However, it should be noted as an important that in this smaller cohort, chorea is more common in HDPC patients, and frequently found in the presence of memory loss (conditional probability 1/0.44).

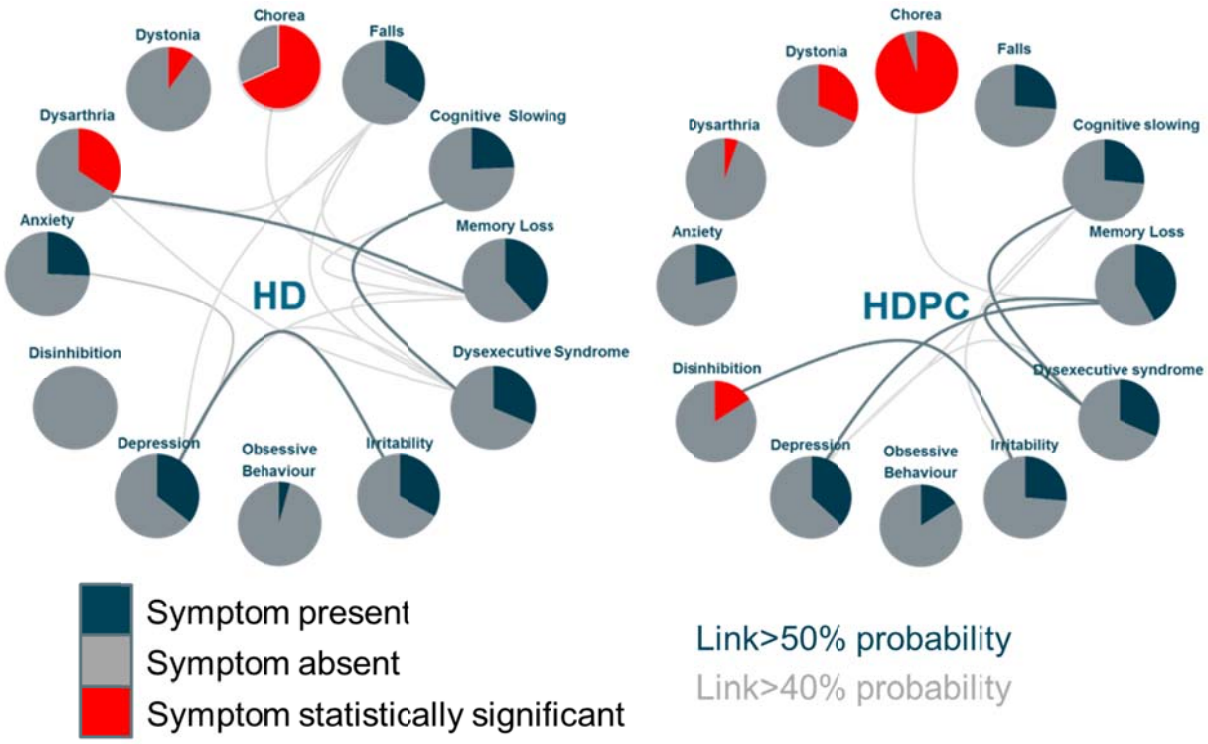

Figure III-11: In the smaller cohort, symptom patterns of HD and HDPC patients

When only the patients presenting to the Neurogenetics clinics are compared, different symptoms are put into focus, reflecting the heterozygous nature of the HDPC syndromes. Compared to Figure III-10, Falls, Obsessive Behaviour, and Anxiety have been added to the graph, while Insomnia, Dysphagia, and Tremor have been removed, to reflect larger differences. The general pattern remains the same: in HDPC patients, links between related symptoms remain strong, while in HD patients, more diverse symptoms are more likely to be present at the same time. Each symptom circle shows the percentage of patients in each group presenting with a given symptom (highlighted in red if $p \leq 0.05$, logistical regression). None of the results withstood Bonferroni's correction for multiple comparisons, but they results are interesting from an exploratory vantage point. The likelihood of a patient presenting with any given symptom pair was calculated using conditional probabilities: if the likelihood is bilaterally ≥ 40%, a faint grey line connects the two symptoms, if it is ≥ 50%, the connecting line is emphasized in dark grey. All statistical analyses were done in R version 3.5.1 (https://www.rproject.org/) using the base package and custom written scripts.

c. Distinguishing factors between HD and HDPC syndromes

The expectations of HD expert clinicians reflected most of the results gathered from the patient analysis. In the larger cohort, the presence of chorea was comparable in both HD and HDPC syndromes and the difference was not significant. Expert clinicians considered chorea a strongly suggestive symptom for HD (although this does not mean they would not consider it typical of other conditions), but it was not a distinguishing factor for patients referred for genetic testing. In fact, chorea was even more common

in HDPC patients in the smaller cohort, and frequently found in combination with memory loss. Similarly, cognitive slowing, irritability, and gait abnormalities/falls were indistinguishable between HD and HDPC syndromes. Dystonia, which was cited as being an indicator of HD, was indeed significantly more common in HDPC syndromes, as was tremor (considered more typical of HDPC syndromes), but neither precluded an HD diagnosis. Despite clinicians' expectations, neither limb weakness, ataxia, neuropathy, nor pain, differed significantly between the cohorts, although this may be due to low prevalence in the study population.

Based on single symptoms, HD and HDPC were difficult to distinguish clinically; differences were minor and in some instances could also reflect differences in timing of presentation or documentation. Eye movements, for example, were usually not documented in the patient records and therefore do not figure in these results. Twenty-five years after the discovery of the *HTT* expansion as the cause of HD, one of the strongest indicators of an HDPC syndrome appears to be the lack of a relevant family history, or indeed any family history of cognitive decline, psychiatric problems or motor symptoms.

4 Chapter 4: Whole-genome sequencing of HD phenocopy syndromes

A. Descriptive statistics of the whole-genome sequencing dataset

50 HDPC samples cases were selected based on the scoring system described in *Materials and Methods: Stratification of the HD phenocopy cohort and sample selection for WGS* and were sent for WGS to the Edinburgh Genomics: Clinical Genomics cluster. 21 patients (42%) were male, with an average AAO of 51.2 years (range 11 years to 77 years), a median GS of 3.5 (GS1 – 15x, GS2 – 4x, GS3 – 1x, GS3.5 – 12x, GS4 – 17x, GS4.5 – 1x), and a mean HDPC score of 2.86/4. 92% of patients presented with a movement disorder (for 56% this included chorea), 76% had cognitive decline and 54% had psychiatric symptoms (behavioural and personality change, agitation, disinhibition, bipolar affective disorder and depression, hallucinations and delusions, and obsessive behaviour). Details and scores for the selected 50 patients are listed in Table III-12. The average coverage per sample was 39.7x.

It is important to remember that the HDPC score is a tool to indicate how HD-like a patient's clinical presentation is in terms of presenting with symptoms across the typical three motor, cognitive, and psychiatric domains. Due to the wide range of potential HD symptoms, especially early on in the disease course, points can be attributed for a variety of symptoms in each category, but these are not themselves weighted as to how HD-like they are, nor are they additive if a patient has more than one symptom per domain. For example one point is awarded equally for chorea or dystonia in the motor domain, one point for executive dysfunction or memory loss in the cognitive domain, although the strength of association with HD of these symptoms is very different. However, since HD patients may plausibly present with any of these symptoms, a point is therefore awarded in the appropriate domain. The strength of the HDPC score lies in its scope covering all three symptom domains, which is one of the hallmarks of HD, and this was a mainstay of patient selection.

After the disease-causing mutation for one sample was discovered in a separate analysis (see following section *Gene expansion results using ExpansionHunter®*, the 49

remaining HDPC samples were uploaded to Ingenuity Variant Analysis (IVA™); for these, the number of variants per sample ranged from 5,230,677 to 6,331,549; on average, 5,334,908 variants were found in each sample using the IVA™ platform.

B. Power calculation and assessment of penetrance

In order to gauge the statistical power of this discovery dataset, I imagined a potentially causal variant to be observed a single time in the HDPC dataset and once on *GnomAD*[124]; using Fisher's exact test, calculated that the result would be statistically significant (p=0.0008). Although this calculation does not take into account multiple testing, it provides a benchmark as to the frequency of likely relevant variants in the discovery dataset.

Penetrance calculations using the same principles as for EOAD and EOFTD described above[24] and a disease prevalence of 2.5/100,000, puts the prevalence of any variant found once in the dataset and once on *GnomAD* (approx. 140,000 controls) at 7.1% (95% CI 0.22% - 218.28%), meaning only novel variants were considered for further analysis.

Table III-12: Details of patient samples selected for whole-genome sequencing

Based on the selection criteria described in *Materials and Methods Stratification of the HD phenocopy cohort and sample selection for WGS,* Table II-8. 50 samples were selected for whole-genome sequencing; their details are described in this table. Sex, age at onset (AAO) and the strength of the family history (modified Goldman score, see Figure II-1) were noted, and the clinical notes scoured for symptoms the patients developed in their lifetime, as well as the last time they presented in clinic at the National Hospital for Neurology and Neurosurgery (NHNN), Queen Square, London . Based on this information and their respective score, patient samples were shortlisted and selected.

HDPC Sample	Sex	AAO	Goldman Score	HDPS Movement Disorder	HDPS Cognitive Decline	HDPS Psychiatric Disturbance	HDPS Total	Neuro-pathology available	Chorea present in patient	Last Clinic within (years)	Scoring Total
1	Female	43	1	1	1	1	3	no	yes	2	85
2	Female	55	1	1	1	1	3	no	yes	2	85
3	Male	59	1	1	1	1	3	no	no	5	81
4	Male	61	1	1	1	0	2	yes	no	2	80
5	Female	69	2	1	1	1	3	no	yes	2	75
6	Female	45	2	1	1	1	3	no	yes	2	75
7	Male	50	3.5	1	1	1	3	no	yes	2	67
8	Female	49	3.5	1	1	1	3	no	yes	2	67
9	Male	59	4	1	1	1	3	no	yes	2	65
10	Male	53	4	1	1	1	3	no	no	2	65
11	Male	70	4	1	1	1	3	no	yes	2	65
12	Female	39	1	1	1	0	2	no	yes	2	65
13	Female	49	4.5	1	1	1	3	no	no	2	65
14	Male	43	4	1	1	1	3	no	yes	2	65
15	Male	64	4	1	1	1	3	no	yes	2	65
16	Female	11	1	1	0	1	2	no	yes	2	65
17	Male	43	3.5	1	1	1	3	no	no	3	64

HDPC Sample	Sex	AAO	Goldman Score	HDPS Movement Disorder	HDPS Cognitive Decline	HDPS Psychiatric Disturbance	HDPS Total	Neuro-pathology available	Chorea present in patient	Last Clinic within (years)	Scoring Total
18	Male	53	3.5	1	1	1	3	no	no	53	63
19	Male	39	3.5	1	1	1	3	no	no	-	62
20	Male	47	4	1	1	1	3	no	no	3	62
21	Female	43	1	0	1	1	2	no	no	-	60
22	Female	69	4	1	1	1	3	no	yes	-	60
23	Male	33	4	1	1	1	3	no	yes	-	60
24	Female	34	2	1	0	1	2	no	no	2	55
25	Female	55	2	1	1	0	2	no	no	2	55
26	Male	75	3.5	1	1	0	2	no	yes	2	47
27	Female	63	3.5	1	1	0	2	no	yes	2	47
28	Female	66	3.5	1	0	1	2	no	yes	2	47
29	Male	44	3.5	1	1	0	2	no	no	2	47
30	Female	62	3.5	1	1	0	2	no	no	2	47
31	Female	34	3.5	1	0	1	2	no	yes	2	47
32	Female	70	3.5	1	0	1	2	no	yes	2	47
33	Female	65	1	1	1	0	2	no	no	5	46
34	Male	16	4	0	1	1	2	no	no	2	45
35	Female	68	4	1	0	1	2	no	yes	2	45
36	Female	46	3	1	1	0	2	no	no	5	45
37	Male	77	4	1	1	0	2	no	yes	2	45
38	Male	51	4	1	1	0	2	no	yes	2	45
39	Male	64	4	1	1	0	2	no	yes	2	45
40	Female	53	4	1	1	0	2	no	yes	2	45

HDPC Sample	Sex	AAO	Goldman Score	HDPS Movement Disorder	HDPS Cognitive Decline	HDPS Psychiatric Disturbance	HDPS Total	Neuro-pathology available	Chorea present in patient	Last Clinic within (years)	Scoring Total
41	Female	46	4	1	1	0	2	no	no	2	45
42	Female	53	4	1	1	0	2	no	yes	2	45
43	Female	58	4	1	1	0	2	no	no	2	45
44	Female	22	1	1	0	0	1	no	yes	2	35
45	Female	52	1	1	0	0	1	no	yes	2	35
46	Female	52	1	1	0	0	1	no	yes	2	35
47	Male	40	1	1	0	0	1	no	no	2	35
48	Male	67	1	0	1	0	1	no	no	2	35
49	Female	35	1	1	0	0	1	no	no	-	30
50	Female	45	1	0	0	0	0	0	no	5	21

C. Gene expansion results using ExpansionHunter®

ExpansionHunter® is a computer algorithm which seeks to identify known repeat expansions using conventional WGS data (see *Materials and Methods: ExpansionHunter®*); it correctly identified a CAG expansion in the Ataxin 1 (*ATXN1*) gene, as can be seen in Figure III-12; this expansion has been confirmed in the Department for Neurogenetics at the National Hospital for Neurology and Neurosurgery (NHNN), Queen Square, London. CAG expansions of more than 39 repeats in exon 1 of the *ATXN1* gene, located on chromosome 6p23 and inherited in an autosomal dominant fashion, cause spino-cerebellar ataxia type 1 (SCA1), which is a known to overlap clinically with HD[103,202]. The *ATXN1* gene encodes a transcriptional corepressor involved in transcription regulation, cell specification and synaptic activity[202]. The male patient's symptoms started at the age of 40 with gait and coordination problems; he had a family history of similar problems and some cognitive decline. Symptoms, family history and gene expansion are consistent with a diagnosis of SCA1, which can cause parkinsonism, dystonia and chorea. No other expansions in the genes known to be linked to expansion disorders were identified in any other samples.

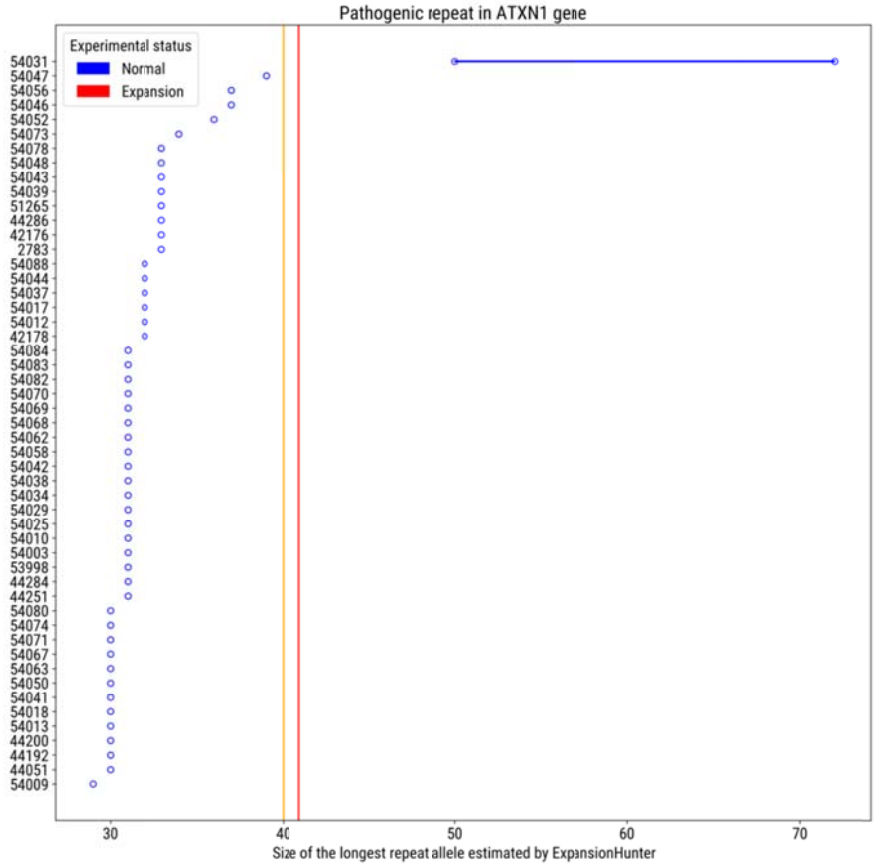

Figure III-12: ExpansionHunter® identifies a repeat expansion in the ATXN1 gene

ExpansionHunter® correctly identified an expansion in the *ATXN1* gene, as can be observed on this output from the programme. All but one sample were estimated to have repeat sizes well below the 39 CAG repeats found in patients with spino-cerebellar ataxia type 1 (SCA1).

D. Ingenuity Variant Analysis®

a. Tolerance score- and gene expression-based analysis

For this analysis, variants and genes were filtered as described in Figure II-3 and *Materials and Methods: Gene- and expression-based analysis*. In addition to the IVA™ filter pipeline, the resulting gene list was manually curated using the RVIS score and GTex expression data to focus on genes intolerant to genetic variation and highly expressed in brain. This resulted in 26 genes with variants with a CADD score >25, low tolerance to genetic variability and high brain tissue expression being retained (see Table III-13Table III-13). The 3 genes most intolerant to genetic variation based on the RVIS score - all in the top 2% most intolerant genes - were *APC*, *CREBBP*, and *MAPK8IP3*.

Two missense variants (in two separate cases), which were absent from GnomAD, were highlighted in Adenomatous polyposis coli (*APC*) (p.S52R and p.A1991T); while *APC* is best known for the role it plays in some familial colon cancers, the tumour suppressor it encodes is an antagonist of the WNT signalling pathway. *APC* is highly expressed in the CNS and among the 0.5% of genes least tolerant of genetic variation. When *APC* causes colon cancer, this is usually due to truncating mutations located in a small region of the gene. WNT signalling has been shown to be downregulated in Alzheimer's disease, but upregulated in other neurodegenerative disorders such as HD and Parkinson's disease (PD)[203,204]. This makes this an interesting finding, all the more as the WNT pathway and its inhibition via *APC* may be a future target in colon cancer therapy.

Furthermore, two exonic and two intronic variants in 4 cases were found in CREB Binding Protein (*CREBBP*), which is among the 0.6% genes least tolerant of genetic variation and highly expressed in brain (see Table III-13Table III-13). One of the variants (p.A1062V) is absent from GnomAD and located in a mutational hotspot; *CREBBP* haploinsufficiency has been shown to be deleterious. The other variants were p.G2200E, c.3256-4dupT, and c.3256-4delT; they were all absent from GnomAD. The *CREBBP* protein has been shown to interact with the Huntington's disease protein[205] and to facilitate its localization to neuronal intranuclear inclusions in a mouse model of HD; in this model of HD the HD protein also affects the transcription of p53-regulated promoters. Reduced *CREBBP* activity is also associated with long-term memory deficits in HD[206] and the mutant HD protein seems to induce cell death in cell culture through transcriptional disruption of cell survival programmes via *CREBBP*[207]

Table III-13: Genes with exonic variants with a CADD score >25, low tolerance to genetic variability and high brain tissue expression

After filtering as described in the Methods section "Gene- and expression based analysis", 26 genes with variants with a CADD score >25, low tolerance to genetic variability and high brain tissue expression were retained and are listed here, in order of increasing tolerance to genetic variation as per the RVIS score and decreasing levels of expression in brain tissue as per the GTEx database[181]. *RIVS* = residual variation intolerance score; *ExAC* = Exome Aggregation Consortium; *RVIS – ExAC_ 0.05%popn* = RVIS based on setting the common minor allele frequency filter at 0.05% of the population on ExAc; *LoF-FDR [ExAC]* = measure expressing rarity of loss of function (LoF) variants on ExAC, with a lower value suggesting intolerance to LoF mutations; *TPM* = transcripts per million reads; *GTEx* = Genotype-Tissue Expression

Gene	#Variants	#Cases	RVIS – ExAC_ 0.05%popn	RVIS LoF-FDR [ExAC]	Highest Median TPM Brain on GTex
APC	2	2	-4.0304 (0.4%)	1.11E-22	360.93
CREBBP	4	4	-3.5495 (0.6%)	1.40E-20	47.51
MAPK8IP3	9	13	-2.9988 (1.1%)	1.03E-10	412.3
SLC12A5	4	4	-1.4557 (6.3%)	0.0000363	204.6
HDLBP	2	2	-1.2016 (9.7%)	5.48E-08	62.73
ATN1	10	14	-1.1970 (9.8%)	0.0000206	506.7
MED15	3	4	-1.0735 (11.6%)	0.000272718	49.6
SEPT2	2	2	-0.6960 (21.1%)	0.03905515	197.4
STAT5B	2	2	-0.6598 (22.1%)	0.000699727	109.3
EPHB6	1	6	-0.5681 (25.1%)	0.000505184	142.1
CSNK2A1	4	6	-0.3978 (31.6%)	0.002716734	42.69
CABP1	1	6	-0.1256 (43.5%)	0.3499545	226.8
RPL14	1	2	0.4669 (70.5%)	0.01900458	547.2
PRKCSH	2	9	0.5552 (73.8%)	0.006809953	129.3
JPH3	4	3	0.6697 (77.7%)	0.000353968	116
NUBP2	1	2	1.0134 (86.6%)	0.4860693	42.35
LOC283710	1	2	N/A	N/A	-
MT-ATP6	3	3	N/A	N/A	$6.936 \times 10^{+4}$

Gene	#Variants	#Cases	RVIS – ExAC_ 0.05%popn	RVIS LoF-FDR [ExAC]	Highest Median TPM Brain on GTex
MT-CO2	2	2	N/A	N/A	$7.966 \times 10^{+4}$
MT-CO3	6	6	N/A	N/A	$7.060 \times 10^{+4}$
MT-CYB	4	4	N/A	N/A	$4.49 \times 10^{+4}$
MT-ND1	5	5	N/A	N/A	$4.323 \times 10^{+4}$
MT-ND2	3	3	N/A	N/A	$4.845 \times 10^{+4}$
MT-ND5	4	4	N/A	N/A	$1.127 \times 10^{+4}$

9 variants in 13 cases (6 exonic, two intronic, and one splice site) were detected in *MAPK8IP3*, which is highly expressed in the brain and among the 1.1% least tolerant genes when it comes to genetic variability (see Table III-13). This included a stop gain variant (p.K584*; CADD score 40) found in 4 cases, one in-frame mutation (p.R307_R309del) found in one case, and 5 frame shift mutations (p.T357fs*30, p.I486fsfs*13, p.P581fs*45, p.D627fs*43, and p.E1316fs), observed 4 times, 5 times, once, twice and once, respectively; all are absent from GnomAD. Mitogen-Activated Protein Kinase 8 Interacting Protein 3 (*MAPK8IP3*) is involved in axonal and synaptic vesicle transport and may regulate kinases of the JNK signalling pathway, thereby acting as a scaffolding protein in neurons; a heterozygous stop gain mutation (p.Tyr37Ter), albeit earlier in the gene than in our case, has been identified in a patient with infant onset developmental delay[208]. The frequency of these heterozygous LoF variants in this dataset, however, diminishes the likelihood that these variants are Mendelian causes of disease; although LoF variants in *MAPK8IP3* are rare on GnomAD (23 total, only three with an allele count over 5 on GnomAD, two observed over 250 times). They may be a contributing factor.

Four variants in four cases were observed in solute carrier family 12 member 5 (*SLC12A5*), which is among the 7% genes least tolerant of genetic variation and highly expressed in brain (see Table III-13). These were three insertions (p.G85dup twice. and p. G84_85dup) and one deletion (p.G85del), which were all located in a repeat region of the relevant codon (GGA). They are absent from GnomAD and predicted to cause LoF and a protein length change. *SLC12A5* plays a role in ion homeostasis in neurons; in mice, LoF in this gene was shown to increase hippocampal hyperexcitability, but

these variants are unlikely to cause disease. Two LoF variants (p.C43fs*16, p.C43fs*23) were found in High Density Lipoprotein Binding Protein (*HDLBP*): they are absent from GnomAD, but no further information for classification was available. *HDLBP* encodes a lipo-protein which binds to high-density cholesterol; however, knockdown of *HDLBP* has been shown to have a protective effect in mice[209]. Ten indels were identified in Atrophin 1 (*ATN1*) (p.Q502dup, 2x p.Q501_Q502dup, 2x p.Q501_Q502del, 2x p.Q502del, p.Q500_Q502dup, p.Q498_Q502dup, p.Q498_Q502del) in 14 cases; they are all absent from GnomAD but located in a repeat region without a known function. While *ATN1* expansions cause SCA1, these indels are not part of this disease mechanism. A further three variants were found in Mediator Complex Subunit 15 (*MED15*) (p.Q191dup, p.Q191del, p.Q191dup) in four cases, which has been linked to several cancers; *MED15* is in the 12% of genes least tolerant to genetic variation and is expressed in brain (see Table III-13). The variants are absent from GnomAD but are all located in the same codon, in a region with a variable number of CAG codons and without further evidence for pathogenicity. One insertion and one deletion in two cases were found in Septin 2 (*SEPT2*) (p.C43fs*16, p.C43fs*23), which is in the 25% of genes least tolerant to genetic variation and highly expressed in brain. *SEPT2* is a cytoskeletal GTPase and scaffolding protein, which interacts with microtubules[210]. The frameshift mutations are predicted to cause a heterozygous LoF and are absent from GnomAD, but haploinsufficiency in this gene is not known to cause disease.

Two variants in two cases were identified in Signal Transducer And Activator Of Transcription 5B (*STAT5B*); they were located in critical and well established functional domains without benign moderation (p.A693T - (SH2) and p.V467A (STAT-DBD)), are absent from GnomAD and predicted pathogenic *in silico* with CADD score of 23.5 and 26.1, respectively. *STAT5B* is also in the 25% of genes least tolerant to genetic variation and highly expressed in brain (see Table III-13); it encodes a transcription factor and part of the JAK2 pathway, which is essential for microglia[211].

All other variants were not in the top 25% of genes least tolerant to genetic variation. One variant was found in in six cases EPH Receptor B6 *(EPHB6)* (p.S177dup), which is

linked to neurofibromatosis and highly expressed in brain (see Table III-13). While the duplicated codon is not found o GnomAD, it is repeated several times in the adjacent codons and there is little evidence for its pathogenicity. Four non-coding variants in six cases were identified in Casein Kinase 2 Alpha 1 (*CSNK2A1*) (c.*34-4817_*34-4816dupTT, c.*34-4816dupT, c.*34-4816dupT, c.*34-4816delT), which is moderately expressed in brain (see Table III-13) and linked to Okur-Chung neurodevelopmental syndrome. While these variants are absent from GnomAD, they are all located in the same codon and there is little evidence for their pathogenicity. Calcium Binding Protein 1 (*CABP1*) (p.A7_C8del) was absent from GnomAD and causes a change in protein length but was predicted benign *in silico*, while *LOC283710* (p.R26fs*11) was located in an uncharacterized region; similarly, Ribosomal Protein L14 (*RPL14*) (p.A155_A159dup) was likely benign based on *in silico* predictions and because it is located in a repetitive region without a known function, although it is absent from GnomAD. Two variants in 9 cases (p.E325dup, p.E324_E325del) were observed in Protein Kinase C Substrate 80K-H (*PRKCSH*), which is highly expressed in brain (see Table III-13). While the variants were absent from GnomAD and predicted to cause a LoF, they were located in the same codon and too frequent to be Mendelian causes of disease. Of four variants identified in Junctophilin-3 (*JPH3*), two (p.A155_A157del and p.A156_A157dup) were recently reported benign[84,212], while p.A155_A157dup and p.A157del were absent from GnomAD; all variants caused a protein length change, but again were located in the same codon. One variant in two cases was found in Nucleotide Binding Protein 2 (*NUBP2)* (p.F105fs*?), which is moderately expressed in brain; it causes a frameshift and is absent from GnomAD, but little more information is available to classify it.

In addition, 25 mitochondrial gene variants in 25 cases were also detected. Mutations in mitochondrial genes can cause very severe but variable neurological disease phenotypes; all these genes are extremely highly expressed in brain (see Table III-13) – as can be expected from an energy-demanding organ – and associated with neurological and neurodegenerative disorders, such as ataxia and polyneuropathy, infantile bilateral striatal necrosis, AD, dystonia and Leigh syndrome. However, little

information is available regarding the population frequency tolerance to genetic variation. Mitochondrial gene expression is altered in Alzheimer's disease[213] in most of the genes that were found to carry mutations in this study. Some variants, for which more evidence of pathogenicity was available, are described in more detail in the next section *Variant-based analysis*.

Of the variants described above, *APC* (p.S52R and p.A1991T), *CREBB* (p.A1062V), *MAPK8IP3* (p.K584*) and *STAT5B* (p.A693T and p.V467A) are high-impact variants likely to affect gene-function. *APC* interference could upregulate the WNT signalling pathway (as it is in HD and PD), *CREBBP* disturbance may affect HD protein localization and p53-regulated promoters, while disruption of *MAPK8IP3* may affect vesicle transport, and *STAT5B* disruption could interfere with microglia function.

b. Gene- and variant-based analysis

In addition to the gene-expression and genetic variation intolerance analysis described above, variants most likely to have a deleterious impact on relevant genes were explored as explained in *Materials and Methods: Variants most likely to be pathogenic*. Variants were filtered through the pipeline described in Figure II-3, screened for genes and variants linked to a pertinent neurological disease, with evidence for the pathogenicity of a particular variant, classified automatically as pathogenic based on the ACMG criteria, as well as variants affecting a splice site, causing a start loss or stop gain, or being predicted damaging *in silico*. Variants which were synonymous, observed more than three times or only once as a single variant in a gene in the HDPC dataset, located in the 20% genes most tolerant of genetic variation, or in genes expressed at less than 40TPM according to GTeX, were excluded. Identified genes were all were absent from GnomAD and were shortlisted using the ToppGene algorithm[214]; they are listed in Table III-14.

