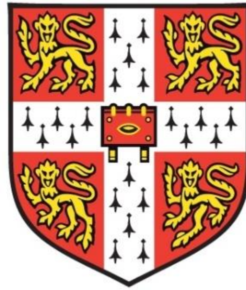
Genetics of Cerebral Small Vessel Disease



Rhea Yan Ying Tan

BSc (Hons)

Supervisor: Professor Hugh S. Markus

Department of Clinical Neurosciences

University of Cambridge

This dissertation is submitted for the degree of

Doctor of Philosophy

Fitzwilliam College

September 2018

Summary

Genetics of Cerebral Small Vessel Disease - Rhea Yan Ying Tan

Cerebral small vessel disease (SVD) is a leading cause of stroke and vascular dementia. The majority of cases are sporadic, occurring in the elderly hypertensive population. However, there also exist patients with familial disease. The most common form is Cerebral Autosomal Dominant Arteriopathy with Subcortical Infarcts and Leukoencephalopathy (CADASIL), caused by mutations in the NOTCH3 gene. In recent years, other genes have also been found to cause familial SVD, such as COL4A1/A2, HTRA1, FOXC1 and TREX1. Genome wide association studies (GWAS) have also revealed loci associated with sporadic SVD strokes and its related features.

This thesis explores the genetic basis of SVD primarily from the angle of the ‘one gene, one disease’ hypothesis. We explore the phenotype of familial SVD using CADASIL as a prototype. We next adopt a candidate gene approach to rare variant discovery using high throughput sequencing (HTS) techniques in two forms: 1) a multi-gene sequencing panel to examine the presence of rare variants in a cohort of 993 presumed-sporadic, early-onset SVD stroke patients, and 2) whole genome sequencing in 118 pedigrees with suspected familial SVD. We also evaluate the prevalence of known disease-causing mutations in the general population using a cohort of whole genome sequenced non-SVD patients, and other control databases.

We demonstrate that a few presumed-sporadic SVD stroke patients may in fact have familial disease that was not previously diagnosed. We show that known and novel rare variants in candidate genes are found in our cohort of familial SVD patients, and suggest a possible role for rare variants in genes associated with related phenotypes and sporadic disease in this cohort. Finally, we identify known disease-causing variants in relatively high frequencies in the population, and show that conclusions on the pathogenicity of variants based on allele frequency and functional analyses may sometimes be misguided, thus highlighting the current limitations we face in the clinical interpretation of variants identified on HTS.

In recent years genetic studies have revealed that pathways in different familial diseases are likely to converge in the pathogenesis of sporadic disease. Further uncovering the genetic basis of undiagnosed cases of familial SVD may shed light on the mechanisms underlying the sporadic form of disease, and may in turn drive the identification of potential therapeutic targets.

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except where declared in the acknowledgments and specifically in the text. The contents of this thesis are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university.

This thesis does not exceed the prescribed word limit for the Clinical Medicine and Clinical Veterinary Medicine Degree Committee (60 000 words).

Rhea Y. Y. Tan

September 2018

Statement of Contribution

The work described in this thesis has been performed primarily in the Department of Clinical Neurosciences and the Department of Haematology at the University of Cambridge.

I have personally screened, identified potential participants and recruited all participants and their relatives into the BRIDGE-Cerebral Small Vessel Disease (SVD) study. Where participants were recruited locally in Cambridge, at St George's Hospital in London or on home visits I have personally performed blood sample collection. I have also performed phenotyping using Human Phenotype Ontology (HPO) codes on all participants.

I have contributed to the design of the SVD-component of the multi-gene sequencing panel, selected gene transcripts on which to report, and curated a database of all genetic variants in monogenic forms of SVD based on published literature, the Human Gene Mutation Database (HGMD), ClinVar and locus-specific databases (LSDBs). I have performed all quantification and QC steps in the ThromboGenomics-SVD study.

I have personally performed analysis and interpretation of all the datasets including statistical analysis. This work was conducted with the support of the following individuals:

Professor Hugh Markus (Professor of Stroke Medicine, Department of Clinical Neurosciences, University of Cambridge) developed the overall concept for this work, established the database of CADASIL patients, and led the collaborations that led to the BRIDGE-SVD and ThromboGenomics-SVD studies. Professor Markus reviewed all participants recruited to the studies, supervised clinical interpretation of genetic results and has been the primary supervisor of this work.

Dr Stefan Gräf (Computational Genome Biologist, Division of Respiratory Medicine and Department of Haematology, University of Cambridge) performed processing of datasets for the BRIDGE-SVD study and supervised the data analysis process. Dr Gräf has been the secondary supervisor of this work.

Dr Kate Downes (Department of Haematology, Cambridge) supervised the sequencing pipeline of the ThromboGenomics-SVD study. Daniel Duarte supervised my work in quantifying and assessing the quality of DNA samples and performed subsequent sequencing steps on the ThromboGenomics platform. Miss Rutendo Mapeta assisted with transcript

selection and curation of variant databases. Dr Matthew Traylor (Department of Clinical Neurosciences, Cambridge), Dr Karyn Megy and Dr Sri Vishnu Vardhan Deevi (Department of Haematology, Cambridge) assisted with variant annotation, filtering and analysis in this study.