Cathepsin B (*CTSB*) c.422-1G>A is heterozygous variant with a CADD of 24.1 and absent from controls; it affects a splice site and decreases splicing efficiency[163], leading to a frameshift and LOF in a gene where it is a known mechanism of disease. *CTSB* is involved in the lysosomal storage pathway and downregulation has been shown to be

a disease modifier in PD[215,216]; it was the gene most likely to be relevant according to the ToppGene shortlist.

Two variants were found in *CREBBP* (p.A1062V, p.G2200E), which were the same as described above; p.A1062V is located in the bromodomain and likely to be high-impact, while for G2200E little information is available. The two variants identified here in *APC* are also the same as described above; *APC* p. A1991T is located in the APC-basic domain where benign variation is not known, while p.S52R is predicted deleterious *in silico* but without further evidence for its pathogenicity.

Table III-14: Variants selected based on likely pathogenicity

Gene Symbol	Chr:Position	Variation Type	Gene Region	Transcript Variant	Protein Variant	Case Samples With Variant	Translation Impact	Classification	CADD Score	RVIS - Percentile ExAC_v2_0.05%popn	Highest Median TPM Gtex CNS	Neurological Disease	Top Gene	Gene # Cases
CTSB	8:11845790	SNV	Splice Site	c.422-1G>A; c.794-1G>A		1		Pathogenic	24.1			Y	1	2
CREBBP	16:3758924	SNV	Exonic	c.3185C>T	p.A1062V	1	missense	Uncertain Significance	25	0.60%	47.51	Y	2	4
CREBBP	16:3728334	SNV	Exonic	c.6713G>A	p.G2200E	1	missense	Uncertain Significance	22.8	0.60%	47.51	Y	2	4
APC	5:112842414	SNV	Exonic	c.6745G>A	p.A1991T	1	missense	Uncertain Significance	< 10	0.42%	360.93	Y	6	2
APC	5:112707873	SNV	5'UTR; Exonic	c.-28C>A; c.156C>A	p.S52R	1	missense	Uncertain 22.4 Significance		0.42%	360.93	Y	6	2
RANBP2	2:108719609	SNV	Exonic	c.3G>C	p.M1I	1	start loss	Pathogenic	23.5	1.77%	31.02	Y	20	2
RANBP2	2:108763792	SNV	Exonic	c.3253A>G	p.K1085E	1	missense	Uncertain Significance	< 10	1.77%	31.02	Y	20	2
CLCN7; LA16c-390E6.4	16:1450645	Insertion	Promoter; Exonic	c.1396_1397insACTCAGGTTAGAGTGT	p.L466fs*4	1	frameshift	Likely Pathogenic		2.45%	337.6	Y	47	2
CLCN7; LA16c-	16:1450601	SNV	Promoter; Exonic	c.1441T>C	p.Y481H	1	missense	Uncertain Significance	27.1	2.45%	337.6	Y	47	2

Gene Symbol	Chr:Position	Variation Type	Gene Region	Transcript Variant	Protein Variant	Case Samples With Variant	Translation Impact	Classification	CADD Score	RVIS - Percentile ExAC_v2_0.05%popn	Highest Median TPM Gtex CNS	Neurological Disease	Top p Gene	Gene # Cases
390E6.4														
ARHGAP32	11:128974462	SNV	Exonic	c.2693A>C	p.K824T	1	missense	Uncertain Significance	26.3	4.69%	25.23	Y	50	2
ARHGAP32	11:128972920	SNV	Exonic; 3'UTR	c.3544T>G	p.F1182V	1	missense	Uncertain Significance	17.88	4.69%	25.23	Y	50	2
MAPK8IP3	16:1767919	SNV	Splice Site	c.3502+1G>C *		1			< 10			Y	61	7
MAPK8IP3	16:1763658	Deletion	Exonic	c.1897delT	p.D633fs*43	2	frameshift			1.06%	412.3	Y	61	7
MAPK8IP3	16:1768776	Deletion	Exonic	c.3945_3946delGA	p.E1323fs?	1	frameshift			1.06%	412.3	Y	61	7
MAPK8IP3	16:1761245	Deletion	Exonic	c.1476delA	p.I493fs*13	1	frameshift			1.06%	412.3	Y	61	7
MAPK8IP3	16:1762869	Insertion	Exonic	c.1740_1741insT	p.P588fs*45	1	frameshift			1.06%	412.3	Y	61	7
MAPK8IP3	16:1761225	SNV	Exonic	c.1438G>T *	p.V487L	1	missense	Uncertain Significance	< 10	1.06%	412.3	Y	61	7
DENND4C	9:19300186	SNV	Splice Site	c.1167-1G>A		1			25.8			N	72	2
DENND4C	9:19336366	SNV	exonic	c.2686C>G	p.Q896E	1	missense	Uncertain Significance	23.3		.3	N	72	2
MT-ND2	M:5277	SNV	Exonic	m.5277T>C	p.F270L	1	missense	Likely	< 10	0.00%	48450	Y	82	2

Gene Symbol	Chr:Position	Variation Type	Gene Region	Transcript Variant	Protein Variant	Case Samples With Variant	Translation Impact	Classification	CADD Score	RVIS - Percentile ExAC_v2_0.05%popn	Highest Median TPM Gtex CNS	Neurological Disease	Top p Gene	Gene # Cases
								Pathogenic						
MT-ND2	M:4812	SNV	Exonic	m.4812G>C	p.V115L	1	missense	Uncertain Significance	15.47	0.00%	48450	Y	82	2
MT-ND1	M:3368	SNV	Exonic	m.3368T>C	p.M21T	1	missense	Likely Pathogenic	< 10	0.00%	43230	Y	117	2
MT-ND1	M:3472	SNV	Exonic	m.3472T>C	p.F56L	1	missense		23.5	0.00%	43230	Y	117	2
MT-ATP6	M:8887	SNV	Exonic	m.8887A>G	p.I121V	1	missense	Uncertain Significance	10.33	0.00%	69360	Y	120	2
MT-ATP6	M:8837	SNV	Exonic	m.8837T>C	p.M104T	1	missense	Uncertain Significance	21.9	0.00%	69360	Y	120	2
HAGH	16:1809796	SNV	Exonic	c.676C>T	p.T262I	1	missense	Uncertain Significance	27.6	62.23%	68.31	N	138	5
HAGH	16:1809835	SNV	Splice Site	c.639-2A>C		3			21.3			N	138	5
DOC2B	17:156240	SNV	Exonic	c.903C>G *	p.Y301*	3	stop gain		< 10	NA	38.4	N	141	20
DOC2B	17:156250	SNV	Exonic	c.893C>A *	p.A298D	1	missense	Uncertain Significance	< 10	NA	38.4	N	141	20
DOC2B	17:161443	SNV	Exonic	c.737T>G *	p.F246C	2	missense	Uncertain Significance	< 10	NA	38.4	N	141	20
DOC2B	17:161443	SNV	Exonic	c.737T>G	p.F246C	1	missense	Uncertain Significance	26.3	NA	38.4	N	141	20

Gene Symbol	Chr:Position	Variation Type	Gene Region	Transcript Variant	Protein Variant	Case Samples With Variant	Translation Impact	Classification	CADD Score	RVIS - Percentile ExAC_v2_0.05%popn	Highest Median TPM Gtex CNS	Neurological Disease	Top Gene	Gene # Cases
DOC2B	17:147460	SNV	Exonic	c.1220G>A	p.G407E	1	missense		< 10	NA	38.4	N	141	20
DOC2B	17:156320	SNV	Exonic	c.823A>C *	p.K275Q	2	missense	Uncertain Significance	< 10	NA	38.4	N	141	20
DOC2B	17:149159	SNV	Exonic	c.957A>T *	p.K319N	2	missense	Uncertain Significance	< 10	NA	38.4	N	141	20
DOC2B	17:162187	SNV	Exonic	c.532A>T	p.N178Y	1	missense	Uncertain Significance	26.6	NA	38.4	N	141	20
DOC2B	17:181384	SNV	Exonic	c.96G>C *	p.Q32H	3	missense	Uncertain Significance	< 10	NA	38.4	N	141	20
DOC2B	17:162177	SNV	Exonic	c.542G>C	p.R181T	2	missense	Uncertain Significance	27.3	NA	38.4	N	141	20
DOC2B	17:148226	SNV	Exonic	c.1049C>G *	p.S350C	2	missense	Uncertain Significance	< 10	NA	38.4	N	141	20
OGFOD1	16:56470770	SNV	Exonic	c.1264G>A	p.E379K	1	missense	Uncertain Significance	22.8	43.94%	17.23	N	181	2
OGFOD1	16:56470684	SNV	Exonic	c.1178G>A	p.S355N	1	missense	Uncertain Significance	23.9	43.94%	17.23	N	181	2
LA16c-312E8.4; CCDC154	16:1438049	Deletion	ncRNA; Splice Site	c.1178_1179+2delCTCT		2				-	11.39	N	254	2

Two variants were observed in RAN Binding Protein 2 (*RANBP2*) (p.M1I and p.K1085E), a gene linked to autosomal-recessive cause of mitochondrial complex I deficiency, and with an RVIS of 3.33% among the genes 10% least tolerant of genetic variation. p.M1I is classified as pathogenic and causes a start loss and LOF in a gene where loss of function (LOF) is a known mechanism of disease, while p.K1085E is predicted damaging *in silico*, but both heterozygous variants in a gene that is recessive for disease. *RANBP2* regulates the transport of proteins through the nuclear pore and attaches to the microtubule scaffolding[217]. Furthermore, p.L466fs*4 in Chloride Voltage-Gated Channel 7 (CLCN7) was classified as likely deleterious because it is a null variant in a gene where LOF is a known mechanism of disease, while p.Y481H is classified as a VUS and predicted damaging *in silico*. Both are compound heterozygous but found in different cases. *CLCN7* encodes a chloride channel and is linked to osteopetrosis[218], but no evidence links it to neurological disease.

Two variants in Rho GTPase Activating Protein 32 (*ARHGAP32*) (p.K824T and p.F1182V) were classified as VUS but predicted damaging *in silico*; p.K824T is suspected of affecting neurite growth and neuronal connectivity in HD[219,220], while little evidence is available for p.F1182V.

Six variants were identified in *MAPK8IP3*; homozygous c.3502+1G>C * affects a splice site expected to result in an in-frame insertion, a damaging effect[163] and LoF, while for p.D633fs*43, p.E1323fs?, p.I493fs*13, p.P588fs*45, and p.V487L little further evidence was available. As described above, *MAPK8IP3* is highly expressed in brain and intolerant to genetic variability; this splice-site variant is different to the stop gain variant p.K584* described above.

Another splice site variant was found in DENN Domain Containing 4C (*DENND4C*) (c.1167-1G>A), which while classified as a VUS, is predicted to result in insertion leading to a frameshift and a 5' Exon Extension[163]. *DENND4C* is involved in localizing glucose transporters to the plasma membrane[221]. Mutations in *RAB11B*, a member of the Member RAS Oncogene Family, which is part of the same pathway, have been linked to a neurodevelopmental syndrome with ataxia, but *DENND4C* itself has not been associated with Mendelian disease[222]. A missense variant (p.Q896E) predicted to

be damaging *in silico* and absent from controls was also identified in *DENND4C*, without any further information for classification available.

Two variants each were observed in three mitochondrial genes, which are all highly expressed in brain and associated with neurodegenerative disease. In Mitochondrially Encoded NADH: Ubiquinone Oxidoreductase Core Subunit 2 (*MT-ND2*), which is associated with optical neuropathy in children, p.F270L is a hemizygous gain of function (GoF) mutation classified as likely pathogenic, while p.V115L is absent from controls; both have been published as benign[223,224]. In Mitochondrially Encoded NADH: Ubiquinone Oxidoreductase Core Subunit 1 (*MT-ND1*), p.M21T was classified as likely pathogenic based on being a hemizygous LoF variant which is absent from controls and observed in cases of diabetes type 2 (DM2) but published as benign in Leigh syndrome[225], while p.F56L was predicted damaging in silico but publications are divided as to whether it is deleterious in Leber's Hereditary Optic Neuropathy[226,227]. In Mitochondrially Encoded ATP Synthase Membrane Subunit 6 (*MT-ATP6*), p.I121V and p.M104T were classified as VUS, but predicted damaging *in silico*; p.I121V has been linked to childhood-onset nystagmus and retinal degeneration[228] but has been published as benign for Leigh syndrome[229], while for p. M104T little additional information is available.

Two variants were observed in Hydroxyacylglutathione Hydrolase (*HAGH*), one causing the loss of a splice site[163] (c.639-2A>C, observed 3 times) and one predicted damaging *in silico* (p.T262I), but both classified as VUS. *HAGH* is involved in glucose metabolism and has been shown to be increased in CSF and plasma in AD[230].

In Double C2-like domain-containing protein beta (*DOC2B*), 11 variants were identified, with one (p.Y301*) causing a stop gain and LoF and 7 variants (p.A298D, 2x p.F246C, p.K319N, p.N178Y, p.Q32H, p.R181T, p.S350C) predicted damaging *in silico*; no additional information was available for classification of any of these variants. *DOC2B* encodes a calcium sensor which promotes vesicle priming and fusion and is downregulated in PD[231].

Two variants with CADD scores over 20 were observed in 2-Oxoglutarate And Iron-Dependent Oxygenase Domain-Containing Protein 1 (*OGFOD1*), one classified as benign (p.E379K) and one as a VUS (p.S355N), with little additional information available. *OGFOD1* controls translation and stress granule formation, which in turn is linked to neurodegenerative disease such as FTD and amyotrophic lateral sclerosis (ALS)[232,233]. In this cohort, a deletion in Coiled-Coil Domain Containing 154 (*CCDC154*) (c.1178_1179+2delCTCT, observed twice)) causes a splice site loss[163] and is absent from controls; *CCDC154* is involved in cell proliferation and downregulated in PD[234].

In summary, in this variant-based analysis, the same variants in functional domains in *CREBBP* and *APC* emerge as in the gene-expression and gene-tolerance based approach described above. In addition, a different highly suspicious variant in *MAPK8IP3* (found in one case) affecting a splice site was observed in this analysis; four more variants affecting splice sites were identified, of which the ones in *CTSB*, *HAGH* and *CCDC154* may have a disease-modifying effect but have not been proven to be Mendelian causes of disease.

c. IVA™ based analyses

In addition to the targeted analyses set out above, the dataset was examined for variants linked to neurological disease, novel variants predicted to be deleterious and novel variants with a link to either HD or the HD phenotype harnessing the IVA™ knowledge database and filters.

i. Variants associated with neurological disease

The 49 HDPC samples were filtered based on the pipeline set out in Figure II-3 and *Materials and Methods: IVA™ filters and analysis strategies: Links to neurological disease*; this returned 25 variants in 10 genes, which are listed in Table III-15. Variants were identified in *ATXN3, ATXN7, CACNA1A, GALC, MTR, PTRH1, SCN1A, SCN9A,* and *TBP*; three variants were predicted to be likely pathogenic based on the ACMG criteria (one in *SCN1A* and two in *ATXN3*), the others as of uncertain significance. All variants were absent from GnomAD.

SCN1A encodes the sodium voltage-gated channel alpha subunit 1, haploinsufficiency of which has been linked to epilepsy and neuropsychiatric comorbidities in infant mice[235], cognitive impairment without seizures[236], and age-related grey-matter and brain activity changes[237]. The corresponding sodium channel Na(v)1.1 has been shown to be reduced in a BACE1-null AD mouse model[238], while overexpression of Na(v)1.1 restored cognition in a human amyloid precursor protein (hAPP)-transgenic AD mouse model[239]. The likely deleterious intronic frameshift insertion (c.2177-11dupT) in *SCN1A* was identified in 14 cases, and despite the classification based on prevalence in cases but not in controls, has been shown not to increase epilepsy or hemiplegic migraine in humans and has been published as benign[240,241]. Three other *SCN1A* variants (p.A423D, c.2553+18_2553+19delAA, and c.2502+17_2502+19delAAA) were classified as VUS; however, p.A423D (found once) is predicted to be a LoF mutation with a CADD score of 28.9. It is located in a functional domain without benign variation and close to a published deleterious variant Y426N[242], is absent from GnomAD, and entails an amino-acid change from a small hydrophobic to a large negatively charged one. c.2553+18_2553+19delAA (observed 9 times), and c.2502+17_2502+19delAAA (observed 8 times) are both deletions in intron 14/15, close to exon 14. According to Human Splicing Finder, both mutations create new deep intronic exonic splicing enhancers (ESE) sites but do not affect splicing. Exon 14 and 15 partially encode the transmembrane segments 2-6 of domain II of the Na(v)1.1 channel, with intron 14/15 located in segment 4. In addition, one stop-gain variant was identified in Sodium Voltage-Gated Channel Alpha Subunit 9 (*SCN9A*) (p.C903*), which also affects the antisense gene *SCN1A-AS1*. It is a frameshift deletion likely to stop production of the protein from this copy of the gene; *SCN9A* is mainly associated with pain disorders, but other sodium channel mutations have been linked to ataxia and may also contribute to psychiatric disorders[243].

Ataxin 3 (*ATXN3*) expansions are known to cause spinocerebellar ataxia type 3 (SCA3); it has been implicated in the regulation of multiple transcriptional pathways relevant to SCA3[244]. *ATXN3* p.G185fs*40 (observed once) and p.G251fs*39 (observed twice) are classified as likely deleterious; p.G185fs*40 is a loss of function (LoF) mutation, while

p.G251fs*39 is a compound heterozygous case, which itself is not predicted to cause LoF but the case has another likely deleterious variant (p.A137_A143dup) on the other chromosome; both are absent from GnomAD. An additional LoF mutation (p.A92_G93insAAAAAAAAAAAAAAA, observed twice) and one predicted not to affect gene function (p.A92_G93insAAAAAAAAAAAAAAA, observed once) were identified in *ATXN3.* Both are located in a repetitive region without known function, but are absent from GnomAD. Since *ATXN3*-null mice do not display signs of obvious neurodegenerative disease[244], a gain-of-function (GoF) mechanism must be assumed for the *ATXN3* expansion disorder gene and LoF mutations may not be relevant.

Table III-15: Variants linked to neurological disease

Gene Symbol	Gene Region	Translation Impact	Chr:Position	Variation Type	Reference / Sample Allele	Transcript ID	Transcript Variant	Protein Variant	Inferred Activity	Case Samples With Variant	Classification	CADD Score
AC010127.1; SCN1A	Intronic		2:166041480-166041481	Insertion	/ A	ENST00000303395.8	c.2177-11dupT		normal	14	Likely Pathogenic	10.45
AC010127.1; SCN1A	Exonic; ncRNA; Intronic; 5'UTR	missense	2:166046879-166046879	SNV	G / T	ENST00000641603.1	c.1268C>A	p.A423D	loss of function	1	Uncertain Significance	28.9
AC010127.1; SCN1A	Intronic		2:166039406-166039407	Deletion	TT / -	ENST00000409050.1	c.2553+18_2553+19delAA		normal	9	Uncertain Significance	12.22
AC010127.1; SCN1A	Intronic		2:166039405-166039407	Deletion	TTT / -	ENST00000409050.1	c.2502+17_2502+19delAAA		normal	8	Uncertain Significance	12.14
ATXN3	Exonic; ncRNA; 3'UTR	frameshift	14:92071010-92071011	Insertion	/ TGCTGCTGCTGCTGCTGCTGCTGCTGCTG	ENST00000340660.10	c.750_751insCAGCAGCAGCAGCAGCAGCAGCAGCAGCA	p.G251fs*39	normal	2	Likely Pathogenic	-

Gene Symbol	Gene Region	Translation Impact	Chr:Position	Variation Type	Reference / Sample Allele	Transcript ID	Transcript Variant	Protein Variant	Inferred Activity	Case Samples With Variant	Classification	CADD Score
ATXN3	Exonic; ncRNA; 3'UTR	frameshift	14:92071010-92071011	Insertion	/ TGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	NR_028456.1	n.898_899insCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA	p.G185fs*40	loss of function	1	Likely Pathogenic	-
ATXN3	Exonic; ncRNA; 3'UTR	in-frame	14:92071010-92071011	Insertion	/ CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	NR_028456.1	c.942_943insCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG	p.A92_G93insAAAAAAAAAAAAAAA	loss of function	2	Uncertain Significance	-
ATXN3	Exonic; ncRNA; 3'UTR	in-frame	14:92071013-92071014	Insertion	/ CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	NR_028456.1	c.942_943insCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA	p.A92_G93insAAAAAAAAAAAAAAA	normal	1	Uncertain Significance	-

Gene Symbol	Gene Region	Translation Impact	Chr:Position	Variation Type	Reference / Sample Allele	Transcript ID	Transcript Variant	Protein Variant	Inferred Activity	Case Samples With Variant	Classification	CADD Score
					GCTGCTG CTGCTG		GCAGCAGC AG					
ATXN7; AC104162.2; THOC7	Promoter; 5'UTR; Intronic		3:63863806-63863807	Insertion	/ GCGGCG	ENST00000474112.5	c.-13_-8dupCGCCGC		normal	1	Uncertain Significance	18.9
ATXN7; SCAANT1	Promoter; Exonic	in-frame	3:63912687-63912688	Insertion	/ GCAGCG	ENST00000626439.1	c.94_95insGGCAGC	p.Q31_Q32insRQ	normal	1	Uncertain Significance	-
CACNA1A	Exonic; 3'UTR	in-frame	19:13207871-13207876	Deletion	CTGCTG / -	NM_001174080.1	c.6976_6981delCAGCAG	p.Q2313_Q2314del	loss of function	4	Uncertain Significance	21.6
CACNA1A	Exonic; 3'UTR	in-frame	19:13207874-13207876	Deletion	CTG / -	NM_001174080.1	c.6940_6942delCAG	p.Q2325del	loss of function	1	Uncertain Significance	21.6
CACNA1A	Exonic; 3'UTR	in-frame	19:13207862-13207864	Deletion	CTG / -	NM_001174080.1	c.6940_6942delCAG	p.Q2325del	loss of function	5	Uncertain Significance	15.7
CACNA1A	Exonic; 3'UTR	in-frame	19:13207861-13207862	Insertion	/ CTGCTG	NM_001174080.1	c.6970_6975dupCAGCAG	p.Q2325_Q2326dup	loss of function	1	Uncertain Significance	
CACNA1A	Exonic; 3'UTR	in-frame	19:13207864-13207865	Insertion	/ CTG	NM_001174080.1	c.6991_6993dupCAG	p.Q2327dup	loss of function	1	Uncertain Significance	-

Gene Symbol	Gene Region	Translation Impact	Chr:Position	Variation Type	Reference / Sample Allele	Transcript ID	Transcript Variant	Protein Variant	Inferred Activity	Case Samples With Variant	Classification	CADD Score
GALC; SHLD2P2	Promoter; Intronic		14:87950753-87950753	Deletion	A / -	ENST00000393568.8	c.1162-4delT		loss of function	3	Uncertain Significance	9.143
MTR	3'UTR		1:236897648-236897648	Deletion	T / -	NM_000254.2	c.*16delT		normal	2	Uncertain Significance	11.51
MTR	3'UTR		1:236897647-236897647	Deletion	T / -	NM_000254.2	c.*16delT		normal	9	Uncertain Significance	10.73
PTRH1; STXBP1	Exonic; 3'UTR	in-frame	9:127695062-127695063	Insertion	/ TGATGA	ENST00000335223.5	c.*73_*78dupATGATG	p.H96_H97dup	loss of function	1	Uncertain Significance	5.256
PTRH1; STXBP1	Exonic; 3'UTR	in-frame	9:127695065-127695066	Insertion	/ TGA	ENST00000335223.5	c.*76_*78dupATG	p.H97dup	loss of function	1	Uncertain Significance	-
SCN9A; AC010127.1; SCN1A-AS1	Exonic; ncRNA	frameshift	2:166277148-166277148	Deletion	A / -	NR_110260.1	n.1008delA	p.C903*	normal	1	Uncertain Significance	-
TBP	Exonic; Intronic	in-frame	6:170561967-	Insertion	/ CAG	ENST00000230354.10	c.219_221dupGCA	p.Q75dup	loss of function	2	Uncertain Significance	19.79

Gene Symbol	Gene Region	Translation Impact	Chr:Position	Variation Type	Reference / Sample Allele	Transcript ID	Transcript Variant	Protein Variant	Inferred Activity	Case Samples With Variant	Classification	CADD Score
			170561968									
TBP	Exonic; Intronic	in-frame	6:170561964-170561965	Insertion	/ CAGCAGCAG	ENST00000230354.10	c.213_221dupGCAGCAGCA	p.Q73_Q75dup	loss of function	1	Uncertain Significance	16.51
TBP	Exonic; Intronic	in-frame	6:170561961-170561962	Insertion	/ CAGCAGCAGCAG	ENST00000230354.10	c.270_281dupGCAGCAGCAGCA	p.Q72_Q75dup	loss of function	1	Uncertain Significance	-
TBP	Exonic; Intronic	in-frame	6:170561970-170561971	Insertion	/ CAG	ENST00000230354.10	c.219_221dupGCA	p.Q75dup	loss of function	1	Uncertain Significance	-

Furthermore, two VUS were identified in Ataxin 7 (*ATXN7*), which is linked to spino-cerebellar ataxia type 7 (SCA7), another expansion disorder. *ATXN7* p.Q31_Q32insRQ and c.-13_-8dupCGCCGC were observed once each and not predicted to affect gene function. Both were absent from GnomAD; *ATXN7* p.Q31_Q32insRQ is an in-frame insertion in a repetitive region without known function, while *ATXN7* c.-13_-8dupCGCCGC is located in the 5'UTR and a promoter region.

A further five VUS were identified in Calcium Voltage-Gated Channel Subunit Alpha1 A (*CACNA1A*), which encodes the transmembrane pore-forming subunit Alpha1a of the calcium voltage-gated channel, another voltage-gated ion channel similar to *SNC1A*. *CACNA1A* has been linked to spinocerebellar ataxia type 6 (SCA6) and episodic ataxia. SCA6 is an autosomal dominant disease caused by a CAG-repeat expansion in exon 47 of the *CACNA1A* gene; however, all five of the variants identified in this study are small LoF indels (three deletions, two insertions of either one or two CAG repeats) located in a non-repeat region within 10bp of each other and observed between once and 5 times. While *CACNA1A* p.Q2327dup has recently been reported as benign, p.Q2313_Q2314del and p.Q2325del have a CADD score or 21.6, placing them in the 0.1% of most likely deleterious variants based on computational evidence.

Galactosylceramidase (*GALC*) c.1162-4delT is a LOF variant found in 3 homozygous cases and is located in a promoter / intronic region; while this variant is more prevalent in cases than controls, it has recently been reported as benign[245] and *in silic*o predictions suggest no impact on the gene product. Homozygous deleterious mutations in *GALC* have been linked to Krabbe's disease and, while the gene has also been linked to adult-onset dementia, this is only the case for heterozygous mutations[246].

One coding change in 5-methyltetrahydrofolate-homocysteine methyltransferase (*MTR*) (c.*16delT) was found as two separate variants, observed 9 times at 1:236897647-236897647 and twice at 1:236897648-236897648. These variants are located in the 3'UTR and are absent from GnomAD; but no further information is

available. *MTR* converts homocysteine to methionine; homocysteine exacerbates the cytotoxicity and increases the production of beta-amyloid[247,248].

Two variants were identified in Peptidyl-TRNA Hydrolase 1 Homolog (*PTRH1*), p.H97dup and p.H96_H97dup, which were both observed once each. These in-frame deletions were both absent from GnomAD and *in silico* evidence suggests little impact on the gene product, however the protein length is affected and they are hence classified as LoF mutations. *PTRH1* is an enzyme which cleaves stalled nascent chains from ribosomes[249]. The same mutation also affects the 3'UTR of syntaxin-binding protein 1 (*STXBP1*), which is linked to a neurodevelopmental disorder[250].

Four VUS were identified in TATA-Box Binding Protein (*TBP*), mutations in which may cause FTD; CAG repeat expansions in exon 3 of *TBP* are known to cause HDL4/SCA17[251]. The four VUS were exonic in-frame CAG insertions all located in exon 3 (2x p.Q75dup, p.Q72_Q75dup, and p.Q73_Q75dup), causing the insertion of between one and three3 extra glutamines; these were observed a total of 6 times. All four VUS are predicted to cause LoF; CADD scores were only outputted for one of two Q75dup (CADD score 19.79) and p.Q73_Q75dup (CADD score 16.51), but haploinsufficiency is not a known disease mechanism for *TBP.* Deletions and insertions in this region are common in ensembl records[252] as normal numbers for CAG repeats in this exon range from 29 to 42; cases of movement disorders consistent with SCA17 may either be idiopathic or suggest that repeat numbers at the upper end of normal ranges can still cause disease[253,254].

Of the variants identified using this filtering method, only *SCN1A* p.A423D, which is located in a functional domain and leads to a substantial change in amino acid, is likely to disrupt protein function of the sodium voltage-gated channel alpha subunit 1; this has been linked to cognitive and neuropsychiatric impairment and linked to AD in mice[235-239].

ii. Variants predicted to be deleterious

Filtering for novel variants predicted to be deleterious, as described in *Materials and Methods: Exploratory variant analysis*, revealed one heterozygous mutation in STIP1 Homology And U-Box Containing Protein 1 (*STUB1*) (c.521C>T; p.T174M), which was classified as likely pathogenic based on it being unreported in controls and computational predictions of loss of function. Homozygous or compound heterozygous mutations in *STUB1* have been linked to Gordon-Holmes syndrome, which causes ataxia, sometime with associated hypogonadism[255], and haploinsufficiency has been shown to accelerate the onset of Huntington's disease in mice with a knock-in human *HTT* gene[256]. Clinically, this patient presented with prominent ataxia and a positive mGS1 family history (mother suspected of HD but diagnosed with cerebellar changes at post-mortem, and two affected sisters). Of interest, this patient (#12) also carried heterozygous mutations in *ERCC2* (associated with Xeroderma pigmentosa) and *SMPD1* (associated with Niemann-Pick disease); however all these diseases are recessive and all mutations were heterozygous. Another patient (#41) also presented with a variant affecting *STUB1* (*STUB1/ LOC105371184/ LA16c-313D11.9*; c.106dupG, p.E108fs*4); this was a null variant in a gene where loss of function is a known disease mechanism, but this patient also has an intermediate allele in *TBP*, which would be compatible with SCA17.

A further three novel variants with a CADD score over 30 - which includes them in the top 0.01% of variants predicted to be deleterious *in silico* – should be highlighted for their links to pertinent neurological disease. One variant was located in a well-established functional domain of CUB And Sushi Multiple Domains 3 (*CSMD3*) (p.R1534L), which is linked to dendrite development, schizophrenia and autism and where loss-of-function and haploinsufficiency are known disease mechanisms[257,258]. Previously reported deleterious mutations in *CSMD3* are deletions and copy-number variations causing haploinsufficiency, while the variant identified in the present sample is a missense mutation. Two variants were found in mutational hotspots in *SCN1A* (p.A423D) and in *SCN1B* (p.R52C), which are both linked to epilepsy and infantile encephalopathy[259,260] and have been described above.

Ten novel heterozygous variants predicted to be deleterious were identified in four genes known to be linked to neurodegenerative disease; they were found in Sphingomyelin phosphodiesterase 1 (*SMPD1*) (p.V36_L37insA and p.V36_L37insALA), *ATN1* (p.Q496_Q502del), *ATXN3* (p.G185fs*31, p.G185fs*32, p.G291fs*38, p.G306fs*39, and p.G185fs*40), *RP11-217B1.2/POLG* (c.624C>A; p.C208*), and *PDGFB* (p.K146fs*10). *ATN1* and *ATXN3* have been covered above. No compound heterozygous cases were identified for the recessive genes with variants predicted to be likely deleterious. While DNA polymerase subunit gamma (*POLG*) is usually considered a recessive gene, a premature stop codon is a known disease mechanism in this gene, and some heterozygous mutations in the *POLG* gene have been associated with progressive external ophthalmoplegia (autosomal dominant form 1). If caused by *POLG* mutations the disease can present with additional features such as neuropathy and depression[261]. In Platelet Derived Growth Factor Subunit B (*PDGFB*), heterozygous loss of function mutations are known to cause brain calcification in humans and in mice[262], leading to neurological problems including motor symptoms and cognitive decline with a wide range of onset from 10 to 55 years. The patient in this case was a male with an AAO of 70 and no family history; he developed behavioural problems and chorea as well as depression and disinhibition. Brain calcifications were observed on his MRI scan.

iii. Variants with a link to the HD (-like) phenotype

Linked to cell functioning in HD

Using the IVA® "Biological Context" filter as described in *Materials and Methods: Exploratory Variant analysis*, potentially disease-causing novel variants were identified in genes that have been shown to have an effect on HD and the HD-phenotype.

One variant in Dynein heavy chain 1 (*DYNC1H*) (p.Q707H) was absent from GnomAD, predicted deleterious *in silico* with a CADD score of 27.3 and expected to cause LoF and haploinsufficiency. Mutations in *DYNC1H1* follow an autosomal-dominant inheritance pattern; however, the clinical phenotype observed in the patient with

tremor, chorea and some cognitive decline is not typical of human diseases previously associated with *DYNC1H*, such as Charcot-Marie-Tooth disease. And yet, a heterozygous *DYNC1H* point mutation in mice has been shown to cause striatal atrophy and an HD-like phenotype. Huntingtin binds and regulates dynein, a role that is impaired in HD[263]; based on the RVIS score, *DYNC1H* is also in the 0.04% of genes that are least tolerant of genetic variation. Another likely deleterious variant with a link to HD was identified in a mutational hotspot in *CREBBP* (p.A1062V); this is the same variant identified through other analyses described above.

Mutations in the recessive excision repair, endonuclease non-catalytic subunit (*ERCC*) genes, which cause Xeroderma pigmentosum (*ERCC2*, *ERCC3* and *ERCC5),* have been linked to HD phenocopy syndromes[264]; however, the variant, *ERCC2* p.Q562*, is heterozygous. Nonetheless, some variants have been shown to cause disease susceptibility (for the skin manifestations) even in the heterozygous state[265].

Linked to the phenotype of HD-like disorders

Based on phenotype-driven ranking (as described in *Materials and Methods: Phenotype-driven ranking*), multiple variants in *ATN1*, Junctophilin-3 (*JPH3*), and ferritin light chain (*FTL*) were returned with the maximum phenotypic association score, which is unsurprising given that these genes are associated with dentatorubral-pallidoluysian atrophy, HDL-2, and neuroferritinopathy, which are all known HD-like syndromes. However, the variants identified in these genes are not the gene expansion pathognomonically associated with these diseases[202], and despite *ATN1* and *JPH3* being assessed to be intolerant of missense mutations (in the 9.34% and 5.54% least tolerant genes, respectively), this renders their impact doubtful.

E. Summary of highly suspicious variants

Based on the analyses set out above, which focused on harnessing the inbuilt power of IVA® to link variants to neurological disease, variants in genes that are highly expressed in the brain and intolerant to genetic variation, variants highly likely to be deleterious, and variants with either a biological link to HD or the HD-phenotype, thirteen variants of particular interest could be identified. These are listed in Table

III-16Table III-16; they are all absent from GnomAD and were found in one sample each; two separate suspicious variants were identified in *APC* and *STAT5B*.

A LoF mutation in a functional domain of *SCN1A* and close to the published deleterious variant Y426N[242] was found in a female patient who developed FTD-like symptoms aged 43 but without seizures or prominent memory problems. She had a strong family history; both her mother and brother developed similar symptoms and her mother died aged 48 in a psychiatric institution, while her sister experienced cognitive decline from her thirties onwards. The variant is absent from controls and predicted deleterious *in silico* and causes a significant change in amino-acid with changes in size and polarity. Based on its RVIS score, *SCN1A* is intolerant to genetic variation, but, despite its links to neurons, is not particularly highly expressed in the brain; rare variants in this gene were found in five cases in this dataset. It is associated with epilepsy and neuropsychiatric comorbidities in infant mice[235], cognitive impairment without seizures[236], and age-related grey-matter and brain activity changes[237]. The corresponding sodium channel Na(v)1.1 has been shown to be reduced in a BACE1-null AD mouse model[238], while overexpression of Na(v)1.1 restored cognition in a human amyloid precursor protein (hAPP)-transgenic AD mouse model[239].

Table III-16: High-impact variants found in the HDPC dataset

Gene Symbol	Case Number	Gene Region	Translation Impact	Chr: Position	Transcript Variant	Protein Variant	Inferred Activity	Classification	CADD Score	Percentile ExAC_0.05% popn	Highest Median TPM Gtex CNS	Topp Gene	Gene # Cases
AC010127.1; SCN1A	21	Exonic; ncRNA; Intronic; 5'UTR	missense	2:166046879-166046879	n.1673C>A	p.A423D	loss of function	Uncertain Significance	28.9	2.4%	15.14	45	5
CTSB	44	Splice Site	SNV	8:11845790	c.422-1G>A		normal	Pathogenic	25.2	66.9%	141.5	1	2
CTSB	24	Exonic	SNV	8:11845789	c.422G>A	p.G141E	normal	Pathogenic	25.2	66.9%	141.5	1	2
APC	3	Exonic	missense	5:112842414	c.6745G>A	p. A1991T		Uncertain Significance	< 10	0.42%	360.93	6	2
APC	44	5'UTR; Exonic	missense	5:112707873	c.-28C>A	p.S52R		Uncertain Significance	22.4	0.42%	360.93	6	2
CREBBP	40	Exonic	missense	16:3758924	c.3185C>T	p.A1062V		Uncertain Significance	25	0.60%	47.51	2	4
MAPK8IP3	22	Splice Site		16:1767919	c.3502+1G>C	*		Uncertain Significance	< 10	1.06%	402.7	61	7
MAPK8IP3	6, 10, 17, 24	Exonic	Stop gain	16:1762879	c.1750A>T	p.K584*	LoF / normal	Uncertain Significance	41.0	1.06%	402.7	61	7
STAT5B	43	Exonic	missense	17:42207558	c.2077G>A	p.A693T	loss of function	Uncertain Significance	23.5	22.09%	106.5	25	2
STAT5B	3	Exonic	missense	17:42216087	c.1400T>C	p.V467A	loss of function	Uncertain Significance	26.1	22.09%	106.5	25	2

Gene Symbol	Case Number	Gene Region	Translation Impact	Chr: Position	Transcript Variant	Protein Variant	Inferred Activity	Classification	CADD Score	Percentile ExAC_0.05% popn	Highest Median TPM Gtex CNS	Topp Gene	Gene # Cases
AC124068.2; POLG	11	Promoter; Exonic	stop gain	15:89333131	c.624C>A; n.-1922G>T	p.C208*	loss of function	Pathogenic	42	53.92%	33.76	-	1
PDGFB	11	Exonic	frameshift	22:39230205	c.435delG	p.K146fs*10	loss of function	Likely Pathogenic	-	59.95%	12.3	-	1
DYNC1H1	2	Exonic	missense	14:101986346	c.2121G>T	p.Q707H	loss of function	Uncertain Significance	27.3	0.0098%	140	-	3

A splice site mutation was identified in *CTSB* (c.422-1G>A) in in a female patient with severe generalized dystonia with prominent chorea, impaired saccade initiation and oculogyric crises movements as well as some pyramidal features beginning at age 22 years; she had a GS of 1 with both her mother and grandmother affected by dementia. While investigating this variant, a further suspicious variant in this gene was identified (p.G141E): an exonic missense mutation with a CADD score of 31.0 in a female patient who developed restless legs, twitching and depression aged 34 years with a strong family history of huntingtonism. *CTSB* was shortlisted as the most relevant gene by the ToppGene algorithm and plays a part the lysosomal storage pathway and downregulation has been shown to be a disease modifier in PD[215,216]. The variants are adjacent and both affect the same, splice-site spanning exon; there is a single stop gain variant in the same codon found on *GnomAD*, but a single count of a variant is generally still compatible with even a very rare disease[28].

Two missense variants were observed once each in *APC* (p.S52R and p.A1991T); both variants were located far from each other and only p.A1991T was predicted deleterious *in silico*. *APC* (p.S52R) was found in the same patient as *CTSB* (c.422-1G>A). *APC* (p.A1991T) was identified in a male patient who developed progressive behavioural decline beginning at 59 years of age. He had frontal deficits, a labile affect, echolalia, memory problems, and visual hallucinations, as well as left-sided extrapyramidal and upper-motor neuron motor signs; he had a GS of 1. *APC* is predicted to be in the 0.5% of genes least tolerant of genetic variation and is highly expressed in brain; ToppGene suggested it as one of top 10 most relevant genes for this analysis. *APC* encodes a tumour suppressor most often associated with familial colon cancers, but which is also an antagonist of the WNT signalling pathway. WNT signalling has been shown to be downregulated in Alzheimer's disease, but upregulated in other neurodegenerative disorders such as HD and Parkinson's disease (PD)[203,204].

Furthermore, a missense variant in a mutational hotspot was found in *CREBBP* (p.A1062V) in a female patient who developed chorea, ataxia and frontal deficits

starting at 52 years of age; she had no family history of neurological disease. *CREBBP* is among the 1.54% genes least tolerant of genetic variation and the second highest ranking gene in the ToppGene list. The *CREBBP* protein is a transcriptional coactivator and has been shown to interact with the HD protein[205] and to facilitate its localization to neuronal intranuclear inclusions in a mouse model of HD; in this model of HD the HD protein also affects the transcription of p53-regulated promoters. *CREBBP* haploinsufficiency has been shown to be deleterious and reduced *CREBBP* activity is associated with long-term memory deficits in HD[206]; the mutant HD protein seems to induce cell death in cell culture through transcriptional disruption of cell survival programmes via *CREBBP*[207], potentially via TAFII130—a TATA-binding protein-associated factor[266]. Since *CREBBP* contains a polyglutamine stretch, it has been considered as a candidate gene alongside *POU3F2*; however since no disease-causing trinucleotide expansions have been identified these genes, they have since fallen somewhat out of favour[267]. In this dataset, suspect mutations were absent from *POU3F2*, but this *CREBBP* variant meets the criteria to be likely deleterious; it is located in in exon 17, in which a splice site mutation in exon 17 was associated with Rubinstein-Taybi syndrome[268], and is part of the disordered region ID3, which has been associated with neuromuscular and neurodegenerative disorders[269]

A frameshift mutation (c.3502+1G>C *) disrupting a splice site was identified in *MAPK8IP3* in a female patient with a background of learning difficulties and epilepsy who developed behavioural changes and chorea age 69 years; she had no relevant family history. A missense variant causing a stop gain (p.K584*) was found in four patients (HDPC Sample numbers 6, 10, 17, and 24). These included the same patient as the *CTSB* (p.G141E) variant, as well as one female patient with chorea, cognitive decline and depression, whose symptoms began at age 45 years old; two of her sisters were also affected, but not her two healthy daughters. Furthermore, it was identified in a male patient who developed FTD-like symptoms, memory problems, irritability, depression and anxiety, as well as dystonia and myoclonus at 53 years of age without any family history. Finally another male patient was found to have this variant: he developed a Parkinsonian syndrome with slowing cognitive faculties, deteriorating

memory, depression and extracampine hallucinations, but normal DaTScan; an uncle also had Parkinson's disease.

MAPK8IP3 is highly intolerant to genetic variation (among the 1.1% least tolerant genes) and highly expressed in brain, although it was only ranked number 61 by the ToppGene algorithm (out of 354). *MAPK8IP3* is involved in axonal and synaptic vesicle transport and may regulate kinases of the JNK signalling pathway, thereby acting as a scaffolding protein in neurons; a heterozygous stop gain mutation (p.Tyr37Ter), albeit earlier in the gene than in this case, has been identified in a patient with infant onset developmental delay[208]. Other mutations, some in the RH1 domain, which is similar to RH2 domain where p.K584* is located, and some the same WD40 domain as c.3502+1G>C *, have also been linked to childhood intellectual disability, ataxia and unstable gait[270]. However, mutations in this gene were quite frequent in the HDPC cohort.

The *APP* and *CREBBP* variants were identified through both the analysis focused on the most likely deleterious variants and the gene-expression and gene-tolerance based analysis, while for *MAPK8IP3* one variant was identified using each approach.

Two missense variants were identified in *STAT5B* and were located functional domains (p.A693T - (SH2, the highly conserved tyrosine phosphorylation site) and p.V467A (STAT-DBD, the DNA-binding domain). p.A693T was found in a female patient without relevant family history, who developed cerebellar ataxia, dysarthria and cognitive impairment from 58 years old. p.V467A was observed in the same male patient with an autosomal dominant family history who was also found to carry the *APC* (p.A1991T) variant and developed frontal deficits, memory and motor problems. *STAT5B* is in the 25% of genes most intolerant to genetic variation and highly expressed in brain; it was the 25th highest ranking gene in the ToppGene recommendations for relevant genes. *STAT5B* encodes a transcription factor and its disruption could interfere with microglial function[211]; tyrosine phosphorylation follows ligand-induced activation and is required for dimerization and nuclear translocation[271,272].

A stop gain variant (p.C208*) was found in *RP11-217B1.2/POLG* in a male patient without significant family history, who developed a behavioural syndrome and fidgety movements aged 70. *POLG* encodes the catalytic subunit of the mitochondrial DNA polymerase gamma; both dominant and recessive mutations are known to cause a wide spectrum of disease including ataxia, progressive external ophthalmoplegia, neuropsychiatric symptoms and parkinsonian features[261,273]. Chorea and dystonia, however, are not commonly observed[273]. A heterozygous LoF *PDGFB* (p.K146fs*10) was identified in the same patient who also had brain calcifications on his MRI. A LoF variant in *PDGFB* causing haploinsufficiency would be consistent with brain calcifications[262] leading to motor symptoms and cognitive decline. Both genes are relatively tolerant to genetic variation and not particularly highly expressed in the brain; neither gene was shortlisted by the ToppGene algorithm.

Finally, a missense variant in *DYNC1H* (p.Q707H) was identified in a female patient with an autosomal dominant family history who was suffering from tremor from age 55 and then developed orolingual, cranial and distal extremity chorea, as well as frontal and subcortical dementia aged 70, she also had had psychotic episode. Heterozygous mutations in *DYNC1H1* can cause Charcot-Marie-Tooth disease; while the patient's phenotype was not consistent with this diagnosis, *DYNC1H* mutations in mice have been shown to cause striatal atrophy and an HD-like phenotype. *DYNC1H* encodes dynein cytoplasmic 1 heavy chain 1 which is a microtubule-linked ATPase involved in axonal transport[274]; huntingtin binds and regulates dynein, a role that is impaired in HD[263]. *DYNC1H* is in the 0.04% of genes that are least tolerant of genetic variation and is highly expressed brain; however, it was not part of the ToppGene ranking.

F. Replication in the Genomics England 100,000 dataset

None of the highly suspicious variants described above could be replicated in the Genomics England 100,000 dataset (search performed 23rd April 2020).

Following the penetrance calculation method described above for EOAD and EOFTD[24,28], and assuming a prevalence of HDPC syndromes of approx. 2.5/100,000 (20% of 12.4/100,000 for HD[94], see above), I was able to estimate the penetrance of a variant found once in the HDPC discovery dataset, once in the approx. 14,000 neurological and neurodevelopmental cases of Genomics England's 100,000 Genomes project, as well as once on *GnomAD*. Based on these numbers, a variant found once in the HDPC dataset and once in the 100,000 Genomes, as well as once on *GnomAD* can be expected to have a penetrance of 0.0503% (95% CI: 0.0024 – 1.04%); the replication analysis of the Genomics England dataset was therefore restricted to novel variants.

Variants in *APC* were not investigated in this step because of confounding effects with cancer genetics. Gene domains of variants identified to be potentially disease causing in the HDPC discovery dataset were identified (see Table III-17) and examined for promising candidate variants in patients in the Neurology and neurodevelopmental disorders subcohort of the Genomics England dataset.

Potential heterozygous candidate variants were identified in three patients with neurodegenerative disease and three patients with motor disorders of the CNS (see Table III-18). One variant was found in *CTSB*, in a female patient with childhood ataxia but it was predicted benign *in silico*. Two variants were identified in *MAPK8IP3* in patients with early onset dementia and a Parkinsonian syndrome, respectively, beginning in their 60s; both variants were predicted probably damaging *in silico*. A *CREBBP* variant was predicted benign *in silico*, while a probably damaging *STAT5B* variant was found in a male patient with motor neurone disease beginning in his 50s. *MAPK8IP3* mutations have been linked to intellectual disability and gait disturbance in children[270]; in this dataset, one patient was identified to carry a mutation in the same WD40 domain as the two patients in the Genomics England dataset; she had a background of learning difficulties and epilepsy, and developed superimposed behavioural changes and choreoathetoid movements, restlessness, agitation and decreased interactions starting age 69 years old. She had no relevant family history. Similarly, the patients identified in the Genomics England dataset developed a

Parkinsonian syndrome and early-onset dementia beginning in their 60s. By contrast, no other novel mutation in the same RH2 domain as the stop gain mutation *MAPK8IP3* p.K584* were observed in the Genomics England dataset. The *STAT5B* mutation p.V467A found in this HDPC dataset in a patient with a behavioural FTD phenotype and additional upper motor signs beginning at age 59 years old, is located in in the DNA-binding domain; a similar mutation in the same domain was identified in a patient included in the Genomics England dataset with motor neurone disease, which would fall on the FTD/ALS spectrum.

Table III-17: Gene regions investigated in the Genomics England dataset

Gene Symbol	Transcript Variant	Protein Variant	Assumed mechanism	Gene region searched
CTSB	c.422-1G>A		Located in splice site spanning intron 8/9 and exon 9, this mutation probably disrupts the splice site[159].	All coding non-synonymous variants in CTSB, of special interest are variants between 8: 11,847,051 and 8:11,845,661 , which spans intron 8/9 and exon 9
CTSB	c.422G>A	p.G141E	Located in first base pair of exon 9.part of a codon split across a splice site. Normally, the codon reads GGC and results in a glycine (G). This change results in a glutamic acid (E) [159].	All coding non-synonymous variants in CTSB, of special interest between 8: 11,847,051 and 8:11,845,661 , which spans intron 8/9 and exon 9
CREBBP	c.3185C>T	p.A1062V	Altered protein function. Exon 17 - a splice site mutation in exon 17 was associated with Rubinstein-Taybi syndrome. Located in the disordered region ID3 between amino acids 674 and 1080, which may be implicated in protein interaction a zinc finger protein ZFP106, involved in RNA processing and it has been associated with neuromuscular and neurodegenerative disorders[268,269]. The variant is located in a low complexity domain which overlaps with the centre of a zinc finger domain according to ensemble[159].	Missense and LOF mutations between amino acids 674 and 1080, i.e.16:3,778,182 and 16:3,767,720
MAPK8IP3	c.3502+1G>C *		Located in first position of intron 28/29. 2/13 patients with childhood intellectual disability also had ataxia, one additional patient had unstable gait. The two patients with ataxia had mutations in the RH1 domain at the very beginning of the protein, but the patient with unstable gait had a mutation in the same domain as our WD40, but earlier; two others with the same mutation had spastic paraparesis[270]	Located just past the WD40 domain which runs from amino acids 971 to 1173, the codon for amino acids 1174 spans exons 28 and 29 and would be disrupted by changed splicing. Check for coding non synonymous and splice site mutations between 16: 1,766,529 and 16:1,768,080 , which includes the following intron and part of exon 30
MAPK8IP3	c.1750A>T	p.K584*	Located in exon 16, in the RH2 domain between amino acids 520 and 594, where missense mutations have been linked to intellectual disability. 2/13 patients with childhood intellectual disability also had ataxia, one additional patient had unstable gait. The two patients with ataxia had mutations in the RH1 domain at the very beginning of the protein, but the patient with unstable gait had a mutation in the same domain as our WD40, but earlier; two others with the same mutation had spastic paraparesis[270].	Coding non-synonymous variants and splice site variants located between 16:1,762,351 and 16:1,763,006

Gene Symbol	Transcript Variant	Protein Variant	Assumed mechanism	Gene region searched
STAT5B	c.2077G>A	p.A693T	Located in last base pair of exon 16 in "PY" domain between amino acids 686 and 699. All STAT molecules have a highly conserved tyrosine phosphorylation site (Y) at or around residue 700, labelled as the phosphotyrosyl segment (PY). Tyrosine phosphorylation follows ligand-induced activation and is required for dimerization and nuclear translocation[271,272].	Coding non-synonymous and splice site variants between 17:42,207,728 and 17:42,202,757
STAT5B	c.1400T>C	p.V467A	Located in exon 12 in the DNA binding domain between amino acids 332 and 471[271].	Missense and LoF in this DNA binding domain between amino acids 332 and 471, i.e. 17:42,218,330 and 17:42,216,014

Table III-18: Variants of interest identified in the Genomics England dataset

Gene	Chr:Position	Ref/Alt	Consequence	Exon (Intron)	cDNA pos.	Protein pos.	Amino acids	PolyPhen	Sex	AAO range	Normalised disease
CTSB	8:11845764	C/G	missense	9/10	931	273	E/D	benign	Female	0-9	Hereditary ataxia
MAPK8IP3	16:1767208	A/G	missense	26/32	3302	1049	I/V	probably damaging	Female	60-69	Early onset and familial Parkinson's Disease
MAPK8IP3	16:1767621	G/A	missense	27/32	3449	1098	G/S	probably damaging	Male	60-69	Early onset dementia
CREBBP	16:3770776	C/G	missense	14/31	3471	892	V/L	benign	Female	0-9	Early onset dystonia
STAT5B	17:42218311	G/T	missense	9/19	1163	337	Q/K	probably damaging	Male	50-59	Amyotrophic lateral sclerosis or motor neuron disease

IV. Discussion

Over the last 28 years, dementia syndromes have progressively been recognized as heterogeneous and multifactorial disorders as the number of known Mendelian causes of disease has increased. Dementia genes have been found to be pleiotropic, linking what were previously entirely separate entities[111,194], and improving our understanding of underlying pathophysiological developments in the process. However, the translation of this knowledge into routine clinical practice has been limited[275], and until recently, only a small number of tests were clinically available. These were often restricted to the genes associated with familial forms of early-onset dementia with Mendelian inheritance, such as *APP*, *PSEN1* and *PSEN2* for AD, and *GRN*, *MAPT*, *FUS* and *VCP* for FTD; studying these genes has permitted crucial insights into the underlying clinical-pathological processes[18,20,33]. However, many familial cases continue without a genetic diagnosis, some variants may be linked to more than one syndrome, like the *C9orf72* expansion[13,29,30,146], and not all of them may be fully-penetrant Mendelian variants[28].

Genome-wide association studies (GWAS) have shown both neurological and psychiatric disorders to be heritable to a substantial degree[276]. While psychiatric diseases share considerable genetic risk, neurological diseases appear to be more genetically distinct from one another[277] but crossover between psychiatric genetic (GWAS) risk and related HD symptoms has been demonstrated[278] in HD. Whether correlations between common traits and disease-specific symptoms are picked up in studies depends on the strength of the underlying association, but also the available data[279]. For the time being, the complex information gleaned from concurrent mutations, genetic modifiers and GWAS polygenic risk scores has no practical utility for healthy people and has therefore not yet entered clinical decision making[32].

1 Insights from the MRC Dementia Gene Panel

A. Predictors of genetic disease in dementia syndromes

In the first chapter, the MRC Dementia Gene Panel aimed to support clinical decision-making by evaluating the frequency of genetically caused disease in patients with AD, FTD, prion disease and dementia with motor symptoms (DemMot). It sought to quantify the effects of clinical information routinely used to select patients for genetic testing, such as sex, age at onset of the disease and family history (stratified as the modified Goldman score (GS)[74,78]), on the likelihood of discovery of a genetic mutation in the context of next-generation sequencing. The results of this series of patients - enriched for those likely to be carrying deleterious mutations, and large enough to inform clinical practice – was published in 2018[24] and supports the more widespread use of gene-panel diagnostics in dementia. Prospectively obtained samples from primary referral centres were included in the analysis to establish transferability of the results.

The number of clinically relevant variants in this series was high in all groups aside from the elderly with a negative family history and those with dementia and motor symptoms that may not be caused by variants in typical dementia genes. Clinical syndrome, age and the strength of the family history were identified as predictive factors that should help guide counselling and decisions about referral for testing. Clinical syndrome was a strong predictor of the chance of detecting a mutation and the gene, but in markedly different ways. Ninety-four percent of suspected prion disease cases with DVs were linked to a single gene, *PRNP*; 93.5% of FTD patients with DVs were linked to three major, and two additional genes associated with FTD syndromes (*C9orf72*, *GRN*, *MAPT*, *SQSTM1*, *VCP*); however, in only 63% of clinically diagnosed AD patients were the DVs in genes linked to AD pathologies (*APP*, *PSEN1* or *PSEN2*). DVs in patients with a dementia-motor syndrome were uncommon and heterogeneous in their associations. These findings have implications for clinical practice: it would be reasonable to refer suspected prion disease patients for testing of *PRNP* alone. For FTD and AD syndromes, the dementia gene-panel approach (+*C9ORF72*) seems sensible owing to the diversity of genes involved and phenotypic heterogeneity. Dementia-

motor syndromes are more challenging however; a low rate of DV discovery either implies that disease relevant variants are not covered by our panel - this would not be surprising as we did not screen genes associated with familial Parkinson's disease or the expansion disorders linked to HD phenocopy syndromes other than *C9orf72* - alternatively, these patients may harbour a low rate of single gene disorders. Despite the prominent role dementia plays in these patients' clinical syndromes, a panel covering typical dementia genes only is of limited use in this cohort; more research is needed to resolve this question.

Age at onset was also a strong predictor of finding a DV. However, this was not an absolute rule, with the rate of DV detection being 13.5% in those with AAO < 65 and 7.2% in those with AAO > 65, which was a surprising finding perhaps related to the selection bias inherent in our referral based sample. Family history remained an important predictor in all age groups. Gene panels are therefore suitable for late-onset dementia where there is evidence of a family history, while only 6 DVs in 233 patients (2.6%) were found in late-onset dementia with a negative family history (GS4): three in *PRNP*, two *C9orf72* expansions and one in *GRN*. If the family history is negative in late-onset dementia, it seems reasonable not to consider gene-panel testing - as would be normal practice at the moment. Dementia gene-panel diagnostics should be considered in all early-onset patients, and late-onset patients with evidence of a genetic disorder in the family history (GS1, 2 or 3)[72]. However, opinions will vary among physicians and patients about what level of risk justifies gene-panel testing; indeed, some clinicians/patients/families may feel that even low risks < 5% of discovering a DV would justify testing.