Mr Glen Brice (St George's Hospital, London) assisted with identifying patients tested at the South West Thames Regional Genetics Service for NOTCH3 mutations. Dr Ahamad Hassan (Leeds Teaching Hospitals), Dr Kirsty Harkness (Sheffield Teaching Hospitals), Professor Keith Muir (University of Glasgow) and Professor David Werring (University College London Institute of Neurology) assisted with identifying a significant number of cases for the BRIDGE-SVD study. The following research coordinators assisted with sample and data collection in a significant number of cases: Mr Dean Waugh (Leeds Teaching Hospitals), Ms Emma Richards (Sheffield Teaching Hospitals), Mrs Angela Bowring (Southern General Hospital, Glasgow) and Miss Sarah Trippier (St George's Hospital, London).

The following also assisted with site-set up and patient screening: Dr Anand Dixit (Newcastle upon Tyne Hospitals), Dr Julian Barwell (University of Leicester), Dr Usman Khan (St George's Hospital, London) Professor Peter Rothwell (University of Oxford), Dr Bartlomiej Piechowski-Jozwiak (King's College London), Professor Pankaj Sharma (Imperial College London) and Dr Julia Rankin (Royal Devon and Exeter).

Dr. Kathy Stirrups assisted with data management and generation. Dr Matthew Traylor, Dr Marta Bleda and Mr Mathias Haimel (Department of Respiratory Medicine, Cambridge) assisted with data analysis. The Cambridge Translational Genomics laboratory assisted with BRIDGE sample processing, and Dr Jonathan Stephens (Department of Haematology, Cambridge) performed Sanger sequencing in a BRIDGE participant. Members of the NIHR BioResource provided clinical and genetic data and assisted with interpretation of results.

The following individuals also assisted with variant or phenotype interpretation: Dr Nathalie Beaufort (Ludwig-Maximillians University) and Dr Tom Van Agtmael (University of Glasgow), Dr Daniel Gale (University College London) and Professor Irene Roberts (University of Oxford). Modelling of NOTCH3 variants in a patient was performed by Dr Gido Gravesteijn and Dr Saskia Lesnik-Oberstein (Leiden University Medical Centre).

Dedication and Thanks

It has been a privilege to have the excellent mentorship of my supervisors, Prof Hugh Markus and Dr Stefan Gräf. Prof Markus provided me with tremendous support and encouragement, as well as invaluable opportunities to meet interesting families and to take our work from clinic, to bench, and back to the clinic. Having started this PhD fresh out of fourth year of medical school I would never have managed to complete this work without his guidance. I am deeply appreciative to Stefan for his tireless help with the bioinformatic aspects of this work, his approachability and his dedication to teaching. I am also indebted to Steve Bevan, Loes Rutten-Jacobs, Matthew Traylor, Kate Downes and Karyn Megy who have provided me with advice, feedback and encouragement throughout the PhD.

My sincere thanks to the members of the Stroke Research Group and the staff in R3 and R2 - Drs Nick Evans and Anna Drazyk who inspire me with their clinical acumen and academic curiosity, and Maria Spillane, Margaret Rawlings, Premlata Telgote and Daniela Soares-Régua for their support in the research clinic. I also thank Elaine Amis, Jo McGee, Sarah Trippier, Dean Waugh, Daniel Duarte, Brie Stark, the CompGenoMed team and the research coordinators at St George's, UCL, Leeds and Oxford, who have made this PhD so enjoyable.

This work would not have been possible without the generosity of the participants, their families and carers. These individuals have put in time and effort in contacting long-lost relatives, and providing us with information and samples. I also thank my funders, the Agency for Science, Technology and Research (A*STAR), who have been incredibly supportive. SVD is a significant problem in Asia and I look forward to continuing my work with A*STAR in this area.

Finally, I owe my deepest gratitude to my family for their unwavering emotional and intellectual support. I cannot thank my parents enough for their sacrifices to provide only the best for us. I am grateful to Rosel for her enthusiasm for all that I do, and her dedication to providing me with personal and gastronomic support. I thank Jun Wei for devoting his weekends to giving me tutorials on R and Python, and for his support through practical advice, love and laughter. This work is dedicated to my grandfather, who has shown remarkable fortitude in the face of ageing and illness. He remains my greatest source of motivation as I work towards making a meaningful contribution to medicine.