Even if no proven disease-modifying treatments are yet available, identifying patients who carry pathogenic variants has many potential benefits. It provides patients with certainty and a firm diagnosis, removing the need for further potentially invasive diagnostic tests and interventions. A genetic diagnosis opens up access to support groups, who often lend strong lobbying and practical support to their case, as well as permitting access to precision medicine and clinical trials, now ongoing for most monogenetic dementia disorders, for example, the HD antisense gene silencing

trials[280]. Genetic status has also been found to be associated with shorter survival times in FTD[281,282], while different *PRNP* variants may have starkly differing prognoses[28], so that the exact identified variant may also have implications for genetic counselling. Last but not least, a positive genetic test also provides the opportunity for siblings or descendants to learn about their own risk of a disease they already know runs in the family, and to plan accordingly including making decisions about accessing preimplantation genetic diagnosis. Being able to estimate the likelihood of identifying a deleterious mutation should help the counselling process and decision-making on clinical gene testing.

In order to best support this process, samples selected for the MRC Dementia Gene Panel aimed to be representative of cases being considered for genetic testing by referring physicians. Selection bias will most likely have influenced the prevalence of DVs found in this study compared with a population-based study or unselected dementia patients; however neither sex nor whether a sample originated in a research centre, the National Prion clinic, an NHS clinical genetic service or in a primary referral cognitive clinic were associated with the likelihood of finding a mutation, strengthening the transferability of the present results to real-life clinical practice.

The MRC Dementia Gene Panel was designed and validated in 2014[113] and could not be modified during the course of the study; the discovery of *TBK1* as a relevant dementia gene causing FTD could therefore not be incorporated. A subset of patients sequenced on the MRC Dementia Gene Panel was sent for exome sequencing, which was not part of this thesis, and a *TBK1* mutation was identified in a single patient. It is therefore unlikely that including *TBK1* on the panel would have meaningfully altered the results, although recent data suggests that it may also increase susceptibility to AD[283] and should be included in future investigations. Since exome sequencing leads to a very small increase in mutation detection, gene panel testing appears most appropriate for typical dementia syndromes as it is more cost-effective and avoids incidental findings.

B. Genotype vs. Phenotype and the pleiotropy of dementia syndromes

A high rate of novel variants and known variants in genes unlikely to be selected for single gene tests based on the clinical syndrome mean that the results of the MRC Dementia Gene Panel study justify broader clinical testing than hitherto customary.

Deleterious and likely deleterious variants (DVs) were observed in excess in cases vs. controls (12.2% vs. 0.4%); in addition, variants classified as potentially deleterious were also significantly more common in cases than in controls (4.1% vs. 1.3%). However testing DemMot patients for variants in dementia genes does not offer the same yield as for typical dementia syndromes such as AD and FTD. While the established DVs were least likely to be identified in DemMot patients, the frequently novel, potentially deleterious variants were most commonly observed in this same, more heterogeneous patient group. Less well characterized variants in dementia genes may therefore confer atypical features; the phenotypic spectrum of genes already known to be associated with neurodegeneration may be broader than has been understood so far and more hitherto unrecognized deleterious variants in these genes are yet to be discovered. The FTD-ALS spectrum, especially where it is linked to *C9orf72* expansions[284], is the archetypical example of how one gene can cause multiple phenotypes. *C9orf72* expansions are also the commonest cause of Huntington's disease phenocopy syndromes[13], a form of often unexplained dementia with motor symptoms.

Overall, 16.9% of cases with an at least likely deleterious variant found through the MRC Dementia Gene Panel could be linked to a gene not typically associated with their clinical syndrome. These were most likely to be identified in cases with a limited family history (GS of 3.5, 4 and 4.5). Put differently, not only are over 16% of all patients with a published deleterious variant in a gene known to be associated with neurodegeneration unlikely to be tested for the correct gene if that gene has to be selected from a list, these cases are also those who are least likely to be tested to start with and the least likely to be followed up. In addition, in the present dataset, 11.3% of all at least likely deleterious variants were novel; these less studied variants may

confer atypical clinical features and explain some of the overlap between genotypes and phenotypes, as demonstrated recently by a study linking various genes associated with neurodegenerative diseases to FTD[285].

These numbers are particularly striking because they emphasize the need for broad clinical testing and with the advent of next-generation sequencing, comprehensive genetic assays on a clinical basis have become feasible[113]. Naturally, a possible bias in referrals due to selection bias may have implications for the amount of pleiotropy found in this study. Although no significant effect of sample origin could be detected in this dataset, our study should not be used to infer the extent of pleiotropy in the broader population. However, these data also suggest that even genes already associated with neurodegeneration are still not fully explored. Classifying novel variants presents unique challenges in clinical genetic testing[132], and more large-scale testing efforts will be required to close this gap.

Synonymous variants in all tested genes were overall not significantly more frequent in cases than controls, even though the panel used in this study covered mostly intronic sequence adjacent to exons and therefore likely to be splicing-relevant. Differential splicing in genes linked to neurodegeneration, such as *MAPT*, *FUS* and others, has been shown to cause dementia[36,286,287]. However, they may be linked to only some of the genes tested in this study or to very specific locations and more differentiated testing may be required to properly assess them in this dataset. One of the weaknesses of typical next-generation sequencing (NGS) is that because of its inherent technical limitations it is not usually suitable to the evaluation of repeat-expansions[288,289]. An innovative software package[187] was used to analyse the average coverage and suggested the presence of CNVs in two AD cases with limited family history, one in *GRN* and one in *MAPT*, respectively. However, these CNVs could not be ascertained with Sanger sequencing and must be classed as artefacts. While highlighting the limits of the technology, the need to exploit the data by all available means nonetheless persists.

C. Single vs. concurrent mutations

Concurrent DVs were found to be significantly enriched in the present dataset with 3.2% of patients with at least one DV affected. This has implications for counselling and testing. Epistasis has previously been shown to be a factor in AD[290,291] and FTD[85,104,105]. In this dataset, there was notable excess in FTD cases (7 out of 11 patients with concurrent DVs), and should be particularly relevant in counselling of these patients. With AD patients often used as a fit-all diagnosis for different dementia syndromes, and FTD patients carrying such a high rate of concurrent mutations, single gene testing risks missing many DVs in known genes. In addition, synergistic effects between novel and less-established variants may also lead to disease, as demonstrated by recent findings in FTD patients[285].

D. ApoE, risk factors and variants with reduced penetrance

While this study focused on Mendelian causes of dementia, selected risk factor were assessed and analysed. As was to be expected, previously published risk factor variants were identified in both cases and controls; overall they were significantly more common in cases than controls.

In the AD cohort, the ApoE genotypes 3/4 and 4/4 were significantly enriched, while ApoE 4/4 was also more common in the FTD cohort. ApoE is the best known risk locus for sporadic AD and ApoE4 contributes most of the risk[65,71]; however, possession of an ApoE4 allele is neither necessary nor sufficient to cause AD, and so clinical genetic testing is not recommended[292].

Progranulin loss-of-function mutations and haploinsufficiency are a well-established disease mechanism in frontotemporal dementia[293,294] and reductions in progranulin levels have been linked to dementia symptoms[295]. This project therefore stipulated that *GRN* missense mutations could be a risk factor for dementia without causing a complete loss of function. Indeed, *GRN* missense mutations were observed in excess in patients with AD, FTD and DemMot, implying that these variants may have a negative effect. Assessing the pathogenicity of individual variants is currently hampered by the rarity of most of the variants involved. In much the same way, heterozygous *TREM2*

variants have been reported in excess in AD[51]; however, in this dataset heterozygous *TREM2* mutations as a whole were not significantly more common in dementia syndromes. Although heterozygous *TREM2* variants appeared to be more frequent in AD patients, this analysis did not reach significance; more powerful analyses may be necessary to distinguish between variants of possibly opposite effects.

More work is also needed to improve information in the literature and databases about the pathogenicity and penetrance of variants. Following the methods set out by Minikel *et al.*[36], all variants were assessed for their frequency in online population databases such as GnomAD[123,124] to assess their likely pathogenicity. Penetrance calculations were performed tor variants in genes associated with AD (*APP*, *PSEN1*, and *PSEN2*) and FTD (*GRN*, *MAPT*, and *VCP*). Based on this assessment, not all purportedly pathogenic variants appear to be compatible with being fully-penetrant Mendelian pathogenic mutations since an excess of potential DVs is seen in population data, incompatible with the observed prevalences of early-onset dementias. Given the prevalence in the population of approx. 30/100,000 and 18/100,000, respectively, of EOAD and EOFTD, and the proportion of genetic cases (see Results Chapter 1 for details), a variant that exceeds a frequency of 1-2 alleles / 125,000 (the approximate size of *GnomAD*[123] at the time of the analysis) in the general population is implausible to be a fully penetrant cause of a rare disease like EOAD or EOFTD. The relative variant frequency in cases and the general population can provide essential information for the interpretation and classification of the variants. Both for EOAD and EOFTD, variants have been reported as potentially deleterious, which are most likely either benign or low penetrance. At least some of these reports of supposedly deleterious variants are likely to have been the basis for genetic counselling. Large-scale studies harnessing the power of next-generation sequencing and big data are vital tools to ensure clinical diagnosis, testing and feed-back is as accurate as possible going forward. Improved sharing of patient genetic data, the availability of large-scale population data, improved *in silico* and *in vitro* modelling, particularly for less commonly involved genes and dementia syndromes should help improve the accuracy of classification[124]. This should be supported by the development of guidelines and provision of funding to

support sharing of clinical and genetic data in databases to further improve the accuracy of classification.

2 ApoE4 – more than just an AD risk factor

Beyond being a risk factor for AD, and potentially FTD, ApoE4 appears to have a disease-modifying effect on FTD-tau[296]. In the UCL FTD cohort, ApoE4 significantly lowered the age at onset in FTD patients with *MAPT* mutations or proven FTLD-tau pathology independently of Aβ by 3.9 years[296]. This is consistent with the results of Shi *et al.* who previously proposed that ApoE4 genotype acts through a toxic gain of function mechanism to exacerbate or modify tau pathology, neuroinflammation, autophagy, and reactive astrocyte activation[147]. They provided extensive evidence in a mouse model of FTD and in neuropathological cases of tauopathies, as well as clinical support in an AD cohort[147]. These results further strengthen their hypothesis providing evidence in human that ApoE4 modifies the clinical phenotype of FTLD-tau, independently of Aβ co-pathology. Indeed, in this dataset, ApoE4 and Aβ co-pathology were associated with opposing effects on age at onset in tau+ FTD cases[296]. The confounding impact of Aβ co-pathology, along with the heterogeneous nature of FTD pathology and the specificity of the effect of ApoE4 on tau pathology may explain why previous studies of ApoE4 in FTD[192-195] did not detect this influence. Despite the robust statistical findings and the large cohort of FTD patients, analysis of pathological and gene mutation defined subgroups inevitably leads to relatively restricted numbers and therefore further replication will be important. Apolipoprotein is a good example of how common variants can have significant and widespread disease-modifying effects, and are of increasing research interest. Having being discovered over 25 years ago[148,149,297], new implications and consequences are still being discovered. Nonetheless, its clinical predictive value remains limited and clinical testing is not recommended.

3 HD and HDPC patients share a clinical phenotype

Discovery rates for reportable causal variants were low for HDPC patients on the MRC Dementia Gene Panel despite the prominent role dementia plays in the DemMot syndrome and the large overlap between the two cohorts. While some patients'

symptoms may be caused by novel variants in established dementia genes, others' may be linked to other non-dementia genes (since genes more typically associated with motor syndromes such as Parkinson's disease were not included), or by other genes altogether. Clinical syndrome strongly predicted the yield of genetic panel testing; however, the definition of HDPC syndromes as *patients in whom an experienced clinician deems HD likely enough to request a gene* test[13,103], while practical and pragmatic, has hitherto been subject to variation between clinicians based on judgement and patient factors. In an attempt to improve the HDPC definition, HD experts were surveyed on what symptoms they would expect in HD and which would make them suspect an HDPC syndrome; these results were juxtaposed with a fastidious comparison of HD and HDPC patients presenting to the same two Neurogenetics clinics.

Based on the survey results, unsurprisingly, chorea was considered to be typical of HD, as were dystonia, gait abnormality and falls, a dysexecutive syndrome and cognitive slowing, irritability, apathy and depression. It is worthwhile remembering, that while the typical manifestation of HD is hyperkinetic and HD is the cause of 90% of genetic chorea[298], it is a heterogeneous disease potentially presenting with relatively pure dystonic, atactic, and psychiatric symptoms; as the disease progresses gait and postural disturbance, as well as falls eventually follow[202]. In addition, the 6% to 10% of patients with juvenile onset usually present with the hypokinetic *Westphal* variant of the HD, sometimes in combination with seizures, further broadening the disease spectrum[299]. Neuropsychiatric symptoms, such as apathy, irritability and executive dysfunction are very common in HD patients, even before motor onset and increase in prevalence as the disease develops[300]. Apathy was even found to predict the rate of cognitive decline in presymptomatic HD patients[301]; after motor onset, dysexecutive functioning may impair patients awareness of their own neuropsychiatric symptoms, such as apathy and disinhibition[302].

By contrast, HD clinician experts considered neuropathy, limb weakness, pain, and tremor to be more indicative of a non-HD syndrome. Even when the need to test for

HD based on a different concurrent symptom was imperative, certain symptoms, such as neuropathy, limb weakness, pain, tremor, and ataxia, would make them expect a negative test result. Neuropathy and ataxia, for example, may be associated with the spinocerebellar ataxias, Friedreich's ataxia or mutations in the *POLG* or *TWNK* genes[13,103,303,304], while limb weakness may be indicative of vascular disease[305,306]. Meanwhile tremor is most often diagnosed as essential tremor; on closer examination, however, these patients with essential tremor often have undiagnosed conditions, such as dystonia and Parkinson's disease, and may also require a more detailed work-up[307]. At the same time, it is worthwhile remembering that not all symptoms found in a patient will necessarily be linked to the primary syndrome; neuropathies, for example, are very common in adults and specifically the elderly and can be triggered by various causes[308].

Meanwhile, comparing HD and HDPC patients presenting to two Neurogenetics clinics at NHNN and GOSH, revealed a different picture, perhaps because referring neurologists had already filtered some of the less convincing HDPC patients or perhaps because some symptoms were overrepresented in the survey due to historical expectations and definitions. HD and HDPC patients presented with comparable levels of chorea, irritability, dysexecutive syndrome, apathy, gait abnormality and falls, cognitive slowing, and depression, which were the symptoms deemed most typical of HD by clinician experts. The exception was dystonia, which, while present in 10% of HD patients was found in 33.3% of HDPC syndromes, although this finding was significant only if no correction for multiple testing was applied. Dystonia made HD clinician experts consider testing for HD, and was not considered a likely indicator of a different syndrome. It is a typical feature of HD[309], and can sometimes be the only the only symptom for some time[310], but pronounced dystonia is also common in HDL2, HDL3 and SCA17/HDL4, as well as in HDPC syndromes associated with *C9orf72*, neuroferritinopathy and brain calcifications[202]. Dystonia has been shown to detrimentally affect day-to-day functioning in HD patients more than chorea[311]; seeking a genetic diagnosis may therefore be particularly useful in these patients.

In this study, insomnia, dysarthria and dysphagia were also significantly more common in HD than in HDPC patients. Insomnia is a well-known symptom of HD, and can have secondary effects on cognitive and emotional symptoms[312]. Circadian disruption is a feature of many neurodegenerative diseases and may be a factor in disease progression[313]. Changes in speech, mainly affecting speech agility, phonatory function and speech rate have been demonstrated to be already present in premanifest HD patients[314], before the onset of motor symptoms adds to levels of dysarthria, but little longitudinal data is available to map speech function to disease progression. Dysphagia on the other hand can be used to track disease progression in HD[315]. Both dysarthria and dysphagia are obviously not unique to HD, but although early corticobulbar involvement is particularly relevant in a catabolic disease it remains poorly understood and treatment options require further study[316]. In contrast, and as predicted, neuropathy, limb weakness, and pain were indeed exclusively observed in HDPC patients, while tremor was seen in HD patients but was significantly more common in patients with HDPC syndromes. Polyneuropathy can be a mark of SCA2 or SCA3, limb weakness often points to a stroke, and tremor has been described in HDPC syndromes associated with *C9orf72*[202], and can also feature in SCA6 and SCA12[317]. Patients with neurodegenerative ataxias have much higher prevalences of tremor than HD patients, but in both cohorts the predominant type is essential tremor[318].

A criticism often levelled at the current definition of HDPC syndromes is that it includes patients that should not have been tested for HD in the first place because their symptoms are not sufficiently suggestive of HD[319]. However, as is obvious from this dataset, HD and HDPC patients presented to clinic with similar ages at onset; the different domains (motor, cognitive and psychiatric) were also affected at similar levels and combined at a comparable frequency. Singularly, family history (codified as the Goldman score), had a distinguishing effect, but may in an individual case be censored or complicated by uncertain paternity[320].

Based on single symptoms and even combinations of symptom domains, HD and HDPC patients were therefore exceedingly difficult to distinguish clinically. What emerged

from the analysis was that patterns may be more eloquent: clinical presentation of patients in whom a diagnosis of HD was considered did differ when simultaneous symptoms were analysed for associations. Relatively early orolingual involvement with dysarthria, dysphagia or choking in patients with cognitive problems or hyperkinetic limb movements appears suggestive of HD, exposing its involvement of diverse clinical domains. This may in part be due to the composition of our cohort of patients, since a comparison of the HDL2 phenotype to HD recently revealed that HDL2 patients presented with more dystonia and dysarthria than HD patients[321]. However, HDL2 is a famously good HD phenocopy and cognitive and psychiatric scores were still comparable between the two groups[321]. In a recent study, dysarthria in HD patients has been shown to fall into four distinct categories based on speech characteristics; different types of dysarthria differed significantly in terms of motor dysfunction, but, unfortunately, cognitive function was not assessed[322]. A literature review came to the conclusion that HD patients may also suffer from primary deficits of language[323], further confounding the motor aspect of speech and the cognitive task of understanding and producing language. Anyhow, given the high rate of aspiration pneumonia causing death in HD patients, corticobulbar rehabilitation should be offered to all patients with significant orolingual involvement early on in their disease course[324]. Interestingly, early insomnia was more common in HD than in HDPC patients, especially in combination with memory loss and depression, but weight loss was not. A link between sleep disturbance, cognitive decline and neuropsychiatric symptoms in HD has been previously reported[325]. While levels of depression and irritability were comparable in the two groups, the link between the two symptoms was stronger in HD rather than an HDPC syndrome. Irritability has previously been shown to be associated with an increase in motor symptoms early on in the HD disease course, and although this was not the case for depression, the authors do not comment on the use of antidepressant medication as a potential cause[326]. The level of neuropsychiatric symptoms appears to be greater in HD patients with a hypokinetic-dominant motor presentation, rather than the chorea-dominant or mixed-motor phenotype[327]. This is despite HD patients experiencing more non-motor symptoms overall than patients with Parkinson's disease.[328]

In contrast, in this study symptoms in HDPC patients seemed to commonly be less diverse and associations were stronger between related symptoms, such as cognitive slowing and irritability. Cognitive and psychiatric symptoms were mostly observed with a similar frequency in the two cohorts, but disinhibition (rather than depression, apathy or anxiety) should point the clinician towards an HDPC syndrome; this was especially true in combination with irritability, cognitive slowing and executive dysfunction. Interestingly, apathy, which has previously been reported to be a core feature of HD[300] and shown to present similarly to a cognitive symptom[279], did not feature in linked symptoms more likely to be found in one syndrome or the other in this dataset. This may be due to difficulties of a retrospective study such as gaps in the clinical documentation, or because of underlying patient heterogeneity.

4 Genetic causes of disease in HD phenocopy patients

A. Clinical and genetic heterogeneity assessed by WGS in 50 HDPC patients

As we have established, HDPC syndromes are clinically very similar to HD; the known HDL syndromes are caused by octapeptide repeats (HDL1 in *PRNP*), trinucleotide expansions (HDL2 in *JPH3* and HDL4 in *TBP* also known as SCA17), and unknown effects mapped to chromosome 4p15 (HDL3)[202]. In addition, there is considerable clinical overlap with other trinucleotide expansion disorders, especially SCA1, SCA2, SCA3, SCA6, SCA7, SCA8, SCA12, DRPLA, and Friedreich's ataxia, as well as other neurodegenerative diseases such as neuroferritinopathy, benign hereditary chorea, primary familial brain calcification, chorea-acanthocytosis, Wilson's disease, ataxia telangiectasia, and aceruloplasminemia[103,202].

C9orf72 expansions have been established as the most common genetic cause[13] of HDPC syndromes, but the diagnostic rate in the discovery study was still only 1.95%[13]. In terms of the diagnostic workup, this justifies testing first for *HTT* and then *C9orf72* expansions, followed by expansions in the *TBP* gene, which cause SCA 17 in approx. 1.1% of HDPC cases[103]; specific symptoms or details from the history may also justify

testing for other conditions, such as seizures for dentatorubral-pallidoluysian atrophy (DRPLA) or African heritage for HDL4/SCA17, although mixed-race may not always be known[329]. Once a patient with an HDPC syndrome has tested negatively for these conditions, the diagnostic rate plummets: as described above, for the MRC Dementia Gene panel, for example, even including *C9orf72* expansions and *PRNP* octapeptide repeat testing, a causal mutation was identified in only 4.1%, made up of 1.6% C9orf72 expansions and 2.5% DVs in other dementia genes[24]. Studies limited to the established HDL genes had even lower yields on genetic testing[330-333], suggesting novel variants or different genes may be causing HDPC syndromes and potentially acting as modifiers in Huntington's disease. A different approach is therefore needed in order to offer HDPC patients the benefits of an established diagnosis.

In all the HDPC patients included in this WGS study, the *HTT* expansion had been excluded in the Department of Neurogenetics at NHNN; following this, they were screened for mutations in known dementia genes, the *C9orf72* expansions and *PRNP* octapeptide repeats before being selected for WGS sequencing. To eliminate other trinucleotide expansions, samples were examined using the ExpansionHunter® programme, which identified one expansion in *ATXN1* in a patient in whom the diagnosis of SCA1 had been confirmed with a clinical test between sample selection and analysis; no other trinucleotide repeat expansions were found in any other samples. Given that HD, the established HDL disorders, and many other neurodegenerative diseases with dementia and motor symptoms are caused by trinucleotide repeats[103,202,317], excluding them is an important step in any diagnostic algorithm. While fragment analysis and southern blot remain the gold standard for the detection of trinucleotide expansions, these are incapable of detecting smaller variants[111,113,131]. Long-read sequencing, such as the ones offered by Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT), have recently tried to marry the SNP detection capabilities of next-generation sequencing with the potential to discover large genetic variations such as copy number variants and trinucleotide repeats, but are neither as accurate nor as reliable as more established platforms[128]; PacBio also remains prohibitively expensive. For the time being, bioinformatics analysis methods,

such as ExpansionHunter®, are one of the most accessible forms of screening WGS data for expansion disorders while also searching for SNPs and smaller indels; they have been employed with some success especially in modern PCR-free WGS datasets[130], but are not yet reliable enough to be employed on a larger scale or for the search for novel repeat expansions.

After trinucleotide expansions had been excluded in the HDPC samples, they were analysed using the IVA™ online platform; this revealed thirteen heterozygous likely deleterious variants in genes either linked to neurological disease, highly expressed in brain and intolerant to variation or with a biological link to HD or the HD-phenotype. However, missense mutations in general are common, with on average 2% of people carrying a mutation in any given gene[334], hence why the assessment of genetic tolerance and likely genetic impact is so important. In this analysis, several non-expansion variants were identified in genes linked to trinucleotide expansion disorders (see Results Chapter 3), but little evidence supported their being deleterious. While missense mutations have been observed as a rare cause of fragile X syndrome[335] and Friedreich's ataxia[336], the repeat expansions' likely trio of effects, namely RNA gain of function, protein loss of function, and gain of function of toxic non-ATG translated dipeptides are now well known[39].

The thirteen likely deleterious variants were found in 12 patients and linked to diverse cellular functions: lysosomal storage (*CTSB*), transcription (*CREBBP*, *STAT5B*) and mitochondrial transcription (*POLG*), WNT signalling (*APC*), neuron-associated functions such as ion channels and axonal transport (*SCN1A, MAPK8IP3, DYNC1H*), and a growth factor (*PDGFB*). Many of these functions are commonly involved in the pathogenesis of neurodegenerative diseases, such as lysosomal dysfunction playing a role in Parkinson's disease[337] and WNT signalling a major factor in AD[338]. No particular gene or cellular pathway was clearly overrepresented in the results potentially indicating a common disease mechanism.

Given that this was a small study in a very heterogeneous study population, it is hardly surprising that no pertinent, known deleterious variants were identified. Novel and understudied variants may play an outsize role in these less well-established syndromes, as was the case in the MRC Dementia Panel study[24] described above; genes not previously established to cause disease may also be involved. Power and penetrance pose a particular challenge in rare diseases like HDPC disorders. While a variant found a single time in the dataset and a single time on *GnomAD* would statistically be significant, it would be estimated to only have a penetrance of 7.1%; the *MAPK8IP3* (p.K584*) variant, which was observed four times, would only be expected to have a penetrance of 28.3% (95% CI 1.96%- 394.1%) if it were observed once on *GnomAD*[123,124]. These calculations justified using a filter of zero occurrences on *GnomAD*, but the numbers observed in our dataset are approaching the limits of how often the same rare variant can be observed in a small sample and still be considered rare. Despite strong family histories of neurodegenerative disease, no patient relatives' samples were available for segregation analysis.

B. Investigating variant causality

a. Variant replication

Principally, all the candidate genes and variants that were discovered require replication and follow up work in order to be proven deleterious. The natural solution to improve the hit rate and statistical power is to build the discovery set and sequence more genomes, but this will be limited by available funding. Despite the up-front cost, a non-hypothesis-driven approach, i.e. WGS, is appropriate since it supports the creation of an optimal resource for further study of not only the exome, but also the non-coding regions, such as promoters, and regulatory sequences down the line; this may be especially relevant because not all of them may already be known. In addition, WGS has been shown to provide more homogeneous coverage with higher genotyping quality, to be less susceptible to allelic drop outs[117], and to identify more variants[117,118].

An alternative would be to perform genotyping assays for single variants in a replication cohort; it is cheap and quick, and avoids having to correct for multiple

testing. However hypothesis testing single variants would may miss similar but not identical variants with equally damaging biological effect and equally compelling evidence. HDPC disorders are more likely to be caused by a set of variants with similar biological effects or meeting certain criteria, such as the splice-site disrupting variants in *MAPT*[36,339] or loss-of-function variants in *GRN*[340,341]. An example would be the two *CTSB* variants c.422-1G>A and p.G141E identified in this study, which affected the same splice-site spanning codon in two patients with similar presentations; either one or both would be missed if only single loci are genotyped, but in combination they warrant further investigation. Since DNA stocks of patients are finite, and patients often not easily available for further sampling, these samples are precious; exhausting them for genotyping is therefore a fraught decision.

Alternatively, it is possible to attempt to follow up on proposed mechanisms using *in vitro* studies and other biological investigations; however, these are complex and time-consuming and may ultimately not be relevant to the disease in question. Replication in a dedicated cohort, if funding allows, or in a publicly-available dataset such as Genomics England is therefore generally preferable.

b. *Establishing gene causality*

Rather than identifying previously published deleterious variants, the primary aim of small WGS discovery studies is often to identify candidate genes, that may play a role in the pathogenesis, and to rule out clinical confounders.

After careful selection of samples and sequencing, samples were checked for known expansion disorder triplet repeat expansion and filtered as described in *Materials and Methods: Whole-genome sequencing data analysis*. Crucial to this step are large public population databases, such as *GnomAD*[123,124], which are tied into the IVA™ platform and facilitate the exclusion of variants above a certain threshold. As this study deals with a rare disorder with a prevalence of approx. 2.5/100,000 (see Results Chapter 3), only variants, which are either absent from genetic population databases or very rare, are taken into account, since deleterious variants are overwhelmingly rare on population databases[342], and the frequency of a variant in population must be in line

with the prevalence of the studied disease[28]. Given the stringent criteria applied before any variants can be called benign or likely benign[343], these were also excluded from the pool of potentially causal mutations, as were synonymous or non-coding mutations since these are much more difficult to prove to have a deleterious impact using a non-computational approach[344,345]. The analysis was focused on the exome and adjacent non-coding stretches where most deleterious variants are thought to be located[346].

As described in Results Chapter 4, variants that remained after stringent filtering were assessed for their likely impact; thirteen variants in eight genes were deemed to be highly suspicious of causing disease (Results Chapter 4, Table III-16). Features that were taken into account included protein expression in the brain (expressed as the Gtex score[181]), tolerance to genetic variation (expressed as the RVIS score[180]), and variant characteristics (according to the ACMG criteria[343]). Of particular interest were *CTSB* c.422-1G>A and *CTSB* p.G141E which were found in the same splice-site spanning codon in two patients with a similar phenotype, *CREBBP* p.A1062V which is a gene that has previously been associated with the HD protein[205-207,266], and *STAT5B* p.A693T which is located in a highly conserved domain required for activation[271]. Yet all of these variants require further evidence to be considered actionable.

Variants identified as potentially deleterious in the HDPC discovery dataset were therefore cross-checked in the Genomics England dataset; however none could be replicated this way. As a next step, variant locations in gene functional domains were precisely defined and checked for novel variants in the in the Genomics England dataset; this revealed five variants, two of which were predicted benign *in silico*. Of the remainder, two variants were missense variants in *MAPK8IP3*, and one was a missense variant in *STAT5B*. The clinical phenotypes associated with the variants in both *MAPK8IP3* and *STAT5B* in the HDPC discovery and the Genomics England datasets complement each other; however since they are not an exact match, they require further replication before they can be considered causal.

It is also useful to exclude any variant or gene found in a particular patient which, even if deleterious and relevant, may be more indicative of a different disease than the one being studied. An example of such clinical confounders in this study is the patient with the concurrent *POLG* (p.C208*) and *PDGFB* (p.K146fs*10) variants who presented with behavioural symptoms and chorea. Although some of his symptoms clinically mimic HD, the lack of cognitive decline and the brain calcifications on the patient's MRI are an indication that this may not be a perfect HD phenocopy. *POLG* mutations are a frequent cause of ataxia[304,347,348], but not dystonia or chorea[273]. On the other hand, *PDGFB* causes 10% of primary familial brain calcification (PFBC)[349], which are typically linked to brain calcifications, cognitive decline, motor symptoms[262], including chorea, and neuropsychiatric problems[349]. In fact, the patient's symptoms are a much better match for PFBC rather than HD, which can serve as a reminder to carefully examine all available clinical information and investigations, but also how very different diseases can overlap clinically.

However, none of the genes with highly suspicious variants were known to be causally linked to an HDPC phenotype. In order to establish a new causal gene or variant, additional evidence is required. Traditionally, linkage or segregation studies in large kindreds are used to prove causality of a given variant[350-352]. Comparing affected and unaffected relatives' genetic information offers a pertinent way to ascertain which mutations may be relevant; trios of parents and child are an established way to identify causal mutations, even *de novo*, in paediatric cases[353], but are rarely feasible in late-onset diseases such as dementia. Alternatively, variants may meet certain criteria in a given gene, such as loss-of-function mutations in *GRN*, where haploinsufficiency is a known disease mechanism in FTD[340], or even across a pathway. The latter, however, presumes that the gene in question and the disease mechanism are already known. None of the highly suspicious variants identified in this study meet these criteria, although some have been shown to interact with the *HTT* protein or are involved in cellular functions thought to be relevant to neurodegenerative disease (see Results Chapter 4 for more details).

So as to be able to confirm the pathogenic effect of a suspicious mutation in a gene not previously associated with a given disease in the absence of segregation data, it is necessary to provide irrefutable statistically valid genetic evidence linking the mutation to a disease phenotype[354]. Attaining biological evidence for the disruptive effect of a mutation can offer additional evidence, but in itself is insufficient proof. It is also a lengthy process involving the in-depth study of pathway and cellular consequences *in vitro* and *in vivo*, as has been done for mutations in *MAPT*[36,355], for example. However, this process is uncertain of success and beyond the scope of this thesis. The gold standard for providing genetic evidence for the pathogenicity of a given variant are the ACMG criteria[343], but none of the highly suspicious variants in this study have sufficient evidence available to be called as a proven deleterious mutation to be reported back to a patient in a clinical setting. In order to be established as clinically proven, results need to be replicated, and genes and variants detected in at least two separate pedigrees, ideally with segregation. With the exception of *MAPK8IP3* p.K584*, all suspicious variants in this study were observed only once, although *CTSB* c.422-1G>A and p.G141E affect the same splice-site spanning codon. None could be replicated in the approx. 14,000 cases with varied neurological disorders in the Genomics England 100,000 Genome project[356]. This may be due to the rarity of HDPC disorders, the heterogeneity of the phenotype, as yet unanalysed non-coding sequence variants, expansions or other undetected structural variants.

C. Genetic modifiers, pleiotropy and polygenicity

Arguments to support both the rare-variants-large-effect and the common-variants-small-effects theory in common diseases can be made convincingly[357], but Mendelian diseases differ from traits determined by common variation in that the effect is usually understood to be dichotomous, either a variant causes disease or not. However, even these seemingly clear cut cases are confounded by incomplete penetrance and mounting evidence that they are also subject to disease-modifying effects of other genetic variants[100,358-361]; for example, late age at onset (AAO) can masquerade as incomplete penetrance, since patients may die of other causes before the disease manifests itself. In addition, as I have shown in Results Chapter 2, ApoE4, an unrelated

genetic factor, influences AAO in FTD-tau patients[296], and in HD, variants in the DNA damage repair pathway influence both AAO and disease progression[100].

In genome-wide association studies (GWAS), polygenicity and pleiotropy have long been recognized as major factors in how complex traits manifest themselves[362,363]. A recent GWAS study revealed that highly pleiotropic genes are most likely to affect general biological functions, especially regulation of transcription[346]. Several of the highly suspicious variants identified in this study were found in genes linked to transcription, such as *CREBBP* (p.A1062V) and *STAT5B* (p.A693T and p.V467A); mutant huntingtin has also been shown to disrupt transcription and DNA repair, although the mechanism of neurotoxicity remains unclear[364]. Beyond GWAS and general traits, pleiotropy may explain some of the missing heritability observed in neurodegenerative diseases; a recent study in FTD patients, identified several rare potentially pathogenic mutations in genes known to cause neurodegeneration but not previously associated with FTD[285], novel variants were enriched in the DemMot cohort in the MRC Dementia Gene Panel study[24], an more generally, rare novel variants have been shown to be associated with disease[365].

Both in the MRC Dementia Gene Panel study and in the HDPC WGS study, several patients carried concurrent mutations, with clinical presentations potentially complicated by interactions between deleterious effects. Concurrent suspicious mutations were noted in three cases: one case carried both the *CTSB* (p.G141E) and a *MAPK8IP3* (p.K584*) variant, another was found to carry both the *CTSB,* (c.422-1G>A) and the *APC* (p.S52R) variant, and another was identified to have both the *APC* (p.A1991T and *STAT5B* (p.V467A) variant. Clinicians and researchers are increasingly more aware of the potential presence of more than one concurrent mutation; in an effort to at least partially automate the search for and the assessment of such double variants, a new computer algorithm has been developed[366]. However, tools like these have not yet been evaluated for their precision and usefulness and are not yet in wide-spread use. It seems likely that epistatic factors are at play in most neurodegenerative diseases leading to decompensation of what may otherwise have been a stable

biological effect; they may play an oversized effect in heterogeneous phenotypes like HDPC disorders.

However, for the time being at least, focus still remains on single causal variants with near complete penetrance. This is simply because genetic modifiers, effects of concurrent mutations and GWAS-derived polygenic risk scores are still poorly understood. Clinical applicability of polygenic risk scores remains severely hampered by their continuing limited ability to predict individual patient symptoms[32,279]. This may in part be due to the exponentially increasing complexity with increasing numbers of genes[346], as well as the effects of differing risks in diverse populations of varied ancestry[367]. In the future, patients may be prospectively sequenced and entered into intervention trials based on their risk scores and other known genetic factors.

5 Conclusion

This thesis has run the gamut from well-established Mendelian causes of dementia to a subcohort of poorly understood and underdiagnosed HDPC patients it has spanned the range of genetic sequencing, including genotyping, fragment analysis, Sanger sequencing, amplicon-based targeted sequencing and whole-genome sequencing. While the first chapter dealt with determining the likelihood of finding one of the known mutations in different dementia syndromes using a targeted gene panel supplemented by additional tests, the second explored the disease-modifying effect of ApoE4, the commonest risk factor for AD, on FTD patients; the third chapter sought to identify distinguishing clinical features for HDPC patients, and finally the fourth chapter explored genetic causes for these underdiagnosed, heterogeneous patients using whole-genome sequencing.

Dementia disorders pose considerable problems to the discerning clinician since they can overlap clinically with each other, making a precise diagnosis difficult, while different genetic causes of dementia may lead to clinically similar syndromes. This project sought to alleviate some of the challenges of selecting patients for clinical genetic testing and interpreting the results. While the traditional clinical predictors for patient selection - age at onset and family history - hold fast and have been better

defined in my sample series, it is becoming increasingly clear that genetic dementia syndromes can manifest into old age and genetic factors may also affect disease progression[100,281]. In addition, evidence is emerging that the genes known to be linked to neurodegeneration have wider-ranging phenotypes than previously thought and that more hitherto unrecognized deleterious variants in these genes are yet to be discovered[116]. More comprehensive genetic testing of a more diverse selection of dementia patients promises yields even in the genes that are already known; the patients most likely to profit from these efforts are those with atypical syndromes and limited family history. One way of assessing the benefits of such a change in strategy would be to compare a new intensified genetic testing regime, such as running a dementia gene panel with *C9orf72* expansion testing on any dementia patient with an AAO under 60 years or a Goldman score of 3.5 or lower, with the established practice of only testing patients with a very strong family history (GS 1 and 2) or other suggestive factors. Such a comparison could be used to assess not only discovery rates of genetic diagnoses, but also the impact on patient quality of life, length of survival and speed of disease progression, as well as financial savings due to reduced other investigations. Novel variants are most likely to be found in these patients, and while this poses unique challenges, it could also offer new insights into pathogenic links and cellular pathways. Concurrent pathogenic mutations were observed in excess in this MRC Dementia Gene Panel dataset, particularly in FTD, suggesting that epistasis of Mendelian variants may contribute significantly to dementia. This large-scale sequencing effort underlines the need for broader clinical testing than was previously customary because of technical and financial constraints, which in the era of next-generation sequencing can finally become a reality. Caution is, however, necessary where the assessment and classification of genetic variants is concerned, as not all published mutations appear to be fully penetrant or even pathogenic. It is imperative to assess the frequency of a variant in the population to prevent false flags to be raised in genetic counselling of patients. More extensive genetic studies of more broadly selected patient cohorts in combination with population-based analyses promise to enhance our understanding of dementia syndromes and improve patient diagnosis and treatment.

In addition to Mendelian causes of disease, common variation, such as the ApoE genotypes, is increasingly recognized as a disease-modifying factor[32,368]. Research interest is growing, but clinical applicability is still severely limited because of the complexity of the numerous genes and loci involved, and because they do not yet offer any clinical benefit to healthy patients. Nonetheless, these insights are broadening our understanding of the underlying processes; with improving and expanding data to mine, and increasing computing capabilities, clinical utility of risk prediction for some traits, such as the likely appearance of specific symptoms[278], may be in the near future. Further research is needed to establish prospective evidence of clinical benefit to patients at the beginning of their disease and the healthy population.

Distinguishing HD and HDPC disorders has been reassuringly difficult, testament to the excellent skills of referring clinicians. HDPC disorders are likely to group highly heterogeneous patients because HD itself is one of the most heterogeneous genetic neurological diseases. When the *HTT* gene expansion was first identified over 25 years ago[98], testing was still relatively arduous and reserved only for the most convincing families and typically only returned a negative result in approx. 1% of suspected HD patients[102,167]. Since then, HD testing has become much more accessible and it has become customary to exclude HD even if a patient's presentation may not be a textbook case. Given that, as shown above, HD and HDPC phenotypes overlap significantly, and that it is so difficult to clinically rule out HD, this is reasonable. Not surprisingly in such a small study of heterogeneous understudied patients, deleterious point mutations in known genes pertinent to the phenotype were not found in the HDPC WGS dataset. Given the results of the MRC Dementia Gene Panel, the novel highly suspicious variants with limited available evidence are likely to be causal in a significant proportion of cases, either as single Mendelian variants or in combination with other epistatic factors, but further confirmatory studies are required.

6 Further work

There are multiple options to both mine more deeply into the data and expand the studies and analyses that the present thesis is based on.

The MRC Dementia Gene Panel study has been explored in terms of predictors of gene discovery and diagnostic rates, but could be analysed further to

- Perform burden testing in the individual cohorts to assess novel / uncertain / synonymous variants in these known dementia genes and to look for protective effects
- Correlate genetic data with clinical presentation, i.e. certain symptoms, speed of disease progression, AAO, etc.

In order to overcome the greatest limitations of the HD - HDPC clinical comparison, which was its retrospective nature and limited documentation of clinical signs, the study could be set up as a prospective comparison, which would allow:

- Detailed documentation of promising symptoms which may be underdocumented, such as eye movements
- Following up patients to analyse how their symptoms develop and how symptom combinations change
- Symptom network analysis over time
- Include followed-up patients in genetic study

The genetic component of the HDPC cohort study could be developed to include:

- Analysis of the WGS dataset with regards to variants in the non-coding sequence, promoters, and regulatory elements
- Burden testing to identify relevant genes
- Exploration of concurrent variants using a recently developed computer algorithm
- CNVs and other structural variants
- Re-contacting of patients with the most likely candidate variants to ask whether any family members would be willing to donate a DNA sample
- If and when a candidate gene does emerge, the remainder of the cohort could be re-analysed for less conspicuous variants in the same gene

- Correlate genetic data with clinical presentation, i.e. certain symptoms, AAO, age at death, disease duration. The HDPC database will be of particular use for this.
- Replication study in an extended UCL HDPC cohort

  Funding for the replication of the highly suspicious variants identified through WGS in the HDPC discovery set has already been secured and a selection of further samples will be whole-genome sequenced and analysed. With a larger dataset, a targeted analysis of non-coding sequences is more likely to be successful, since it depends heavily on computational and statistical methods. Confirmation of the involvement of new genes and variants would permit to further the biological pathways involved in the pathogenesis of HD and HDPC disorders, which may be exploited for therapeutic purposes later on.

V. Appendix

Table V-1: The MRC dementia gene panel comprised 17 genes

The MRC dementia gene panel comprised 117 regions from 17 genes which are set out below with their respective start and end points based on the genomic location and which exons correspond to this location. Intronic regions flanking these exons were also included.

Gene	Chromosome	Start	End	Exon	Gene Section
APP	chr21	27253982	27254092	Exon 18/18	1
APP	chr21	27264025	27264190	Exon 17/18	2
APP	chr21	27269876	27269995	Exon 16/18	3
APP	chr21	27277326	27277399	Exon 15/18	4
APP	chr21	27326894	27327013	Exon 13/18	5
APP	chr21	27327931	27328079	Exon 12/18	6
APP	chr21	27348257	27348351	Exon 10/18	7
APP	chr21	27354647	27354800	Exon 9/18	8
APP	chr21	27425542	27425674	Exon 4/18	9
APP	chr21	27462249	27462398	Exon 3/18	10
APP	chr21	27542872	27542972	Exon 1/18	11
CHMP2B	chr3	87302545	87302670	Exon 4/5	1
CHMP2B	chr3	87302853	87302983	Exon 5/5	2
CSF1R	chr5	1.49E+08	1.49E+08	Exon 22/22	1
CSF1R	chr5	1.49E+08	1.49E+08	Exon 21/22	2
CSF1R	chr5	1.49E+08	1.49E+08	Exon 20/22	3
CSF1R	chr5	1.49E+08	1.49E+08	Exon 19/22	4
CSF1R	chr5	1.49E+08	1.49E+08	Exon 18/22	5
CSF1R	chr5	1.49E+08	1.49E+08	Exon 17/22	6
CSF1R	chr5	1.49E+08	1.49E+08	Exon 16/22	7
CSF1R	chr5	1.49E+08	1.49E+08	Exon 15/22	8
CSF1R	chr5	1.49E+08	1.49E+08	Exon 14/22	9
CSF1R	chr5	1.49E+08	1.49E+08	Exon 13/22	10
CSF1R	chr5	1.49E+08	1.49E+08	Exon 12/22	11
FUS	chr16	31193825	31193995	Exon 3/15	1
FUS	chr16	31195521	31195727	Exon 5/15	2
FUS	chr16	31196251	31196510	Exon 6/15	3
FUS	chr16	31201587	31201729	Exon 12/15	4
FUS	chr16	31202054	31202173	Exon 13/15	5
FUS	chr16	31202275	31202441	Exon 14/15	6
FUS	chr16	31202711	31202770	Exon 15/15	7
GRN	chr17	42422462	42422732	Exon 1/13	1
GRN	chr17	42426497	42426929	Exon 2-3/13	2

Gene	Chromosome	Start	End	Exon	Gene Section
GRN	chr17	42427026	42427129	Exon 4/13	3
GRN	chr17	42427587	42427718	Exon 5/13	4
GRN	chr17	42427801	42427955	Exon 6/13	5
GRN	chr17	42428050	42428178	Exon 7/13	6
GRN	chr17	42428396	42428541	Exon 8/13	7
GRN	chr17	42428722	42428838	Exon 9/13	8
GRN	chr17	42428909	42429173	Exon 10/13	9
GRN	chr17	42429374	42429626	Exon 11/13	10
GRN	chr17	42429700	42429949	Exon 12/13	11
GRN	chr17	42430020	42430176	Exon 13/13	12
ITM2B	chr13	48835236	48835401	Exon 6/6	1
MAPT	chr17	44039655	44039886	Exon 2/14	1
MAPT	chr17	44068777	44069002	Exon 9/14	2
MAPT	chr17	44073716	44074080	Exon 10/14	3
MAPT	chr17	44087627	44087818	Exon 11/14	4
MAPT	chr17	44091560	44091740	Exon 12/14	5
MAPT	chr17	44095935	44096146	Exon 13/14	6
MAPT	chr17	44101273	44101587	Exon 14/14	7
NOTCH3	chr19	15302762	15303119	Exon 4/33	1
NOTCH3	chr19	15303179	15303340	Exon 3/33	2
PRNP	chr20	4679858	4680638	Exon 2/2	1
PSEN1	chr14	73614719	73614824	Exon 3/12	1
PSEN1	chr14	73637496	73637765	Exon 4/12	2
PSEN1	chr14	73640265	73640425	Exon 5/12	3
PSEN1	chr14	73653552	73653638	Exon 6/12	4
PSEN1	chr14	73659343	73659582	Exon 7/12	5
PSEN1	chr14	73664730	73664847	Exon 8/12	6
PSEN1	chr14	73673085	73673190	Exon 9/12	7
PSEN1	chr14	73678468	73678660	Exon 10/12	8
PSEN1	chr14	73683825	73683962	Exon 11/12	9
PSEN1	chr14	73685833	73686007	Exon 12/12	10
PSEN2	chr1	2.27E+08	2.27E+08	Exon 5/13	1
PSEN2	chr1	2.27E+08	2.27E+08	Exon 6/13	2
PSEN2	chr1	2.27E+08	2.27E+08	Exon 7/13	3
PSEN2	chr1	2.27E+08	2.27E+08	Exon 8/13	4
PSEN2	chr1	2.27E+08	2.27E+08	Exon 13/13	5
SERPINI1	chr3	1.68E+08	1.68E+08	Exon 2/9	1
SERPINI1	chr3	1.68E+08	1.68E+08	Exon 3/9	2
SERPINI1	chr3	1.68E+08	1.68E+08	Exon 4/9	3

Gene	Chromosome	Start	End	Exon	Gene Section
SERPINI1	chr3	1.68E+08	1.68E+08	Exon 5/9	4
SERPINI1	chr3	1.68E+08	1.68E+08	Exon 6/9	5
SERPINI1	chr3	1.68E+08	1.68E+08	Exon 7/9	6
SERPINI1	chr3	1.68E+08	1.68E+08	Exon 8/9	7
SERPINI1	chr3	1.68E+08	1.68E+08	Exon 9/9	8
SQSTM1	chr5	1.79E+08	1.79E+08	Exon 1/8	1
SQSTM1	chr5	1.79E+08	1.79E+08	Exon 2/8	2
SQSTM1	chr5	1.79E+08	1.79E+08	Exon 3/8	3
SQSTM1	chr5	1.79E+08	1.79E+08	Exon 4/8	4
SQSTM1	chr5	1.79E+08	1.79E+08	Exon 5/8	5
SQSTM1	chr5	1.79E+08	1.79E+08	Exon 6/8	6
SQSTM1	chr5	1.79E+08	1.79E+08	Exon 7/8	7
SQSTM1	chr5	1.79E+08	1.79E+08	Exon 8/8	8
TARDBP	chr1	11073764	11074032	Exon 2/6	1
TARDBP	chr1	11076892	11077074	Exon 3/6	2
TARDBP	chr1	11078781	11078940	Exon 4/6	3
TARDBP	chr1	11080477	11080666	Exon 5/6	4
TARDBP	chr1	11082172	11082722	Exon 6/6	5
TREM2	chr6	41126492	41126528	Exon 5/5	1
TREM2	chr6	41126602	41126814	Exon 4/5	2
TREM2	chr6	41127521	41127630	Exon 3/5	3
TREM2	chr6	41128992	41129361	Exon 2/5	4
TREM2	chr6	41130772	41130924	Exon 1/5	5
TYROBP	chr19	36395461	36395546	Exon 5/5	1
TYROBP	chr19	36398111	36398176	Exon 4/5	2
TYROBP	chr19	36398339	36398492	Exon 3/5	3
TYROBP	chr19	36398623	36398674	Exon 2/5	4
TYROBP	chr19	36399061	36399221	Exon 1/5	5
VCP	chr9	35057104	35057229	Exon 17/17	1
VCP	chr9	35057364	35057537	Exon 16/17	2
VCP	chr9	35059052	35059226	Exon 15/17	3
VCP	chr9	35059481	35059808	Exon 14/17	4
VCP	chr9	35060301	35060532	Exon 13/17	5
VCP	chr9	35060789	35060930	Exon 12/17	6
VCP	chr9	35061003	35061186	Exon 11/17	7
VCP	chr9	35061565	35061696	Exon 10/17	8
VCP	chr9	35061991	35062145	Exon 9/17	9
VCP	chr9	35062205	35062357	Exon 8/17	10
VCP	chr9	35062966	35063087	Exon 7/17	11

Gene	Chromosome	Start	End	Exon	Gene Section
VCP	chr9	35064142	35064292	Exon 6/17	12
VCP	chr9	35065239	35065388	Exon 5/17	13
VCP	chr9	35066663	35066824	Exon 4/17	14
VCP	chr9	35067879	35068070	Exon 3/17	15
VCP	chr9	35068239	35068369	Exon 2/17	16
VCP	chr9	35072325	35073256	Exon 1/17	17

Table V-2: Genes very tolerant to genetic variation based on their RVIS score and uploaded as an Ingenuity filter

Gene	Gene	Gene	Gene	Gene
ANKLE1	*EFNA2*	*IFT140*	*OR1L3*	*SIRT5*
AR	*EME2*	*IGFALS*	*PCDHA4*	*SMPD1*
C17orf97	*ERAP2*	*LMF1*	*PCDHA8*	*SSTR5*
CCDC157	*FDXR*	*MRPL9*	*PCDHGA4*	*TAS1R1*
CENPP	*FOXD4*	*MRPS34*	*PDCD6*	*TMIE*
CEP89	*GALC*	*NDUFB2*	*PODXL*	*TPSD1*
CHTF18	*HRC*	*NME3*	*RAET1E*	*TPSG1*
DNAH11	*HRCT1*	*NUBP2*	*RSPH14*	*ZGRF1*
DUOX2	*IFNAR2*	*OAZ3*	*SEC16B*	*ZNF28*

Table V-3: Genes with a GTeX expression of below 10 TPM in brain and uploaded as an Ingenuity filter

Gene	Gene	Gene	Gene
ADAMTSL1	*FRMD3*	*PCDHA5*	*SLC19A1*
AHRR	*FTCDNL1*	*PCDHA6*	*SLC3A1*
ANKDD1B	*GAGE1 (includes others)*	*PCDHB16*	*SOX8*
ANKRD13D	*GOLGA6L1 (includes others)*	*PCDHB9*	*SRSF10*
ASPN	*HEATR1*	*PCDHGA1*	*SSTR5-AS1*
ATXN3	*HSPB11*	*PCDHGA10*	*STAT5B*
C1orf141	*HTT-AS*	*PCDHGA5*	*SYCP2*
C9orf3	*JPH3*	*PCDHGA6*	*TM4SF19*
CBWD1	*KDM5A*	*PCDHGA7*	*TM4SF19-AS1*
CCDC77	*KIFC1*	*PCDHGA8*	*TMEM161B*
CDCA7L	*KLF14*	*PCDHGB1*	*TMEM247*
CEACAM20	*KNTC1*	*PCDHGB2*	*TMEM99*
CHODL	*KRT8P50*	*PCDHGB3*	*TRIM61*
CHRNA1	*LRCH2*	*PCDHGB4*	*TRIM64/TRIM64B*
CHRNA3	*LRP1B*	*PCDHGB6*	*TRIM64C*
COL4A3BP	*MADCAM1*	*PI4K2B*	*U7*
COLCA1	*MAP3K1*	*PIGL*	*UNGP3*
CRYZL2P-SEC16B	*MYCBPAP*	*POLK*	*ZNF234*
DBT	*NDUFB2-AS1*	*PTRH1*	*ZNF277*
DCAF8L2	*NEBL-AS1*	*PYROXD1*	*ZNF283*
DDX20	*NEURL2*	*RAET1E-AS1*	*ZNF384*
DQX1	*OR4E1*	*RBMXL3*	*ZNF595*
EIF2AK3	*OR4F17 (includes others)*	*RSPH10B/RSPH10B2*	*ZNF714*
FAM110C	*OR7A17*	*SAMD9L*	*ZNF718*
FAM218A	*OR8G1*	*SCGB1C1/SCGB1C2*	*ZNF728*
FAM35CP	*PCDHA1*	*SCN9A*	
FOXO6	*PCDHA2*	*SEPSECS*	

Table V-4: Neurodegenerative disease genes uploaded to Ingenuity as a filter

Gene	Gene	Gene	Gene	Gene	Gene	Gene
AARS	*BSCL2*	*DST*	*GLB1*	*MFSD8*	*PRPS1*	
ABCA7	*C12orf65*	*DYNC1H1*	*GM2A*	*MSH3*	*PRRT2*	
ABCD1	*C19orf12*	*DYT1*	*GNAL*	*NAGA*	*PSAP*	
ADCY5	*C20ORF54*	*DYT13*	*GNPTAB*	*NAGPA*	*PSEN1*	*SURF1*
AFG3L2	*C9orf72*	*DYT14*	*GNS*	*NDUFS1*	*PSEN2*	*TAF1*
AGA	*CACNA1A*	*DYT15*	*GRID2*	*NDUFS4*	*RAB3GAP1*	*TARDBP*
AIMP1	*CACNA1a*	*DYT6*	*GRN*	*NDUFS7*	*RAB7*	*TBC1D20*
ALDH3A2	*CACNA1G*	*DYT7*	*GUSB*	*NDUFS8*	*RARS*	*TBK1*
ALG3	*CACNB4*	*DYT8*	*HEXA*	*NDUFV1*	*RNF216*	*TBK-1*
ALS2	*CACNL1A4*	*EEF2*	*HEXB*	*NEFH*	*SACS*	*TBP*
AMACR	*CCDC88C*	*EFG1*	*HGSNAT*	*NEU1*	*SCA25*	*TDP43*
ANG	*CCT5*	*EGR2*	*HNRNPA1*	*NGF*	*SCN2A*	*TGM6*
ANO3	*CHCHD10*	*ELOVL4*	*HNRNPA2B1*	*NKX2-1*	*SCN9A*	*TH*
ANXA11	*CHMP2B*	*ELOVL5*	*HSP22*	*NOP56*	*SDHA*	*THAP1*
APP	*CLCN2*	*ELP1*	*HSP27*	*NOTCH3*	*SERPINI1*	*TIMM8A*
ARSA	*CLN3*	*ERBB3*	*HTRA1*	*NPC1*	*SETX*	*TMEM126A*
ASAH1	*CLN6*	*EWSR1*	*HYAL1*	*NPC2*	*SGCE*	*TMEM240*
ASPA	*CLN8*	*FA2H*	*IDS*	*NR2F1*	*SGSH*	*TP*
ATL1	*COQ2*	*FAM134B*	*IDUA*	*NTRK1*	*SLC12A6*	*TPP1*
ATM	*COQ8A*	*FGD4*	*IFRD1*	*OPA1*	*SLC16A2*	*TREM2*
ATN1	*COQ9*	*FGF14*	*IGHMBP2*	*OPA3*	*SLC17A5*	*TSC1*
ATP12	*COX10*	*FIG4*	*ITM2B*	*OPTN*	*SLC20A2*	*TSC2*
ATP1A3	*CSF1R*	*FMRI*	*ITPR1*	*PANK2*	*SLC25A4*	*TTBK2*
ATP7A	*CTDP1*	*FTL*	*JPH3*	*PARK2*	*SLC2A1*	*TTC3*
ATP7B	*CTNS*	*FTX*	*KCNA1*	*PARK7*	*SLC52A2*	*TUBA4A*
ATPAF2	*CTSA*	*FUCA1*	*KCNC3*	*PDGFB*	*SLC52A3*	*TUFM*
ATXN1	*CTSD*	*FUS*	*KCND3*	*PDGFRB*	*SMN1*	*TWNK*
ATXN10	*CTSK*	*GAA*	*KIF7*	*PDYN*	*SMPD1*	*TYROBP*
ATXN10	*CYP27A1*	*GALC*	*LAMP2*	*PEO1*	*SNCA*	*UBQLN2*
ATXN2	*DAO*	*GALNS*	*LIPA*	*PFN1*	*SOD1*	*UBQLN2*
ATXN3	*DARS*	*GARS*	*LMNB1*	*PINK1*	*SPG11*	*UCHL-1*
ATXN7	*DARS2*	*GBA*	*LRPPRC*	*PLP1*	*SPG21*	*VAMP1*
ATXN7	*DCTN1*	*GBA*	*LRRK2*	*POLG*	*SPG7*	*VAPB*
ATXN8	*DJ-1*	*GBA2*	*MAN2A1*	*POLG2*	*SPR*	*VCP*
ATXN8	*DNAJB2*	*GCH1*	*MANBA*	*PPP2R2B*	*SPTBN2*	*VPS13A*
ATXN80S	*DNAJC5*	*GDAP1*	*MAPT*	*PPT1*	*SPTLC1*	*VPS35*
BAG3	*DNCT1*	*GFAP*	*MATR3*	*PRKCG*	*SPTLC2*	*WDR45*
BCS1L	*DNM2*	*GJB1*	*MECP2*	*PRNP*	*SQSTM1*	*WNK1*
BEAN1	*DNMT1*	*GLA*	*MFN2*	*PRPH*	*SUMF-1*	*XK*
						XP-A

Table V-5: Tabulation of novel DVs.

Comments refer to the main lines of evidence used to justify the classification based on rules defined by Table II-6 and Table II-7.

Pathogenicity	Times observed	AAO	Type	Gene	Alleles / GnomAD	HGVSProtein (or coding if no change in protein)	Comment
Likely Deleterious	3	65.5, 55.5, 55.6	missense	*CSF1R*	15	p.Leu868Arg	No variant at this position on *ExAC.* Leu 868Pro reported pathogenic
Deleterious	3	64, 62	frame-shift	*GRN*	0	p.Gln130Serfs	Frameshift mutations in GRN cause haploinsufficiency and disease. Several mutations at this codon causing deleterious frameshift mutations have been reported on Molgen
Deleterious	3	?, 25.5, 64	frame-shift	*GRN*	0	p.Ser78Phefs	Not on *ExAC.*, causes a premature stop codon. frameshift mutations in GRN reported pathogenic, pathogenic frameshift mutation on Molgen in the adjacent codon
Deleterious	2	?, ?	frame-shift	*GRN*	0	p.Ser129Lysfs	frameshift mutation in GRN, adjacent to two other frameshift mutation at codon 130 reported deleterious
Likely Deleterious	2	52, 52,	missense	*MAPT*	0	p.Gln351Arg	Not on Molgen, not on *ExAC.*, not on exome variant server. In silico predictions conflicting but mostly damaging. Amino acid change form uncharged polarised to positively charged, only slightly bigger.
Likely Deleterious	2	50, 59	missense	*MAPT*	0	p.Gly271Arg	Not on Molgen, not on *ExAC.*, not on EVS, not on google. Big amino acid change between a synonymous (benign) and a pathogenic mutation

Pathogenicity	Times observed	AAO	Type	Gene	Alleles / GnomAD	HGVSProtein (or coding if no change in protein)	Comment
Likely Deleterious	1	44	missense	*CSF1R*	0	p.Ala891Pro	Not on *ExAC.*, not on EVS. In silico not consistent. Mutation located in intracellular tyrosine kinase domain Clinical description fits Hereditary Diffuse Leukoencephalopathy with spheroids. Big amino acid change from small hydrophobic alanine to big special case proline
Deleterious	1	39.87123	frame-shift	*CSF1R*	0	p.Asp829Valfs	Not on *GnomAD*. Adjacent to a deleterious mutation and many more in the immediate vicinity in the tyrosine kinase domain where most deleterious mutations are located. Big amino acid change with charge change from positively charged to hydrophobic in the catalytic intracellular domain with an additional frameshift. On other chromosome from Glu694Lys variant in same patient, could be additive / recessive.
Likely Deleterious	1	41	missense	*CSF1R*	0	p.His776Tyr	Not previously reported, predicted damaging in silico but not strongly, not in *ExAC.* In mutation hotspot according to Guerreiro et al. 2013, adjacent mutations either similar or less pronounced than this big amino acid change going from charged to hydrophobic
Deleterious	1	-	frame-shift	*GRN*	0	c.522_523insTGTGAAGACAGGGTGCACTGCTGTC	Not an artefact. Frameshift mutation relatively early in the gene. Loss of function and haploinsufficiency known disease mechanism in GRN

Pathogenicity	Times observed	AAO	Type	Gene	Alleles / GnomAD	HGVSProtein (or coding if no change in protein)	Comment
Deleterious	1	55.61096	frame-shift	*GRN*	0	p.Asp254Valfs	This result confirms the diagnosis of a GRN-related dementia. This 2-bp duplication in granulin exon 8 causes a frameshift and a premature STOP, 3 codons downstream. Although this specific sequence change has not been previously reported, several pathogenic frameshift mutations in GRN have been previously described in the literature1.
Likely Deleterious	1	77.5	intronic	*GRN*	0	p.c.264+1G>A	mutation in a splice site, after input of exon , Human Splice Finder predict broken WT site and Alteration of the WT donor site, most probably affecting splicing.
Deleterious	1	58.89315	frame-shift	*GRN*	0	p.Cys260Valfs	This result confirms the diagnosis of a GRN-related dementia. This 1-bp duplication in granulin exon 8 causes a frameshift and a premature stop 14 codons downstream. Although this specific sequence change has not been previously reported, several pathogenic frameshift mutations in GRN have been previously described in the literature1.
Deleterious	1	67.14521	frame-shift	*GRN*	0	p.Cys482Ter	This result confirms the diagnosis of a GRN-related dementia. This substitution in granulin exon 12 causes a premature STOP codon. Although this specific sequence change has not been previously reported, several pathogenic nonsense and frameshift mutations in GRN causing similar protein effects have been described in the literature1.

Pathogenicity	Times observed	AAO	Type	Gene	Alleles / GnomAD	HGVSProtein (or coding if no change in protein)	Comment
Deleterious	1	60	frame-shift	*GRN*	0	p.Leu187Argfs	This result supports the diagnosis of a GRN-related dementia. This 1-bp deletion in granulin exon 6 causes a frameshift and a premature stop 69 codons downstream. Although this specific sequence change has not been previously reported, several pathogenic frameshift mutations in GRN have been previously described in the literature
Deleterious	1	72.4274	frame-shift	*GRN*	0	p.Met1	This result is consistent with a diagnosis of GRN-related dementia. The c.1179G>A p.? Variant has not been previously reported in the literature, though it is likely to be pathogenic. It occurs at a highly conserved splice site at an intron-exon boundary and is therefore predicted to affect GRN splicing. Analysis of affected family members would assist in the interpretation of this result.
Likely Deleterious	1	59	in-frame	*GRN*	0	p.Ser449_Thr455del	Not on *ExAC.*, not on google. Deletion slightly earlier in the gene has been described to cause young onset neurodegeneration, but insufficient information about exon 11. CADD phred 22.8
Deleterious	1	-	missense	*MAPT*	0	p.Gly303Ser	Gly303Val is called path on Molgen, 1 family. Described in a pedigree with segregation
Likely Deleterious	1	58.5	missense	*MAPT*	0	p.His362Tyr	Not on *ExAC.*, not on google. In silico prediction disagree, but big amino acid change with big charge change and next to a pathogenic mutation with a smaller amino acid change without charge change

Pathogenicity	Times observed	AAO	Type	Gene	Alleles / GnomAD	HGVSProtein (or coding if no change in protein)	Comment
Deleterious	1	56.5287	missens	*NOTCH3*	0	p.Arg640Cys	This result is consistent with a diagnosis of Cerebral Autosomal Dominant Arteriopathy with Subcortical Infarcts and Leukoencephalopathy (CADASIL). The c.1918C>T p. (Arg640Cys) variant has not been previously reported in the literature, though it is likely to be pathogenic. In keeping with many disease-causing mutations in NOTCH3, this substitution involves a change to a Cysteine amino acid within an EGF-like domain of the NOTCH3 protein. This variant has been reported to the Leiden Open Variation Database (LOVD) 1, having been detected in a patient with CADASIL. It is also recorded in the ExAC variation database2 at a global minor allele frequency of 0.0025%. In order to confirm a diagnosis of CADASIL, review of brain MRI for the characteristic white matter intensities is recommended.
		7	e				
Likely Deleterious	1	47	missens e	*PSEN 1*	0	p.Ala137Thr	Not on Molgen, but in mutation hotspot. Not on *ExAC.*, not on exome variant server. In silico predictions conflicting
Likely Deleterious	1	57.5	intronic	*PSEN 1*	0	c.869-1G>A	CADD score 26.7, not on *ExAC.*, intronic mutation. Human Splice Finder predicts a broken WT site and Alteration of the WT acceptor site, most probably affecting splicing.
Likely Deleterious	1	49	missens e	*PSEN 1*	0	p.Gln222Pro	Not on *ExAC.*. not on Molgen, but 2 mutations at same locus pathogenic (one change to Arg, one to His); these are both charged amino acid in a transmembrane domain and Proline is a special case, but with a very different sidechain from Gln

Pathogenicity	Times observed	AAO	Type	Gene	Alleles / GnomAD	HGVSProtein (or coding if no change in protein)	Comment
Likely Deleterious	1	38.5	in-frame	*PSEN 1*	0	p.Leu171Tyr	Not on *ExAC*. Similar mutation (L166H) reported pathogenic at same location, this variant has no charge change but a change in bulkiness.
Likely Deleterious	1	43.5	missense	*PSEN 1*	0	p.Pro433Ser	Not on *ExAC.*. Predicted deleterious. on the very edge of the intermembrane domain, next to several deleterious mutations, but only to one side, quite significant amino acid change to a much smaller one
Likely Deleterious	1	55	missense	*PSEN 1*	0	p.Thr122Ala	Not on *ExAC*. amino acid change from polar to hydrophobic in the 1st luminal part of PSEN1, close and adjacent to pathogenic mutations with no charge change
Deleterious	1	54.50685	missense	*PSEN 1*	0	p.Val142Ile	This result is consistent with a diagnosis of Alzheimer's disease in this patient. The PSEN1 c.424G>A p. (Val142Ile) variant has not been previously reported in the literature or to public databases of genetic variation (ExAC, EVS, 1000G), though it is likely to be pathogenic. It causes a missense change at a highly conserved amino acid located within the second transmembrane domain of the protein, a region in which several pathogenic missense mutations have been previously reported. Analysis of affected family members would assist in the interpretation of this result.
Likely Deleterious	1	64.5	missense	*PSEN 1*	0	p.Val393Phe	not on *ExAC.*, big amino acid change in size, deleterious mutations in both adjacent codons (with smaller amino acid changes), predicted deleterious in silico, CADD phredd 35

Pathogenicity	Times observed	AAO	Type	Gene	Alleles / GnomAD	HGVSProtein (or coding if no change in protein)	Comment
Likely Deleterious	1	30	missense	*PSEN 2*	0	p.Val150Met	Not in ExAC, known pathogenic mutation at 148 v148i (similar amino acid), reported in young onset Alzheimer patient with a family history in a thesis
Likely Deleterious	1	65	missense	*VCP*	0	p.Pro137Ser	Mutation at same location has been described in various pedigrees related to dementia with Paget's disease or with myopathy. The amino acid change caused by this variant compared to the published one is slightly less dramatic, but still very similar with a change from proline to an uncharged polar side chain instead of a hydrophobic one

References

1 Winblad, B. *et al.* Defeating Alzheimer's disease and other dementias: a priority for European science and society. *Lancet Neurol* **15**, 455-532, doi:10.1016/S1474-4422(16)00062-4 (2016).

2 Hou, Y. *et al.* Ageing as a risk factor for neurodegenerative disease. *Nature reviews. Neurology* **15**, 565-581, doi:10.1038/s41582-019-0244-7 (2019).

3 Jansen, I. E. *et al.* Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat Genet* **51**, 404-413, doi:10.1038/s41588-018-0311-9 (2019).

4 Gatz, M. *et al.* Role of genes and environments for explaining Alzheimer disease. *Arch Gen Psychiatry* **63**, 168-174, doi:10.1001/archpsyc.63.2.168 (2006).

5 Hardy, J. & Selkoe, D. J. The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science* **297**, 353-356, doi:10.1126/science.1072994 (2002).

6 Bateman, R. J. *et al.* Clinical and biomarker changes in dominantly inherited Alzheimer's disease. *N Engl J Med* **367**, 795-804, doi:10.1056/NEJMoa1202753 (2012).

7 Bateman, R. J. *et al.* The DIAN-TU Next Generation Alzheimer's prevention trial: Adaptive design and disease progression model. *Alzheimers Dement* **13**, 8-19, doi:10.1016/j.jalz.2016.07.005 (2017).

8 Alector. (GlobeNewsWire, 2019).

9 Tabrizi, S. J., Ghosh, R. & Leavitt, B. R. Huntingtin Lowering Strategies for Disease Modification in Huntington's Disease. *Neuron* **101**, 801-819, doi:10.1016/j.neuron.2019.01.039 (2019).

10 Atkins, E. R. & Panegyres, P. K. The clinical utility of gene testing for Alzheimer's disease. *Neurol Int* **3**, e1, doi:10.4081/ni.2011.e1 (2011).

11 Falk, N., Cole, A. & Meredith, T. J. Evaluation of Suspected Dementia. *Am Fam Physician* **97**, 398-405 (2018).

12 Irwin, D. J. *et al.* Frontotemporal lobar degeneration: defining phenotypic diversity through personalized medicine. *Acta Neuropathol* **129**, 469-491, doi:10.1007/s00401-014-1380-1 (2015).

13 Hensman Moss, D. J. *et al.* C9orf72 expansions are the most common genetic cause of Huntington disease phenocopies. *Neurology* **82**, 292-299, doi:10.1212/WNL.0000000000000061 (2014).

14 Rohrer, J. D. *et al.* C9orf72 expansions in frontotemporal dementia and amyotrophic lateral sclerosis. *Lancet Neurol* **14**, 291-301, doi:10.1016/S1474-4422(14)70233-9 (2015).

15 Collaborators, G. B. D. D. Global, regional, and national burden of Alzheimer's disease and other dementias, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet. Neurology* **18**, 88-106, doi:10.1016/S1474-4422(18)30403-4 (2019).

16 Simpson, J. R. DSM-5 and neurocognitive disorders. *The journal of the American Academy of Psychiatry and the Law* **42**, 159-164 (2014).

17 Sachdev, P. S. *et al.* Classifying neurocognitive disorders: the DSM-5 approach. *Nature reviews. Neurology* **10**, 634-642, doi:10.1038/nrneurol.2014.181 (2014).
18 Rossor, M. N., Fox, N. C., Mummery, C. J., Schott, J. M. & Warren, J. D. The diagnosis of young-onset dementia. *The Lancet. Neurology* **9**, 793-806, doi:10.1016/S1474-4422(10)70159-9 (2010).
19 Wolters, F. J. *et al.* Parental family history of dementia in relation to subclinical brain disease and dementia risk. *Neurology*, doi:10.1212/WNL.0000000000003871 (2017).
20 Rohrer, J. D. *et al.* The heritability and genetics of frontotemporal lobar degeneration. *Neurology* **73**, 1451-1456, doi:10.1212/WNL.0b013e3181bf997a (2009).
21 Ngo, J. & Holroyd-Leduc, J. M. Systematic review of recent dementia practice guidelines. *Age Ageing* **44**, 25-33, doi:10.1093/ageing/afu143 (2015).
22 Mitsui, J. & Tsuji, S. Genomic aspects of sporadic neurodegenerative diseases. *Biochemical and biophysical research communications* **452**, 221-225, doi:10.1016/j.bbrc.2014.07.098 (2014).
23 Goldman, J. S. *et al.* An algorithm for genetic testing of frontotemporal lobar degeneration. *Neurology* **76**, 475-483, doi:10.1212/WNL.0b013e31820a0d13 (2011).
24 Koriath, C. *et al.* Predictors for a dementia gene mutation based on gene-panel next-generation sequencing of a large dementia referral series. *Molecular psychiatry*, doi:10.1038/s41380-018-0224-0 (2018).
25 Goldman, J. S. Genetic testing and counseling in the diagnosis and management of young-onset dementias. *The Psychiatric clinics of North America* **38**, 295-308, doi:10.1016/j.psc.2015.01.008 (2015).
26 Bird, T. D. & Smith, C. O. Clinical approach to the patient with neurogenetic disease. *Handbook of clinical neurology* **147**, 3-9, doi:10.1016/B978-0-444-63233-3.00001-4 (2018).
27 Deleon, J. & Miller, B. L. Frontotemporal dementia. *Handbook of clinical neurology* **148**, 409-430, doi:10.1016/B978-0-444-64076-5.00027-2 (2018).
28 Minikel, E. V. *et al.* Quantifying prion disease penetrance using large population control cohorts. *Science translational medicine* **8**, 322ra329, doi:10.1126/scitranslmed.aad5169 (2016).
29 Harms, M. *et al.* C9orf72 hexanucleotide repeat expansions in clinical Alzheimer disease. *JAMA neurology* **70**, 736-741, doi:10.1001/2013.jamaneurol.537 (2013).
30 Lesage, S. *et al.* C9orf72 repeat expansions are a rare genetic cause of parkinsonism. *Brain : a journal of neurology* **136**, 385-391, doi:10.1093/brain/aws357 (2013).
31 Giri, M., Zhang, M. & Lu, Y. Genes associated with Alzheimer's disease: an overview and current status. *Clinical interventions in aging* **11**, 665-681, doi:10.2147/CIA.S105769 (2016).
32 Escott-Price, V. *et al.* Common polygenic variation enhances risk prediction for Alzheimer's disease. *Brain : a journal of neurology* **138**, 3673-3684, doi:10.1093/brain/awv268 (2015).

33 Cacace, R., Sleegers, K. & Van Broeckhoven, C. Molecular genetics of early-onset Alzheimer's disease revisited. *Alzheimer's & dementia : the journal of the Alzheimer's Association* **12**, 733-748, doi:10.1016/j.jalz.2016.01.012 (2016).
34 Rovelet-Lecrux, A. *et al.* APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. *Nature genetics* **38**, 24-26, doi:10.1038/ng1718 (2006).
35 Guven, G. *et al.* Peripheral GRN mRNA and Serum Progranulin Levels as a Potential Indicator for Both the Presence of Splice Site Mutations and Individuals at Risk for Frontotemporal Dementia. *J Alzheimers Dis* **67**, 159-167, doi:10.3233/JAD-180599 (2019).
36 Sposito, T. *et al.* Developmental regulation of tau splicing is disrupted in stem cell-derived neurons from frontotemporal dementia patients with the 10 + 16 splice-site mutation in MAPT. *Human molecular genetics* **24**, 5260-5269, doi:10.1093/hmg/ddv246 (2015).
37 Neueder, A. *et al.* The pathogenic exon 1 HTT protein is produced by incomplete splicing in Huntington's disease patients. *Sci Rep* **7**, 1307, doi:10.1038/s41598-017-01510-z (2017).
38 Bates, G. P. *et al.* Huntington disease. *Nat Rev Dis Primers* **1**, 15005, doi:10.1038/nrdp.2015.5 (2015).
39 Balendra, R. & Isaacs, A. M. C9orf72-mediated ALS and FTD: multiple pathways to disease. *Nature reviews. Neurology* **14**, 544-558, doi:10.1038/s41582-018-0047-2 (2018).
40 Santa-Maria, I. *et al.* The MAPT H1 haplotype is associated with tangle-predominant dementia. *Acta Neuropathol* **124**, 693-704, doi:10.1007/s00401-012-1017-1 (2012).
41 Forrest, S. L., Kril, J. J. & Halliday, G. M. Cellular and regional vulnerability in frontotemporal tauopathies. *Acta Neuropathol* **138**, 705-727, doi:10.1007/s00401-019-02035-7 (2019).
42 Archer, H. A. *et al.* Memory complaints and increased rates of brain atrophy: risk factors for mild cognitive impairment and Alzheimer's disease. *International journal of geriatric psychiatry* **25**, 1119-1126, doi:10.1002/gps.2440 (2010).
43 Barnes, J. *et al.* Differentiating AD from aging using semiautomated measurement of hippocampal atrophy rates. *NeuroImage* **23**, 574-581, doi:10.1016/j.neuroimage.2004.06.028 (2004).
44 Fox, N. C. *et al.* Presymptomatic hippocampal atrophy in Alzheimer's disease. A longitudinal MRI study. *Brain : a journal of neurology* **119 (Pt 6)**, 2001-2007 (1996).
45 Lambert, M. A. *et al.* Estimating the burden of early onset dementia; systematic review of disease prevalence. *European journal of neurology* **21**, 563-569, doi:10.1111/ene.12325 (2014).
46 Dujardin, S. *et al.* Tau molecular diversity contributes to clinical heterogeneity in Alzheimer's disease. *Nature medicine*, doi:10.1038/s41591-020-0938-9 (2020).
47 Prince, M. W., A., Guerchet, M., Ali, GC., Wu, YT, Prina, M., Alzheimer's Disease International. Alzheimer's Disease International: World Alzheimer Report 2015. (Alzheimer's Disease International (ADI), London, 2015).

48 Schindler, S. E. & Fagan, A. M. Autosomal Dominant Alzheimer Disease: A Unique Resource to Study CSF Biomarker Changes in Preclinical AD. *Front Neurol* **6**, 142, doi:10.3389/fneur.2015.00142 (2015).
49 van Vliet, D. *et al.* Time to diagnosis in young-onset dementia as compared with late-onset dementia. *Psychological medicine* **43**, 423-432, doi:10.1017/S0033291712001122 (2013).
50 Jones, L. *et al.* Genetic evidence implicates the immune system and cholesterol metabolism in the aetiology of Alzheimer's disease. *PloS one* **5**, e13950, doi:10.1371/journal.pone.0013950 (2010).
51 Guerreiro, R. *et al.* TREM2 variants in Alzheimer's disease. *The New England journal of medicine* **368**, 117-127, doi:10.1056/NEJMoa1211851 (2013).
52 Ryman, D. C. *et al.* Symptom onset in autosomal dominant Alzheimer disease: a systematic review and meta-analysis. *Neurology* **83**, 253-260, doi:10.1212/WNL.0000000000000596 (2014).
53 Fox, N. C. *et al.* Clinicopathological features of familial Alzheimer's disease associated with the M139V mutation in the presenilin 1 gene. Pedigree but not mutation specific age at onset provides evidence for a further genetic factor. *Brain* **120 (Pt 3)**, 491-501, doi:10.1093/brain/120.3.491 (1997).
54 Arboleda-Velasquez, J. F. *et al.* Resistance to autosomal dominant Alzheimer's disease in an APOE3 Christchurch homozygote: a case report. *Nature medicine* **25**, 1680-1683, doi:10.1038/s41591-019-0611-3 (2019).
55 Thordardottir, S. *et al.* Reduced penetrance of the PSEN1 H163Y autosomal dominant Alzheimer mutation: a 22-year follow-up study. *Alzheimers Res Ther* **10**, 45, doi:10.1186/s13195-018-0374-y (2018).
56 Rossor, M. N., Fox, N. C., Beck, J., Campbell, T. C. & Collinge, J. Incomplete penetrance of familial Alzheimer's disease in a pedigree with a novel presenilin-1 gene mutation. *Lancet* **347**, 1560, doi:10.1016/s0140-6736(96)90715-1 (1996).
57 Liang, Y. *et al.* A cognitive chameleon: lessons from a novel MAPT mutation case. *Neurocase* **20**, 684-694, doi:10.1080/13554794.2013.826697 (2014).
58 Mead, S., Lloyd, S. & Collinge, J. Genetic Factors in Mammalian Prion Diseases. *Annu Rev Genet* **53**, 117-147, doi:10.1146/annurev-genet-120213-092352 (2019).
59 Zarea, A. *et al.* Seizures in dominantly inherited Alzheimer disease. *Neurology* **87**, 912-919, doi:10.1212/WNL.0000000000003048 (2016).
60 Ryan, N. S. *et al.* Clinical phenotype and genetic associations in autosomal dominant familial Alzheimer's disease: a case series. *Lancet Neurol* **15**, 1326-1335, doi:10.1016/S1474-4422(16)30193-4 (2016).
61 Bernardi, L. *et al.* Novel PSEN1 and PGRN mutations in early-onset familial frontotemporal dementia. *Neurobiology of aging* **30**, 1825-1833, doi:10.1016/j.neurobiolaging.2008.01.005 (2009).
62 Ramos, E. M. *et al.* Genetic screening of a large series of North American sporadic and familial frontotemporal dementia cases. *Alzheimers Dement* **16**, 118-130, doi:10.1002/alz.12011 (2020).
63 Wallon, D. *et al.* The French series of autosomal dominant early onset Alzheimer's disease cases: mutation spectrum and cerebrospinal fluid

biomarkers. *Journal of Alzheimer's disease : JAD* **30**, 847-856, doi:10.3233/JAD-2012-120172 (2012).

64 Canevelli, M. *et al.* Familial Alzheimer's disease sustained by presenilin 2 mutations: systematic review of literature and genotype-phenotype correlation. *Neurosci Biobehav Rev* **42**, 170-179, doi:10.1016/j.neubiorev.2014.02.010 (2014).

65 Wolf, A. B. *et al.* Apolipoprotein E as a beta-amyloid-independent factor in Alzheimer's disease. *Alzheimers Res Ther* **5**, 38, doi:10.1186/alzrt204 (2013).

66 Belloy, M. E., Napolioni, V. & Greicius, M. D. A Quarter Century of APOE and Alzheimer's Disease: Progress to Date and the Path Forward. *Neuron* **101**, 820-838, doi:10.1016/j.neuron.2019.01.056 (2019).

67 Uddin, M. S. *et al.* APOE and Alzheimer's Disease: Evidence Mounts that Targeting APOE4 may Combat Alzheimer's Pathogenesis. *Mol Neurobiol* **56**, 2450-2465, doi:10.1007/s12035-018-1237-z (2019).

68 Bertram, L., McQueen, M. B., Mullin, K., Blacker, D. & Tanzi, R. E. Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. *Nat Genet* **39**, 17-23, doi:10.1038/ng1934 (2007).

69 Maestre, G. *et al.* Apolipoprotein E and Alzheimer's disease: ethnic variation in genotypic risks. *Annals of neurology* **37**, 254-259, doi:10.1002/ana.410370217 (1995).

70 Kawamata, J., Tanaka, S., Shimohama, S., Ueda, K. & Kimura, J. Apolipoprotein E polymorphism in Japanese patients with Alzheimer's disease or vascular dementia. *Journal of neurology, neurosurgery, and psychiatry* **57**, 1414-1416, doi:10.1136/jnnp.57.11.1414 (1994).

71 Reiman, E. M. *et al.* Exceptionally low likelihood of Alzheimer's dementia in APOE2 homozygotes from a 5,000-person neuropathological study. *Nature communications* **11**, 667, doi:10.1038/s41467-019-14279-8 (2020).

72 Onyike, C. U. & Diehl-Schmid, J. The epidemiology of frontotemporal dementia. *International review of psychiatry* **25**, 130-137, doi:10.3109/09540261.2013.776523 (2013).

73 Coyle-Gilchrist, I. T. *et al.* Prevalence, characteristics, and survival of frontotemporal lobar degeneration syndromes. *Neurology* **86**, 1736-1743, doi:10.1212/WNL.0000000000002638 (2016).

74 Rohrer, J. D. *et al.* Clinical and neuroanatomical signatures of tissue pathology in frontotemporal lobar degeneration. *Brain : a journal of neurology* **134**, 2565-2581, doi:10.1093/brain/awr198 (2011).

75 Poos, J. M. *et al.* Cognitive profiles discriminate between genetic variants of behavioral frontotemporal dementia. *J Neurol*, doi:10.1007/s00415-020-09738-y (2020).

76 Moore, K. M. *et al.* Age at symptom onset and death and disease duration in genetic frontotemporal dementia: an international retrospective cohort study. *The Lancet. Neurology*, doi:10.1016/S1474-4422(19)30394-1 (2019).

77 Mahoney, C. J. *et al.* Frontotemporal dementia with the C9ORF72 hexanucleotide repeat expansion: clinical, neuroanatomical and neuropathological features. *Brain : a journal of neurology* **135**, 736-750, doi:10.1093/brain/awr361 (2012).

78 Goldman, J. S. *et al.* Comparison of family histories in FTLD subtypes and related tauopathies. *Neurology* **65**, 1817-1819, doi:10.1212/01.wnl.0000187068.92184.63 (2005).
79 Po, K. *et al.* Heritability in frontotemporal dementia: more missing pieces? *J Neurol* **261**, 2170-2177, doi:10.1007/s00415-014-7474-9 (2014).
80 Wood, E. M. *et al.* Development and validation of pedigree classification criteria for frontotemporal lobar degeneration. *JAMA neurology* **70**, 1411-1417, doi:10.1001/jamaneurol.2013.3956 (2013).
81 Hutton, M. *et al.* Association of missense and 5'-splice-site mutations in tau with the inherited dementia FTDP-17. *Nature* **393**, 702-705, doi:10.1038/31508 (1998).
82 Cruts, M. *et al.* Null mutations in progranulin cause ubiquitin-positive frontotemporal dementia linked to chromosome 17q21. *Nature* **442**, 920-924, doi:10.1038/nature05017 (2006).
83 Greaves, C. V. & Rohrer, J. D. An update on genetic frontotemporal dementia. *J Neurol* **266**, 2075-2086, doi:10.1007/s00415-019-09363-4 (2019).
84 Ferrari, R. *et al.* Frontotemporal dementia: insights into the biological underpinnings of disease through gene co-expression network analysis. *Mol Neurodegener* **11**, 21, doi:10.1186/s13024-016-0085-4 (2016).
85 Lashley, T. *et al.* A pathogenic progranulin mutation and C9orf72 repeat expansion in a family with frontotemporal dementia. *Neuropathol Appl Neurobiol* **40**, 502-513, doi:10.1111/nan.12100 (2014).
86 Thompson, A. *et al.* Behavioral and psychiatric symptoms in prion disease. *The American journal of psychiatry* **171**, 265-274, doi:10.1176/appi.ajp.2013.12111460 (2014).
87 Armstrong, R. A. Factors determining disease duration in Alzheimer's disease: a postmortem study of 103 cases using the Kaplan-Meier estimator and Cox regression. *BioMed research international* **2014**, 623487, doi:10.1155/2014/623487 (2014).
88 Rudge, P., Hyare, H., Green, A., Collinge, J. & Mead, S. Imaging and CSF analyses effectively distinguish CJD from its mimics. *Journal of neurology, neurosurgery, and psychiatry* **89**, 461-466, doi:10.1136/jnnp-2017-316853 (2018).
89 Mead, S. & Rudge, P. CJD mimics and chameleons. *Pract Neurol* **17**, 113-121, doi:10.1136/practneurol-2016-001571 (2017).
90 Lloyd, S. E., Mead, S. & Collinge, J. Genetics of prion diseases. *Current opinion in genetics & development* **23**, 345-351, doi:10.1016/j.gde.2013.02.012 (2013).
91 Jaunmuktane, Z. *et al.* Evidence of amyloid-beta cerebral amyloid angiopathy transmission through neurosurgery. *Acta Neuropathol* **135**, 671-679, doi:10.1007/s00401-018-1822-2 (2018).
92 Mead, S. Prion disease genetics. *European journal of human genetics : EJHG* **14**, 273-281, doi:10.1038/sj.ejhg.5201544 (2006).
93 Mead, S. *et al.* Inherited prion disease with 5-OPRI: phenotype modification by repeat length and codon 129. *Neurology* **69**, 730-738, doi:10.1212/01.wnl.0000267642.41594.9d (2007).
94 Rawlins, M. Huntington's disease out of the closet? *Lancet* **376**, 1372-1373, doi:10.1016/S0140-6736(10)60974-9 (2010).

95 Ghosh, R. & Tabrizi, S. J. Clinical Features of Huntington's Disease. *Adv Exp Med Biol* **1049**, 1-28, doi:10.1007/978-3-319-71779-1_1 (2018).
96 Ghosh, R. & Tabrizi, S. J. Huntington disease. *Handbook of clinical neurology* **147**, 255-278, doi:10.1016/B978-0-444-63233-3.00017-8 (2018).
97 Gusella, J. F. *et al.* A polymorphic DNA marker genetically linked to Huntington's disease. *Nature* **306**, 234-238, doi:10.1038/306234a0 (1983).
98 A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. The Huntington's Disease Collaborative Research Group. *Cell* **72**, 971-983, doi:10.1016/0092-8674(93)90585-e (1993).
99 Rubinsztein, D. C. *et al.* Phenotypic characterization of individuals with 30-40 CAG repeats in the Huntington disease (HD) gene reveals HD cases with 36 repeats and apparently normal elderly individuals with 36-39 repeats. *Am J Hum Genet* **59**, 16-22 (1996).
100 Bettencourt, C. *et al.* DNA repair pathways underlie a common genetic mechanism modulating onset in polyglutamine diseases. *Annals of neurology* **79**, 983-990, doi:10.1002/ana.24656 (2016).
101 Becanovic, K. *et al.* A SNP in the HTT promoter alters NF-kappaB binding and is a bidirectional genetic modifier of Huntington disease. *Nature neuroscience* **18**, 807-816, doi:10.1038/nn.4014 (2015).
102 Andrew, S. E. *et al.* Huntington disease without CAG expansion: phenocopies or errors in assignment? *American journal of human genetics* **54**, 852-863 (1994).
103 Wild, E. J. *et al.* Huntington's disease phenocopies are clinically and genetically heterogeneous. *Movement disorders : official journal of the Movement Disorder Society* **23**, 716-720, doi:10.1002/mds.21915 (2008).
104 Mignarri, A. *et al.* Double trouble? Progranulin mutation and C9ORF72 repeat expansion in a case of primary non-fluent aphasia. *Journal of the neurological sciences* **341**, 176-178, doi:10.1016/j.jns.2014.03.030 (2014).
105 Testi, S., Tamburin, S., Zanette, G. & Fabrizi, G. M. Co-occurrence of the C9ORF72 expansion and a novel GRN mutation in a family with alternative expression of frontotemporal dementia and amyotrophic lateral sclerosis. *J Alzheimers Dis* **44**, 49-56, doi:10.3233/JAD-141794 (2015).
106 Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 5463-5467, doi:10.1073/pnas.74.12.5463 (1977).
107 *Human Genome Project*, <https://www.genome.gov/human-genome-project/Completion-FAQ> (
108 Venter, J. C. *et al.* The sequence of the human genome. *Science* **291**, 1304-1351, doi:10.1126/science.1058040 (2001).
109 Schouten, J. P. *et al.* Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic acids research* **30**, e57, doi:10.1093/nar/gnf056 (2002).
110 Method of the year. *Nature methods* **5**, 1 (2008).
111 Beck, J. *et al.* Large C9orf72 hexanucleotide repeat expansions are seen in multiple neurodegenerative syndromes and are more frequent than expected in the UK population. *American journal of human genetics* **92**, 345-353, doi:10.1016/j.ajhg.2013.01.011 (2013).

112 Nho, K. *et al.* Whole-exome sequencing and imaging genetics identify functional variants for rate of change in hippocampal volume in mild cognitive impairment. *Molecular psychiatry* **18**, 781-787, doi:10.1038/mp.2013.24 (2013).
113 Beck, J. *et al.* Validation of next-generation sequencing technologies in genetic diagnosis of dementia. *Neurobiology of aging* **35**, 261-265, doi:10.1016/j.neurobiolaging.2013.07.017 (2014).
114 Pottier, C. *et al.* Genome-wide analyses as part of the international FTLD-TDP whole-genome sequencing consortium reveals novel disease risk factors and increases support for immune dysfunction in FTLD. *Acta Neuropathol* **137**, 879-899, doi:10.1007/s00401-019-01962-9 (2019).
115 DeJesus-Hernandez, M. *et al.* Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* **72**, 245-256, doi:10.1016/j.neuron.2011.09.011 (2011).
116 Blauwendraat, C. *et al.* The wide genetic landscape of clinical frontotemporal dementia: systematic combined sequencing of 121 consecutive subjects. *Genetics in medicine : official journal of the American College of Medical Genetics* **20**, 240-249, doi:10.1038/gim.2017.102 (2018).
117 Björn N., P. S., Sigurgeirsson B., Lundeberg J., Gréen H., Sahlén P. Comparison of Variant Calls from Whole Genome and Whole Exome Sequencing Data Using Matched Samples. *Journal of Next Generation Sequencing & Applications* **5**, doi:10.4172/2469-9853.1000154 (2018).
118 Belkadi, A. *et al.* Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 5473-5478, doi:10.1073/pnas.1418631112 (2015).
119 Fogel, B. L. *et al.* Exome sequencing in the clinical diagnosis of sporadic or familial cerebellar ataxia. *JAMA neurology* **71**, 1237-1246, doi:10.1001/jamaneurol.2014.1944 (2014).
120 Xu, Y. *et al.* The Whole Exome Sequencing Clarifies the Genotype- Phenotype Correlations in Patients with Early-Onset Dementia. *Aging Dis* **9**, 696-705, doi:10.14336/AD.2018.0208 (2018).
121 Galatolo, D., Tessa, A., Filla, A. & Santorelli, F. M. Clinical application of next generation sequencing in hereditary spinocerebellar ataxia: increasing the diagnostic yield and broadening the ataxia-spasticity spectrum. A retrospective analysis. *Neurogenetics* **19**, 1-8, doi:10.1007/s10048-017-0532-6 (2018).
122 Fogel, B. L. Genetic and genomic testing for neurologic disease in clinical practice. *Handbook of clinical neurology* **147**, 11-22, doi:10.1016/B978-0-444-63233-3.00002-6 (2018).
123 Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434-443, doi:10.1038/s41586-020-2308-7 (2020).
124 Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285-291, doi:10.1038/nature19057 (2016).
125 Ku, C. S. *et al.* Exome sequencing: Dual role as a discovery and diagnostic tool. *Annals of neurology* **71**, 5-14, doi:10.1002/ana.22647 (2012).

126 Rabbani, B., Tekin, M. & Mahdieh, N. The promise of whole-exome sequencing in medical genetics. *Journal of human genetics* **59**, 5-15, doi:10.1038/jhg.2013.114 (2014).
127 Cuccaro, D., De Marco, E. V., Cittadella, R. & Cavallaro, S. Copy Number Variants in Alzheimer's Disease. *J Alzheimers Dis* **55**, 37-52, doi:10.3233/JAD-160469 (2017).
128 Ebbert, M. T. W. *et al.* Long-read sequencing across the C9orf72 'GGGGCC' repeat expansion: implications for clinical use and genetic discovery efforts in human disease. *Mol Neurodegener* **13**, 46, doi:10.1186/s13024-018-0274-4 (2018).
129 Kumar, K. R., Cowley, M. J. & Davis, R. L. Next-Generation Sequencing and Emerging Technologies. *Semin Thromb Hemost* **45**, 661-673, doi:10.1055/s-0039-1688446 (2019).
130 Dolzhenko, E. *et al.* Detection of long repeat expansions from PCR-free whole-genome sequence data. *Genome research* **27**, 1895-1903, doi:10.1101/gr.225672.117 (2017).
131 Renton, A. E. *et al.* A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* **72**, 257-268, doi:10.1016/j.neuron.2011.09.010 (2011).
132 Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in medicine : official journal of the American College of Medical Genetics* **17**, 405-424, doi:10.1038/gim.2015.30 (2015).
133 Amendola, L. M. *et al.* Performance of ACMG-AMP Variant-Interpretation Guidelines among Nine Laboratories in the Clinical Sequencing Exploratory Research Consortium. *American journal of human genetics* **98**, 1067-1076, doi:10.1016/j.ajhg.2016.03.024 (2016).
134 Pepin, M. G. *et al.* The challenge of comprehensive and consistent sequence variant interpretation between clinical laboratories. *Genetics in medicine : official journal of the American College of Medical Genetics* **18**, 20-24, doi:10.1038/gim.2015.31 (2016).
135 Yang, Y. *et al.* Molecular findings among patients referred for clinical whole-exome sequencing. *Jama* **312**, 1870-1879, doi:10.1001/jama.2014.14601 (2014).
136 Yang, Y. *et al.* Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *The New England journal of medicine* **369**, 1502-1511, doi:10.1056/NEJMoa1306555 (2013).
137 Shahmirzadi, L. *et al.* Patient decisions for disclosure of secondary findings among the first 200 individuals undergoing clinical diagnostic exome sequencing. *Genetics in medicine : official journal of the American College of Medical Genetics* **16**, 395-399, doi:10.1038/gim.2013.153 (2014).
138 Hicks, J. K. *et al.* Patient Decisions to Receive Secondary Pharmacogenomic Findings and Development of a Multidisciplinary Practice Model to Integrate Results Into Patient Care. *Clin Transl Sci* **11**, 71-76, doi:10.1111/cts.12493 (2018).

139 Daack-Hirsch, S. *et al.* 'Information is information': a public perspective on incidental findings in clinical and research genome-based testing. *Clinical genetics* **84**, 11-18, doi:10.1111/cge.12167 (2013).
140 Kalia, S. S. *et al.* Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. *Genetics in medicine : official journal of the American College of Medical Genetics* **19**, 249-255, doi:10.1038/gim.2016.190 (2017).
141 in *MRS JUSTICE YIP DBE England and Wales High Court (Queen's Bench Division) Decisions* (2020).
142 Florida. Supreme, C. Pate v. Threlkel. *Wests South Report* **661**, 278-282 (1995).
143 Hodgson, J. *et al.* Outcomes of a randomised controlled trial of a complex genetic counselling intervention to improve family communication. *European journal of human genetics : EJHG* **24**, 356-360, doi:10.1038/ejhg.2015.122 (2016).
144 Wiens, M. E., Wilson, B. J., Honeywell, C. & Etchegary, H. A family genetic risk communication framework: guiding tool development in genetics health services. *J Community Genet* **4**, 233-242, doi:10.1007/s12687-012-0134-9 (2013).
145 Owen, J. *et al.* Predictive testing for inherited prion disease: report of 22 years experience. *European journal of human genetics : EJHG* **22**, 1351-1356, doi:10.1038/ejhg.2014.42 (2014).
146 Ji, A. L., Zhang, X., Chen, W. W. & Huang, W. J. Genetics insight into the amyotrophic lateral sclerosis/frontotemporal dementia spectrum. *Journal of medical genetics* **54**, 145-154, doi:10.1136/jmedgenet-2016-104271 (2017).
147 Shi, Y. *et al.* ApoE4 markedly exacerbates tau-mediated neurodegeneration in a mouse model of tauopathy. *Nature* **549**, 523-527, doi:10.1038/nature24016 (2017).
148 Strittmatter, W. J. *et al.* Apolipoprotein E: high-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 1977-1981, doi:10.1073/pnas.90.5.1977 (1993).
149 Corder, E. H. *et al.* Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* **261**, 921-923, doi:10.1126/science.8346443 (1993).
150 Rohrer, J. D. *et al.* The heritability and genetics of frontotemporal lobar degeneration. *Neurology* **73**, 1451-1456 (2009).
151 Beck, J. *et al.* A distinct clinical, neuropsychological and radiological phenotype is associated with progranulin gene mutations in a large UK series. *Brain : a journal of neurology* **131**, 706-720, doi:10.1093/brain/awm320 (2008).
152 Genomes Project, C. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74, doi:10.1038/nature15393 (2015).
153 Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-575, doi:10.1086/519795 (2007).

154 Freischmidt, A. *et al.* Haploinsufficiency of TBK1 causes familial ALS and fronto-temporal dementia. *Nature neuroscience* **18**, 631-636, doi:10.1038/nn.4000 (2015).
155 Wadsworth, J. D. *et al.* Molecular diagnosis of human prion disease. *Methods Mol Biol* **459**, 197-227, doi:10.1007/978-1-59745-234-2_14 (2008).
156 Warner, J. P. *et al.* A general method for the detection of large CAG repeat expansions by fluorescent PCR. *Journal of medical genetics* **33**, 1022-1026 (1996).
157 Tsai, M. F. *et al.* PrimerZ: streamlined primer design for promoters, exons and human SNPs. *Nucleic acids research* **35**, W63-65, doi:10.1093/nar/gkm383 (2007).
158 Untergasser, A. *et al.* Primer3--new capabilities and interfaces. *Nucleic acids research* **40**, e115, doi:10.1093/nar/gks596 (2012).
159 Yates, A. *et al.* Ensembl 2016. *Nucleic acids research* **44**, D710-716, doi:10.1093/nar/gkv1157 (2016).
160 *UCSC genome browser* < http://genome.ucsc.edu/> (2017).
161 Kent, W. J. *et al.* The human genome browser at UCSC. *Genome research* **12**, 996-1006, doi:10.1101/gr.229102. Article published online before print in May 2002 (2002).
162 McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* **17**, 122, doi:10.1186/s13059-016-0974-4 (2016).
163 Desmet, F. O. *et al.* Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic acids research* **37**, e67, doi:10.1093/nar/gkp215 (2009).
164 Pubmed.
165 Wilson, E. B. Probable Inference, the Law of Succession, and Statistical Inference. *Journal of the American Statistical Association* **Vol. 22,**, 209-212 (1927).
166 Bates, G., Tabrizi, S., Jones, L. *Huntington's disease*. 4th edn, 512 Pages (Oxford Press, 2014).
167 Kremer, B. *et al.* A worldwide study of the Huntington's disease mutation. The sensitivity and specificity of measuring CAG repeats. *The New England journal of medicine* **330**, 1401-1406, doi:10.1056/NEJM199405193302001 (1994).
168 Koriath, C., Wild, E., Tabrizi, S. *Huntington's Disease and HD-like syndromes*, <https://www.surveymonkey.co.uk/r/TL32BB7> (
169 Bioinformatics, B. C. <http://bcb.io/> (
170 Faust, G. G. & Hall, I. M. SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics* **30**, 2503-2505, doi:10.1093/bioinformatics/btu314 (2014).
171 McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research* **20**, 1297-1303, doi:10.1101/gr.107524.110 (2010).
172 Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nature methods* **7**, 248-249, doi:10.1038/nmeth0410-248 (2010).
173 Ng, P. C. & Henikoff, S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic acids research* **31**, 3812-3814 (2003).

174 Lee, W., Zhang, Y., Mukhyala, K., Lazarus, R. A. & Zhang, Z. Bi-directional SIFT predicts a subset of activating mutations. *PloS one* **4**, e8311, doi:10.1371/journal.pone.0008311 (2009).
175 Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic acids research* **47**, D886-D894, doi:10.1093/nar/gky1016 (2019).
176 Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome research* **20**, 110-121, doi:10.1101/gr.097857.109 (2010).
177 Yeo, G. & Burge, C. B. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *Journal of computational biology : a journal of computational molecular cell biology* **11**, 377-394, doi:10.1089/1066527041410418 (2004).
178 Eng, L. *et al.* Nonclassical splicing mutations in the coding and noncoding regions of the ATM Gene: maximum entropy estimates of splice junction strengths. *Human mutation* **23**, 67-76, doi:10.1002/humu.10295 (2004).
179 Consortium, E. P. A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS biology* **9**, e1001046, doi:10.1371/journal.pbio.1001046 (2011).
180 Petrovski, S., Wang, Q., Heinzen, E. L., Allen, A. S. & Goldstein, D. B. Genic intolerance to functional variation and the interpretation of personal genomes. *PLoS genetics* **9**, e1003709, doi:10.1371/journal.pgen.1003709 (2013).
181 Project, G. <https://gtexportal.org/> (
182 Skoglund, L. *et al.* Novel progranulin mutation detected in 2 patients with FTLD. *Alzheimer disease and associated disorders* **25**, 173-178, doi:10.1097/WAD.0b013e3181fbc22c (2011).
183 Koriath, C. A. *et al.* The clinical, neuroanatomical, and neuropathologic phenotype of TBK1-associated frontotemporal dementia: A longitudinal case report. *Alzheimers Dement (Amst)* **6**, 75-81, doi:10.1016/j.dadm.2016.10.003 (2017).
184 Baets, J. *et al.* Defects of mutant DNMT1 are linked to a spectrum of neurological disorders. *Brain* **138**, 845-861, doi:10.1093/brain/awv010 (2015).
185 van Blitterswijk, M. *et al.* Evidence for an oligogenic basis of amyotrophic lateral sclerosis. *Human molecular genetics* **21**, 3776-3784, doi:10.1093/hmg/dds199 (2012).
186 Ahmed, Z. *et al.* A Huntington's disease phenocopy characterized by pallido-nigro-luysian degeneration with brain iron accumulation and p62-positive glial inclusions. *Neuropathol Appl Neurobiol* **36**, 551-557, doi:10.1111/j.1365-2990.2010.01093.x (2010).
187 Fowler A, M. S., Ruark E et al. . Accurate clinical detection of exon copy number variants in a targeted NGS panel using DECoN [version 1; referees: 1 approved]. *Wellcome Open Res* **2016, 1:20**, doi:doi: 10.12688/wellcomeopenres.10069.1 (2016).
188 Luukkainen, L., Bloigu, R., Moilanen, V. & Remes, A. M. Epidemiology of Frontotemporal Lobar Degeneration in Northern Finland. *Dementia and geriatric cognitive disorders extra* **5**, 435-441, doi:10.1159/000440858 (2015).
189 Reiman, E. M. *et al.* Fibrillar amyloid-beta burden in cognitively normal people at 3 levels of genetic risk for Alzheimer's disease. *Proceedings of the National*

Academy of Sciences of the United States of America **106**, 6820-6825, doi:10.1073/pnas.0900345106 (2009).
190 Holtzman, D. M., Herz, J. & Bu, G. Apolipoprotein E and apolipoprotein E receptors: normal biology and roles in Alzheimer disease. *Cold Spring Harb Perspect Med* **2**, a006312, doi:10.1101/cshperspect.a006312 (2012).
191 Liu, C. C. *et al.* ApoE4 Accelerates Early Seeding of Amyloid Pathology. *Neuron* **96**, 1024-1032 e1023, doi:10.1016/j.neuron.2017.11.013 (2017).
192 Mishra, A. *et al.* Gene-based association studies report genetic links for clinical subtypes of frontotemporal dementia. *Brain : a journal of neurology* **140**, 1437-1446, doi:10.1093/brain/awx066 (2017).
193 Rostgaard, N. *et al.* TMEM106B and ApoE polymorphisms in CHMP2B-mediated frontotemporal dementia (FTD-3). *Neurobiology of aging* **59**, 221 e221-221 e227, doi:10.1016/j.neurobiolaging.2017.06.026 (2017).
194 Ferrari, R. *et al.* Genetic architecture of sporadic frontotemporal dementia and overlap with Alzheimer's and Parkinson's diseases. *Journal of neurology, neurosurgery, and psychiatry* **88**, 152-164, doi:10.1136/jnnp-2016-314411 (2017).
195 Jairani, P. S., Aswathy, P. M., Gopala, S., Verghese, J. & Mathuranath, P. S. Interaction with the MAPT H1H1 Genotype Increases Dementia Risk in APOE epsilon4 Carriers in a Population of Southern India. *Dement Geriatr Cogn Disord* **42**, 255-264, doi:10.1159/000447446 (2016).
196 Zhao, N. *et al.* APOE epsilon2 is associated with increased tau pathology in primary tauopathy. *Nature communications* **9**, 4388, doi:10.1038/s41467-018-06783-0 (2018).
197 Braak, H. & Braak, E. Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathol* **82**, 239-259, doi:10.1007/BF00308809 (1991).
198 Thal, D. R., Rub, U., Orantes, M. & Braak, H. Phases of A beta-deposition in the human brain and its relevance for the development of AD. *Neurology* **58**, 1791-1800, doi:10.1212/wnl.58.12.1791 (2002).
199 Hussain, Z. *et al.* "Clinical Profile of Genetically Proven Huntington's Disease Patients from Eastern India". *Ann Indian Acad Neurol* **23**, 195-200, doi:10.4103/aian.AIAN_505_19 (2020).
200 Balint, B. *et al.* Dystonia. *Nat Rev Dis Primers* **4**, 25, doi:10.1038/s41572-018-0023-6 (2018).
201 McColgan, P. & Tabrizi, S. J. Huntington's disease: a clinical review. *European journal of neurology* **25**, 24-34, doi:10.1111/ene.13413 (2018).
202 Schneider, S. A. & Bird, T. Huntington's Disease, Huntington's Disease Look-Alikes, and Benign Hereditary Chorea: What's New? *Movement disorders clinical practice* **3**, 342-354, doi:10.1002/mdc3.12312 (2016).
203 Vallee, A. & Lecarpentier, Y. Alzheimer Disease: Crosstalk between the Canonical Wnt/Beta-Catenin Pathway and PPARs Alpha and Gamma. *Frontiers in neuroscience* **10**, 459, doi:10.3389/fnins.2016.00459 (2016).
204 Tapia-Rojas, C. & Inestrosa, N. C. Loss of canonical Wnt signaling is involved in the pathogenesis of Alzheimer's disease. *Neural regeneration research* **13**, 1705-1710, doi:10.4103/1673-5374.238606 (2018).
205 Steffan, J. S. *et al.* The Huntington's disease protein interacts with p53 and CREB-binding protein and represses transcription. *Proceedings of the National*

Academy of Sciences of the United States of America **97**, 6763-6768, doi:10.1073/pnas.100110097 (2000).

206 Giralt, A. *et al.* Long-term memory deficits in Huntington's disease are associated with reduced CBP histone acetylase activity. *Human molecular genetics* **21**, 1203-1216, doi:10.1093/hmg/ddr552 (2012).

207 Nucifora, F. C., Jr. *et al.* Interference by huntingtin and atrophin-1 with cbp-mediated transcription leading to cellular toxicity. *Science* **291**, 2423-2428, doi:10.1126/science.1056784 (2001).

208 (UDN), T. U. D. N. <https://undiagnosed.hms.harvard.edu/genes/mapk8ip3/> (

209 Humphries, C. & Kohli, M. A. Rare Variants and Transcriptomics in Alzheimer disease. *Curr Genet Med Rep* **2**, 75-84, doi:10.1007/s40142-014-0035-9 (2014).

210 Uniprot. *UniProtKB - Q15019 (SEPT2_HUMAN)*, <https://www.uniprot.org/uniprot/Q15019> (

211 Natarajan, C., Sriram, S., Muthian, G. & Bright, J. J. Signaling through JAK2-STAT5 pathway is essential for IL-3-induced activation of microglia. *Glia* **45**, 188-196, doi:10.1002/glia.10316 (2004).

212 ClinVar. *NM_020655.4(JPH3):c.382+760CTG[16] AND Huntington disease-like 2*, <https://www.ncbi.nlm.nih.gov/clinvar/RCV000625421.1/> (

213 Lunnon, K. *et al.* Mitochondrial genes are altered in blood early in Alzheimer's disease. *Neurobiology of aging* **53**, 36-47, doi:10.1016/j.neurobiolaging.2016.12.029 (2017).

214 Chen, J., Bardes, E. E., Aronow, B. J. & Jegga, A. G. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic acids research* **37**, W305-311, doi:10.1093/nar/gkp427 (2009).

215 Blauwendraat, C. *et al.* Genetic modifiers of risk and age at onset in GBA associated Parkinson's disease and Lewy body dementia. *Brain : a journal of neurology* **143**, 234-248, doi:10.1093/brain/awz350 (2020).

216 Man, S. M. & Kanneganti, T. D. Regulation of lysosomal dynamics and autophagy by CTSB/cathepsin B. *Autophagy* **12**, 2504-2505, doi:10.1080/15548627.2016.1239679 (2016).

217 GeneticsHomeReference. *RANBP2 gene - RAN binding protein 2*, <https://ghr.nlm.nih.gov/gene/RANBP2> (

218 Omim. *OSTEOPETROSIS, AUTOSOMAL DOMINANT 2; OPTA2*, <https://www.omim.org/entry/166600> (

219 Liu, E. Y., Cali, C. P. & Lee, E. B. RNA metabolism in neurodegenerative disease. *Dis Model Mech* **10**, 509-518, doi:10.1242/dmm.028613 (2017).

220 Johnson, R. & Buckley, N. J. Gene dysregulation in Huntington's disease: REST, microRNAs and beyond. *Neuromolecular Med* **11**, 183-199, doi:10.1007/s12017-009-8063-4 (2009).

221 Yoshimura, S., Gerondopoulos, A., Linford, A., Rigden, D. J. & Barr, F. A. Family-wide characterization of the DENN domain Rab GDP-GTP exchange factors. *The Journal of cell biology* **191**, 367-381, doi:10.1083/jcb.201008051 (2010).

222 Deal, S. L. & Yamamoto, S. Unraveling Novel Mechanisms of Neurodegeneration Through a Large-Scale Forward Genetic Screen in Drosophila. *Front Genet* **9**, 700, doi:10.3389/fgene.2018.00700 (2018).

223 Khan, N. A. *et al.* Leber's Hereditary Optic Neuropathy-Specific Mutation m.11778G>A Exists on Diverse Mitochondrial Haplogroups in India. *Invest Ophthalmol Vis Sci* **58**, 3923-3930, doi:10.1167/iovs.16-20695 (2017).
224 ClinVar. *NC_012920.1:m.4812G>C AND Leigh syndrome*, <https://www.ncbi.nlm.nih.gov/clinvar/RCV000853824/> (
225 ClinVar. *NC_012920.1:m.3368T>C AND Leigh syndrome*, <https://www.ncbi.nlm.nih.gov/clinvar/47910421/> (
226 Sheremet, N. L. *et al.* Previously Unclassified Mutation of mtDNA m.3472T>C: Evidence of Pathogenicity in Leber's Hereditary Optic Neuropathy. *Biochemistry (Mosc)* **81**, 748-754, doi:10.1134/S0006297916070117 (2016).
227 Finsterer, J. & Zarrouk-Mahjoub, S. Pathogenicity of the LHON variant m.3472T>C is uncertain. *Journal of the neurological sciences* **382**, 164-165, doi:10.1016/j.jns.2017.09.045 (2017).
228 Widgren, P. *et al.* Genetic aetiology of ophthalmological manifestations in children - a focus on mitochondrial disease-related symptoms. *Acta Ophthalmol* **94**, 83-91, doi:10.1111/aos.12897 (2016).
229 ClinVar. *NC_012920.1:m.8887A>G AND Leigh syndrome*, <https://www.ncbi.nlm.nih.gov/clinvar/47911130/> (
230 Whelan, C. D. *et al.* Multiplex proteomics identifies novel CSF and plasma biomarkers of early Alzheimer's disease. *Acta Neuropathol Commun* **7**, 169, doi:10.1186/s40478-019-0795-2 (2019).
231 Kouroupi, G. *et al.* Defective synaptic connectivity and axonal neuropathology in a human iPSC-based model of familial Parkinson's disease. *Proceedings of the National Academy of Sciences of the United States of America* **114**, E3679-E3688, doi:10.1073/pnas.1617259114 (2017).
232 Singleton, R. S. *et al.* OGFOD1 catalyzes prolyl hydroxylation of RPS23 and is involved in translation control and stress granule formation. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 4031-4036, doi:10.1073/pnas.1314482111 (2014).
233 Fan, A. C. & Leung, A. K. RNA Granules and Diseases: A Case Study of Stress Granules in ALS and FTLD. *Adv Exp Med Biol* **907**, 263-296, doi:10.1007/978-3-319-29073-7_11 (2016).
234 Dong, W. *et al.* Proteomics and bioinformatics approaches for the identification of plasma biomarkers to detect Parkinson's disease. *Exp Ther Med* **18**, 2833-2842, doi:10.3892/etm.2019.7888 (2019).
235 Ricobaraza, A. *et al.* Epilepsy and neuropsychiatric comorbidities in mice carrying a recurrent Dravet syndrome SCN1A missense mutation. *Sci Rep* **9**, 14172, doi:10.1038/s41598-019-50627-w (2019).
236 Bender, A. C. *et al.* Focal Scn1a knockdown induces cognitive impairment without seizures. *Neurobiol Dis* **54**, 297-307, doi:10.1016/j.nbd.2012.12.021 (2013).
237 Meier, S. *et al.* SCN1A affects brain structure and the neural activity of the aging brain. *Biol Psychiatry* **72**, 677-683, doi:10.1016/j.biopsych.2012.03.017 (2012).
238 Kim, D. Y., Gersbacher, M. T., Inquimbert, P. & Kovacs, D. M. Reduced sodium channel Na(v)1.1 levels in BACE1-null mice. *J Biol Chem* **286**, 8106-8116, doi:10.1074/jbc.M110.134692 (2011).

239 Martinez-Losa, M. *et al.* Nav1.1-Overexpressing Interneuron Transplants Restore Brain Rhythms and Cognition in a Mouse Model of Alzheimer's Disease. *Neuron* **98**, 75-89 e75, doi:10.1016/j.neuron.2018.02.029 (2018).
240 ClinVar. *NM_006920.6(SCN1A):c.2144-24dup AND Epilepsy*, <https://www.ncbi.nlm.nih.gov/clinvar/RCV000386585.1/> (2016).
241 ClinVar. *NM_006920.6(SCN1A):c.2144-24dup AND Familial hemiplegic migraine*, 2016).
242 Nabbout, R. *et al.* Spectrum of SCN1A mutations in severe myoclonic epilepsy of infancy. *Neurology* **60**, 1961-1967, doi:10.1212/01.wnl.0000069463.41870.2f (2003).
243 Meisler, M. H. & Kearney, J. A. Sodium channel mutations in epilepsy and other neurological disorders. *The Journal of clinical investigation* **115**, 2010-2017, doi:10.1172/JCI25466 (2005).
244 Zeng, L. *et al.* Loss of the Spinocerebellar Ataxia type 3 disease protein ATXN3 alters transcription of multiple signal transduction pathways. *PloS one* **13**, e0204438, doi:10.1371/journal.pone.0204438 (2018).
245 ClinVar. *NM_000153.4(GALC):c.1162-4del AND Galactosylceramide beta-galactosidase deficiency*, <https://www.ncbi.nlm.nih.gov/clinvar/RCV000292380.1/> (
246 Scott-Hewitt, N. J. *et al.* Heterozygote galactocerebrosidase (GALC) mutants have reduced remyelination and impaired myelin debris clearance following demyelinating injury. *Human molecular genetics* **26**, 2825-2837, doi:10.1093/hmg/ddx153 (2017).
247 Kruman, II *et al.* Folic acid deficiency and homocysteine impair DNA repair in hippocampal neurons and sensitize them to amyloid toxicity in experimental models of Alzheimer's disease. *J Neurosci* **22**, 1752-1762 (2002).
248 Pacheco-Quinto, J. *et al.* Hyperhomocysteinemic Alzheimer's mouse model of amyloidosis shows increased brain amyloid beta peptide levels. *Neurobiol Dis* **22**, 651-656, doi:10.1016/j.nbd.2006.01.005 (2006).
249 Kuroha, K., Zinoviev, A., Hellen, C. U. T. & Pestova, T. V. Release of Ubiquitinated and Non-ubiquitinated Nascent Chains from Stalled Mammalian Ribosomal Complexes by ANKZF1 and Ptrh1. *Mol Cell* **72**, 286-302 e288, doi:10.1016/j.molcel.2018.08.022 (2018).
250 Stamberger, H. *et al.* STXBP1 encephalopathy: A neurodevelopmental disorder including epilepsy. *Neurology* **86**, 954-962, doi:10.1212/WNL.0000000000002457 (2016).
251 Olszewska, D. A. *et al.* Autosomal Dominant Gene Negative Frontotemporal Dementia-Think of SCA17. *Cerebellum* **18**, 654-658, doi:10.1007/s12311-018-0998-2 (2019).
252 Ensembl. *TBP-201*, <https://www.ensembl.org/Homo_sapiens/Transcript/Exons?db=core;g=ENSG00000112592;r=6:170554379-170572870;t=ENST00000230354> (
253 Origone, P. *et al.* Spinocerebellar ataxia 17: full phenotype in a 41 CAG/CAA repeats carrier. *Cerebellum Ataxias* **5**, 7, doi:10.1186/s40673-018-0086-x (2018).
254 Shin, J. H. *et al.* The Pathogenic Role of Low Range Repeats in SCA17. *PloS one* **10**, e0135275, doi:10.1371/journal.pone.0135275 (2015).

255 Hayer, S. N. *et al.* STUB1/CHIP mutations cause Gordon Holmes syndrome as part of a widespread multisystemic neurodegeneration: evidence from four novel mutations. *Orphanet journal of rare diseases* **12**, 31, doi:10.1186/s13023-017-0580-x (2017).
256 Miller, V. M. *et al.* CHIP suppresses polyglutamine aggregation and toxicity in vitro and in vivo. *J Neurosci* **25**, 9152-9161, doi:10.1523/JNEUROSCI.3001-05.2005 (2005).
257 Mizukami, T., Kohno, T. & Hattori, M. CUB and Sushi multiple domains 3 regulates dendrite development. *Neuroscience research* **110**, 11-17, doi:10.1016/j.neures.2016.03.003 (2016).
258 Magri, C. *et al.* New copy number variations in schizophrenia. *PloS one* **5**, e13422, doi:10.1371/journal.pone.0013422 (2010).
259 Online Mendelian Inheritance in Man, O. J. H. U., Baltimore, MD. *MIM Number: * 182389*, <https://omim.org/> (
260 Online Mendelian Inheritance in Man, O. J. H. U., Baltimore, MD. *MIM Number: {MIM number}:* 600235*, <https://omim.org/> (
261 Online Mendelian Inheritance in Man, O. J. H. U., Baltimore, MD. *MIM Number: # 157640*, <https://omim.org/> (
262 Keller, A. *et al.* Mutations in the gene encoding PDGF-B cause brain calcifications in humans and mice. *Nature genetics* **45**, 1077-1082, doi:10.1038/ng.2723 (2013).
263 Braunstein, K. E. *et al.* A point mutation in the dynein heavy chain gene leads to striatal atrophy and compromises neurite outgrowth of striatal neurons. *Human molecular genetics* **19**, 4385-4398, doi:10.1093/hmg/ddq361 (2010).
264 Garcia-Moreno, H. *et al.* Xeroderma pigmentosum is a definite cause of Huntington's disease-like syndrome. *Ann Clin Transl Neur* **5**, 102-108, doi:10.1002/acn3.511 (2018).
265 Wright, C. F. *et al.* Assessing the Pathogenicity, Penetrance, and Expressivity of Putative Disease-Causing Variants in a Population Setting. *American journal of human genetics* **104**, 275-286, doi:10.1016/j.ajhg.2018.12.015 (2019).
266 Tsuji, S. [Dentatorubral-pallidoluysian atrophy (DRPLA)--discovery of the disease, DRPLA gene and the pathophysiology]. *Rinsho Shinkeigaku* **40**, 1287-1289 (2000).
267 Costa, M. D. C. *et al.* Exclusion of mutations in the PRNP, JPH3, TBP, ATN1, CREBBP, POU3F2 and FTL genes as a cause of disease in Portuguese patients with a Huntington-like phenotype. *Journal of human genetics* **51**, 645-651, doi:10.1007/s10038-006-0001-9 (2006).
268 Roelfsema, J. H. *et al.* Genetic heterogeneity in Rubinstein-Taybi syndrome: mutations in both the CBP and EP300 genes cause disease. *American journal of human genetics* **76**, 572-580, doi:10.1086/429130 (2005).
269 Contreras-Martos, S. *et al.* Linking functions: an additional role for an intrinsically disordered linker domain in the transcriptional coactivator CBP. *Sci Rep* **7**, 4676, doi:10.1038/s41598-017-04611-x (2017).
270 Platzer, K. *et al.* De Novo Variants in MAPK8IP3 Cause Intellectual Disability with Variable Brain Anomalies. *American journal of human genetics* **104**, 203-212, doi:10.1016/j.ajhg.2018.12.008 (2019).

271 de Araujo, E. D. *et al.* Structural and functional consequences of the STAT5B(N642H) driver mutation. *Nature communications* **10**, 2517, doi:10.1038/s41467-019-10422-7 (2019).
272 Tan, S. H. & Nevalainen, M. T. Signal transducer and activator of transcription 5A/B in prostate and breast cancers. *Endocr Relat Cancer* **15**, 367-390, doi:10.1677/ERC-08-0013 (2008).
273 Rahman, S. & Copeland, W. C. POLG-related disorders and their neurological manifestations. *Nature reviews. Neurology* **15**, 40-52, doi:10.1038/s41582-018-0101-0 (2019).
274 NCBI. *DYNC1H1 dynein cytoplasmic 1 heavy chain 1 [Homo sapiens (human)]*, <https://www.ncbi.nlm.nih.gov/gene/1778> (
275 Stevens, J. C. *et al.* Familial Alzheimer's disease and inherited prion disease in the UK are poorly ascertained. *Journal of neurology, neurosurgery, and psychiatry* **82**, 1054-1057, doi:10.1136/jnnp.2009.199653 (2011).
276 Polderman, T. J. *et al.* Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature genetics* **47**, 702-709, doi:10.1038/ng.3285 (2015).
277 Brainstorm, C. *et al.* Analysis of shared heritability in common disorders of the brain. *Science* **360**, doi:10.1126/science.aap8757 (2018).
278 Ellis, N. *et al.* Genetic Risk Underlying Psychiatric and Cognitive Symptoms in Huntington's Disease. *Biol Psychiatry* **87**, 857-865, doi:10.1016/j.biopsych.2019.12.010 (2020).
279 Coleman, J. R. I. Shared Genetic Risk Between Psychiatric and Cognitive Symptoms in Huntington's Disease and in the General Population. *Biol Psychiatry* **87**, e25-e27, doi:10.1016/j.biopsych.2020.02.1180 (2020).
280 Pharmaceuticals, I. Ionis Pharmaceuticals Receives Orphan Drug Designation for HTTRx. *Biopharm Int* **29**, 57-57 (2016).
281 Caswell, C. *et al.* Genetic predictors of survival in behavioral variant frontotemporal degeneration. *Neurology* **93**, e1707-e1714, doi:10.1212/WNL.0000000000008387 (2019).
282 Borroni, B. *et al.* Genetic background predicts poor prognosis in frontotemporal lobar degeneration. *Neuro-degenerative diseases* **8**, 289-295, doi:10.1159/000322790 (2011).
283 Verheijen, J. *et al.* Common and rare TBK1 variants in early-onset Alzheimer disease in a European cohort. *Neurobiol Aging* **62**, 245 e241-245 e247, doi:10.1016/j.neurobiolaging.2017.10.012 (2018).
284 Ling, S. C., Polymenidou, M. & Cleveland, D. W. Converging mechanisms in ALS and FTD: disrupted RNA and protein homeostasis. *Neuron* **79**, 416-438, doi:10.1016/j.neuron.2013.07.033 (2013).
285 Ciani, M. *et al.* The Missing Heritability of Sporadic Frontotemporal Dementia: New Insights from Rare Variants in Neurodegenerative Candidate Genes. *Int J Mol Sci* **20**, doi:10.3390/ijms20163903 (2019).
286 Ishigaki, S. *et al.* Altered Tau Isoform Ratio Caused by Loss of FUS and SFPQ Function Leads to FTLD-like Phenotypes. *Cell reports* **18**, 1118-1131, doi:10.1016/j.celrep.2017.01.013 (2017).

287 Tan, Q. *et al.* Extensive cryptic splicing upon loss of RBM17 and TDP43 in neurodegeneration models. *Human molecular genetics* **25**, 5083-5093, doi:10.1093/hmg/ddw337 (2016).
288 Metzker, M. L. Sequencing technologies - the next generation. *Nature reviews. Genetics* **11**, 31-46, doi:10.1038/nrg2626 (2010).
289 Loomis, E. W. *et al.* Sequencing the unsequenceable: expanded CGG-repeat alleles of the fragile X gene. *Genome research* **23**, 121-128, doi:10.1101/gr.141705.112 (2013).
290 de Bem, C. M. *et al.* The synergistic risk effect of apolipoprotein epsilon4 and DNA (cytosine-5-)-methyltransferase 3 beta (DNMT3B) haplotype for Alzheimer's disease. *Molecular biology reports* **43**, 653-658, doi:10.1007/s11033-016-3999-6 (2016).
291 Hohman, T. J. *et al.* Discovery of gene-gene interactions across multiple independent data sets of late onset Alzheimer disease from the Alzheimer Disease Genetics Consortium. *Neurobiology of aging* **38**, 141-150, doi:10.1016/j.neurobiolaging.2015.10.031 (2016).
292 Langlois, C. M. *et al.* Alzheimer's Prevention Initiative Generation Program: Development of an APOE genetic counseling and disclosure process in the context of clinical trials. *Alzheimers Dement (N Y)* **5**, 705-716, doi:10.1016/j.trci.2019.09.013 (2019).
293 Ghidoni, R., Paterlini, A., Albertini, V., Binetti, G. & Benussi, L. Losing protein in the brain: the case of progranulin. *Brain research* **1476**, 172-182, doi:10.1016/j.brainres.2012.01.075 (2012).
294 Petkau, T. L. & Leavitt, B. R. Progranulin in neurodegenerative disease. *Trends in neurosciences* **37**, 388-398, doi:10.1016/j.tins.2014.04.003 (2014).
295 Zanardini, R., Ciani, M., Benussi, L. & Ghidoni, R. Molecular Pathways Bridging Frontotemporal Lobar Degeneration and Psychiatric Disorders. *Frontiers in aging neuroscience* **8**, 10, doi:10.3389/fnagi.2016.00010 (2016).
296 Koriath, C. *et al.* ApoE4 lowers age at onset in patients with frontotemporal dementia and tauopathy independent of amyloid-beta copathology. *Alzheimers Dement (Amst)* **11**, 277-280, doi:10.1016/j.dadm.2019.01.010 (2019).
297 Roses, A. D. & Saunders, A. M. APOE is a major susceptibility gene for Alzheimer's disease. *Curr Opin Biotechnol* **5**, 663-667, doi:10.1016/0958-1669(94)90091-4 (1994).
298 Schneider, S. A., Walker, R. H. & Bhatia, K. P. The Huntington's disease-like syndromes: what to consider in patients with a negative Huntington's disease gene test. *Nat Clin Pract Neurol* **3**, 517-525, doi:10.1038/ncpneuro0606 (2007).
299 Stout, J. C. Juvenile Huntington's disease: left behind? *The Lancet. Neurology* **17**, 932-933, doi:10.1016/S1474-4422(18)30334-X (2018).
300 Martinez-Horta, S. *et al.* Neuropsychiatric symptoms are very common in premanifest and early stage Huntington's Disease. *Parkinsonism Relat Disord* **25**, 58-64, doi:10.1016/j.parkreldis.2016.02.008 (2016).
301 Andrews, S. C. *et al.* Apathy predicts rate of cognitive decline over 24 months in premanifest Huntington's disease. *Psychological medicine*, 1-7, doi:10.1017/S0033291720000094 (2020).

302 Andrews, S. C. *et al.* Executive impairment is associated with unawareness of neuropsychiatric symptoms in premanifest and early Huntington's disease. *Neuropsychology* **32**, 958-965, doi:10.1037/neu0000479 (2018).
303 Kume, K. *et al.* Middle-age-onset cerebellar ataxia caused by a homozygous TWNK variant: a case report. *BMC Med Genet* **21**, 68, doi:10.1186/s12881-020-01002-4 (2020).
304 Henao, A. I. *et al.* Characteristic brain MRI findings in ataxia-neuropathy spectrum related to POLG mutation. *Neuroradiol J* **29**, 46-48, doi:10.1177/1971400915621324 (2016).
305 Kim, J. S. Delayed onset mixed involuntary movements after thalamic stroke: clinical, radiological and pathophysiological findings. *Brain : a journal of neurology* **124**, 299-309, doi:10.1093/brain/124.2.299 (2001).
306 Fukui, T. *et al.* Hemiballism-hemichorea induced by subcortical ischemia. *Can J Neurol Sci* **20**, 324-328 (1993).
307 Amlang, C. J., Trujillo Diaz, D. & Louis, E. D. Essential Tremor as a "Waste Basket" Diagnosis: Diagnosing Essential Tremor Remains a Challenge. *Front Neurol* **11**, 172, doi:10.3389/fneur.2020.00172 (2020).
308 Sommer, C. *et al.* Polyneuropathies. *Dtsch Arztebl Int* **115**, 83-90, doi:10.3238/arztebl.2018.083 (2018).
309 van de Zande, N. A. *et al.* Clinical characterization of dystonia in adult patients with Huntington's disease. *European journal of neurology* **24**, 1140-1147, doi:10.1111/ene.13349 (2017).
310 Andriuta, D., Tir, M., Godefroy, O. & Krystkowiak, P. Huntington's Disease Revealed by Familial Cervical Dystonia. *Movement disorders clinical practice* **3**, 415-416, doi:10.1002/mdc3.12311 (2016).
311 Carlozzi, N. E. *et al.* How different aspects of motor dysfunction influence day-to-day function in huntington's disease. *Movement disorders : official journal of the Movement Disorder Society* **34**, 1910-1914, doi:10.1002/mds.27866 (2019).
312 Herzog-Krzywoszanska, R. & Krzywoszanski, L. Sleep Disorders in Huntington's Disease. *Front Psychiatry* **10**, 221, doi:10.3389/fpsyt.2019.00221 (2019).
313 Abbott, S. M. & Videnovic, A. Chronic sleep disturbance and neural injury: links to neurodegenerative disease. *Nat Sci Sleep* **8**, 55-61, doi:10.2147/NSS.S78947 (2016).
314 Chan, J. C. S., Stout, J. C. & Vogel, A. P. Speech in prodromal and symptomatic Huntington's disease as a model of measuring onset and progression in dominantly inherited neurodegenerative diseases. *Neurosci Biobehav Rev* **107**, 450-460, doi:10.1016/j.neubiorev.2019.08.009 (2019).
315 Manor, Y. *et al.* Dysphagia characteristics in Huntington's disease patients: insights from the Fiberoptic Endoscopic Evaluation of Swallowing and the Swallowing Disturbances Questionnaire. *CNS Spectr* **24**, 413-418, doi:10.1017/S1092852918001037 (2019).
316 Pizzorni, N., Pirola, F., Ciammola, A. & Schindler, A. Management of dysphagia in Huntington's disease: a descriptive review. *Neurol Sci*, doi:10.1007/s10072-020-04265-0 (2020).
317 Den Dunnen, W. F. A. Trinucleotide repeat disorders. *Handbook of clinical neurology* **145**, 383-391, doi:10.1016/B978-0-12-802395-2.00027-4 (2017).

318 Rudzinska, M., Krawczyk, M., Wojcik-Pedziwiatr, M., Szczudlik, A. & Tomaszewski, T. Tremor in neurodegenerative ataxias, Huntington disease and tic disorder. *Neurol Neurochir Pol* **47**, 232-240, doi:10.5114/ninp.2013.35585 (2013).

319 Feigin, A. & Talbot, K. Expanding the genetics of huntingtonism. *Neurology* **82**, 286-287, doi:10.1212/WNL.0000000000000067 (2014).

320 Hercher, L. & Jamal, L. An old problem in a new age: Revisiting the clinical dilemma of misattributed paternity. *Appl Transl Genom* **8**, 36-39, doi:10.1016/j.atg.2016.01.004 (2016).

321 Anderson, D. G. *et al.* Comparison of the Huntington's Disease like 2 and Huntington's Disease Clinical Phenotypes. *Movement disorders clinical practice* **6**, 302-311, doi:10.1002/mdc3.12742 (2019).

322 Diehl, S. K. *et al.* Motor speech patterns in Huntington disease. *Neurology* **93**, e2042-e2052, doi:10.1212/WNL.0000000000008541 (2019).

323 Gagnon, M., Barrette, J. & Macoir, J. Language Disorders in Huntington Disease: A Systematic Literature Review. *Cogn Behav Neurol* **31**, 179-192, doi:10.1097/WNN.0000000000000171 (2018).

324 Burnip, E., Wallace, E., Gozdzikowska, K. & Huckabee, M. L. A Systematic Review of Rehabilitation for Corticobulbar Symptoms in Adults with Huntington's Disease. *J Huntingtons Dis* **9**, 1-12, doi:10.3233/JHD-190384 (2020).

325 Bellosta Diago, E. *et al.* Circadian rhythm and autonomic dysfunction in presymptomatic and early Huntington's disease. *Parkinsonism Relat Disord* **44**, 95-100, doi:10.1016/j.parkreldis.2017.09.013 (2017).

326 van Duijn, E. *et al.* Course of irritability, depression and apathy in Huntington's disease in relation to motor symptoms during a two-year follow-up period. *Neuro-degenerative diseases* **13**, 9-16, doi:10.1159/000343210 (2014).

327 Julayanont, P., Heilman, K. M. & McFarland, N. R. Early-motor phenotype relates to neuropsychiatric and cognitive disorders in huntington's disease. *Movement disorders : official journal of the Movement Disorder Society*, doi:10.1002/mds.27980 (2020).

328 Aldaz, T. *et al.* Non-motor symptoms in Huntington's disease: a comparative study with Parkinson's disease. *Journal of neurology* **266**, 1340-1350, doi:10.1007/s00415-019-09263-7 (2019).

329 Rodrigues, G. R. *et al.* Clinical and genetic analysis of 29 Brazilian patients with Huntington's disease-like phenotype. *Arq Neuropsiquiatr* **69**, 419-423, doi:10.1590/s0004-282x2011000400002 (2011).

330 Stevanin, G. *et al.* Huntington's disease-like phenotype due to trinucleotide repeat expansions in the TBP and JPH3 genes. *Brain : a journal of neurology* **126**, 1599-1603, doi:10.1093/brain/awg155 (2003).

331 Koutsis, G. *et al.* Genetic screening of Greek patients with Huntington's disease phenocopies identifies an SCA8 expansion. *Journal of neurology* **259**, 1874-1878, doi:10.1007/s00415-012-6430-9 (2012).

332 Sulek-Piatkowska, A. *et al.* Searching for mutation in the JPH3, ATN1 and TBP genes in Polish patients suspected of Huntington's disease and without mutation in the IT15 gene. *Neurol Neurochir Pol* **42**, 203-209 (2008).

333 Keckarevic, M. *et al.* Yugoslav HD phenocopies analyzed on the presence of mutations in PrP, ferritin, and Jp-3 genes. *Int J Neurosci* **115**, 299-301, doi:10.1080/00207450590519571 (2005).
334 Andrews, T. D., Sjollema, G. & Goodnow, C. C. Understanding the immunological impact of the human mutation explosion. *Trends Immunol* **34**, 99-106, doi:10.1016/j.it.2012.12.001 (2013).
335 Myrick, L. K. *et al.* Fragile X syndrome due to a missense mutation. *European journal of human genetics : EJHG* **22**, 1185-1189, doi:10.1038/ejhg.2013.311 (2014).
336 Zuhlke, C. H. *et al.* Extension of the mutation spectrum in Friedreich's ataxia: detection of an exon deletion and novel missense mutations. *European journal of human genetics : EJHG* **12**, 979-982, doi:10.1038/sj.ejhg.5201257 (2004).
337 Klein, A. D. & Mazzulli, J. R. Is Parkinson's disease a lysosomal disorder? *Brain : a journal of neurology* **141**, 2255-2262, doi:10.1093/brain/awy147 (2018).
338 Palomer, E., Buechler, J. & Salinas, P. C. Wnt Signaling Deregulation in the Aging and Alzheimer's Brain. *Front Cell Neurosci* **13**, 227, doi:10.3389/fncel.2019.00227 (2019).
339 Lisowiec, J., Magner, D., Kierzek, E., Lenartowicz, E. & Kierzek, R. Structural determinants for alternative splicing regulation of the MAPT pre-mRNA. *RNA Biol* **12**, 330-342, doi:10.1080/15476286.2015.1017214 (2015).
340 Galimberti, D., Fenoglio, C. & Scarpini, E. Progranulin as a therapeutic target for dementia. *Expert Opin Ther Targets* **22**, 579-585, doi:10.1080/14728222.2018.1487951 (2018).
341 Sellami, L. *et al.* Distinct Neuroanatomical Correlates of Neuropsychiatric Symptoms in the Three Main Forms of Genetic Frontotemporal Dementia in the GENFI Cohort. *Journal of Alzheimer's disease : JAD* **65**, 147-163, doi:10.3233/JAD-180053 (2018).
342 Kobayashi, Y. *et al.* Pathogenic variant burden in the ExAC database: an empirical approach to evaluating population data for clinical variant interpretation. *Genome Med* **9**, 13, doi:10.1186/s13073-017-0403-7 (2017).
343 Directors, A. B. o. ACMG policy statement: updated recommendations regarding analysis and reporting of secondary findings in clinical genome-scale sequencing. *Genetics in medicine : official journal of the American College of Medical Genetics* **17**, 68-69, doi:10.1038/gim.2014.151 (2015).
344 Gao, L. *et al.* Identifying noncoding risk variants using disease-relevant gene regulatory networks. *Nat Commun* **9**, 702, doi:10.1038/s41467-018-03133-y (2018).
345 Ritchie, G. R., Dunham, I., Zeggini, E. & Flicek, P. Functional annotation of noncoding sequence variants. *Nat Methods* **11**, 294-296, doi:10.1038/nmeth.2832 (2014).
346 Watanabe, K. *et al.* A global overview of pleiotropy and genetic architecture in complex traits. *Nature genetics* **51**, 1339-1348, doi:10.1038/s41588-019-0481-0 (2019).
347 Hakonen, A. H. *et al.* Mitochondrial DNA polymerase W748S mutation: a common cause of autosomal recessive ataxia with ancient European origin. *American journal of human genetics* **77**, 430-441, doi:10.1086/444548 (2005).

348 Schicks, J., Synofzik, M., Schulte, C. & Schols, L. POLG, but not PEO1, is a frequent cause of cerebellar ataxia in Central Europe. *Movement disorders : official journal of the Movement Disorder Society* **25**, 2678-2682, doi:10.1002/mds.23286 (2010).
349 GeneticsHomeReference. *Primary familial brain calcification*, <https://ghr.nlm.nih.gov/condition/primary-familial-brain-calcification#genes> (2020).
350 Goldin, L. R., Cox, N. J., Pauls, D. L., Gershon, E. S. & Kidd, K. K. The detection of major loci by segregation and linkage analysis: a simulation study. *Genet Epidemiol* **1**, 285-296, doi:10.1002/gepi.1370010307 (1984).
351 Goate, A. *et al.* Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. *Nature* **349**, 704-706, doi:10.1038/349704a0 (1991).
352 Mastromoro, G. *et al.* Unusual Segregation of APP Mutations in Monogenic Alzheimer Disease. *Neuro-degenerative diseases* **19**, 96-100, doi:10.1159/000502906 (2019).
353 Mestek-Boukhibar, L. *et al.* Rapid Paediatric Sequencing (RaPS): comprehensive real-life workflow for rapid diagnosis of critically ill children. *Journal of medical genetics* **55**, 721-728, doi:10.1136/jmedgenet-2018-105396 (2018).
354 MacArthur, D. G. *et al.* Guidelines for investigating causality of sequence variants in human disease. *Nature* **508**, 469-476, doi:10.1038/nature13127 (2014).
355 Verheyen, A. *et al.* Genetically Engineered iPSC-Derived FTDP-17 MAPT Neurons Display Mutation-Specific Neurodegenerative and Neurodevelopmental Phenotypes. *Stem Cell Reports* **13**, 434-435, doi:10.1016/j.stemcr.2019.07.007 (2019).
356 GenomicsEngland. (Genomics England, 2019).
357 Gibson, G. Rare and common variants: twenty arguments. *Nature reviews. Genetics* **13**, 135-145, doi:10.1038/nrg3118 (2012).
358 Chen, Y. *et al.* PINK1 and Parkin are genetic modifiers for FUS-induced neurodegeneration. *Human molecular genetics* **25**, 5059-5068, doi:10.1093/hmg/ddw310 (2016).
359 Jain, N. & Chen-Plotkin, A. S. Genetic Modifiers in Neurodegeneration. *Curr Genet Med Rep* **6**, 11-19, doi:10.1007/s40142-018-0133-1 (2018).
360 Kaltenbach, L. S. *et al.* Huntingtin interacting proteins are genetic modifiers of neurodegeneration. *PLoS genetics* **3**, e82, doi:10.1371/journal.pgen.0030082 (2007).
361 Lavoy, S., Chittoor-Vinod, V. G., Chow, C. Y. & Martin, I. Genetic Modifiers of Neurodegeneration in a Drosophila Model of Parkinson's Disease. *Genetics* **209**, 1345-1356, doi:10.1534/genetics.118.301119 (2018).
362 Canela-Xandri, O., Rawlik, K. & Tenesa, A. An atlas of genetic associations in UK Biobank. *Nature genetics* **50**, 1593-1599, doi:10.1038/s41588-018-0248-z (2018).
363 Visscher, P. M. *et al.* 10 Years of GWAS Discovery: Biology, Function, and Translation. *American journal of human genetics* **101**, 5-22, doi:10.1016/j.ajhg.2017.06.005 (2017).

364 Gao, R. *et al.* Mutant huntingtin impairs PNKP and ATXN3, disrupting DNA repair and transcription. *Elife* **8**, doi:10.7554/eLife.42988 (2019).
365 Cho, Y. *et al.* Prevalence of Rare Genetic Variations and Their Implications in NGS-data Interpretation. *Sci Rep* **7**, 9810, doi:10.1038/s41598-017-09247-5 (2017).
366 Papadimitriou, S. *et al.* Predicting disease-causing variant combinations. *Proceedings of the National Academy of Sciences of the United States of America* **116**, 11878-11887, doi:10.1073/pnas.1815601116 (2019).
367 Martin, A. R. *et al.* Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. *American journal of human genetics* **100**, 635-649, doi:10.1016/j.ajhg.2017.03.004 (2017).
368 Verghese, P. B., Castellano, J. M. & Holtzman, D. M. Apolipoprotein E in Alzheimer's disease and other neurological disorders. *Lancet Neurol* **10**, 241-252, doi:10.1016/S1474-4422(10)70325-2 (2011).