Table of contents

Summary	1
Declaration	2
Statement of Contribution	3
Dedication and Thanks	5
Table of contents	6
List of abbreviations	7
Chapter 1: Introduction	11
Chapter 2: Methodology	61
Chapter 3: CADASIL – Migraine, encephalopathy and stroke	87
Chapter 4: A candidate gene study in presumed sporadic, early-onset small vessel disease	113
Chapter 5: Rare variants in suspected familial SVD: a candidate gene approach using whole genome sequencing	171
Chapter 6: The frequency of known mutations in the general population	255
Chapter 7: Conclusions and future directions	293
References	305
Conference oral presentations	342
Publications	342
Awards and travel grants	343

List of abbreviations

ADAM17	A disintegrin and metalloproteinase 17
AGS	Aicardi-Goutiéres Syndrome
APOE	Apolipoprotein Ɛ
APP	Amyloid Precursor Protein
ARS	Axenfeld-Rieger Syndrome
BBB	Blood brain barrier
BMET	Brief Memory and Executive Test
BMI	Body mass index
BWA	Burrows-Wheeler Aligner
CAA	Cerebral amyloid angiopathy
CADD	Combined Annotation Dependent Depletion
CATGO	Cambridge Translational Genomics
CCDS	Consensus coding sequence
CADASIL	Cerebral Autosomal Dominant Arteriopathy with Subcortical Infarcts and Leukoencephalopathy
CARASAL	Cathepsin A related arteriopathy with strokes and Leukoencephalopathy
CARASIL	Cerebral Autosomal Recessive Arteriopathy with Subcortical Infarcts and Leukoencephalopathy
CECR1	Cats Eye Syndrome Chromosome Region Candidate 1
CHARGE	Cohorts for Heart and Aging Research in Genomic Epidemiology
CNV	Copy Number Variant
CRMCC1	Cerebroretinal Microangiopathy with Calcifications and Cysts 1
CRV	cerebroretinal vasculopathy
CSF	Cerebrospinal fluid
CT	Computed tomography
DADA2	Deficiency of adenosine deaminase 2
DECIPHER	DatabasE of genomiC varIation and Phenotype in Humans using Ensembl Resources
DHPLC	Denaturing high-performance liquid chromatography

dNTP	Deoxynucleoside triphosphate
DWI	Diffusion weighted imaging
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
EGFr	Epidermal growth factor-like repeat
eQTL	Expression quantitative trait loci
ExAC	Exome Aggregation Consortium
FHM	Familial hemiplegic migraine
FLAIR	Fluid attenuation inversion recovery
FOXC1	Forkhead box transcription factor 1
GERP	Genomic Evolutionary Rate Profiling
GnomAD	Genome Aggregation Database
GOM	Granular Osmiophilic Material
GRE	Gradient echo
GWAS	Genome-wide Association Studies
HERNS	Hereditary endotheliopathy, retinopathy, nephropathy and stroke
HGMD	Human Gene Mutation Database
HPO	Human Phenotype Ontology
HTS	High Throughput Sequencing
HVR	Hereditary vascular retinopathy
ICH	Intracerebral haemorrhage
ICHD	International Classification of Headache Disorders
LCC	Cerebral microangiopathy, leukoencephalopathy with calcifications and cysts
LOVD	Leiden Open Variation Database
LSDB	Locus-Specific Databases
MAF	Minor Allele Frequency
MEB	Muscle-Eye-Brain disease
MoCA	Montreal Cognitive Assessment
MRI	Magnetic resonance imaging
OR	Odds ratio

PCR	Polymerase Chain Reaction
PITX2	Paired-like homeodomain transcription factor 2/ pituitary homeobox 2
QC	Quality control
RVCL-S	Retinal Vasculopathy with Cerebral Leukodystrophy and Systemic Manifestations
SD	Standard deviation
SIFT	Sorting Intolerant from Tolerant
SNP	Single nucleotide polymorphisms
SNV	Single Nucleotide Variant
SVD	Small vessel disease
TBC1D7	TBC1 Domain Family Member 7
TGF	Transforming Growth Factor
TIA	Transient ischaemic attack
TIMP3	Tissue inhibitor of metalloproteinase 3
TOAST	Trial of Org 10172 in Acute Stroke Treatment
TSC	Tuberous sclerosis complex
UTR	Untranslated region
VUCS	Variants of uncertain clinical significance
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing
WMH	White matter hyperintensity
WWS	Walker-Warburg Syndrome

Chapter 1: Introduction

Outline

1.1 Overview

1.2 Stroke and Cerebral SVD

1.2.1 Arteriosclerotic SVD

1.2.2 Cerebral amyloid angiopathy

1.3 Genetics of Stroke

1.3.1 Monogenic causes of stroke

1.3.2 Monogenic causes of SVD

1.4 The role of the extracellular matrix in the pathogenesis of SVD: new insights from monogenic SVD

1.4.1 Matrisome involvement in CADASIL

1.4.2 Convergence of disease mechanisms in CADASIL and CARASIL: relevance to sporadic SVD

1.4.3 Collagen: involvement of the most abundant matrisome protein

1.5 Other possible mechanisms of disease – insights from genetics

1.5.1 Blood Brain Barrier development and integrity

1.5.2 Impairment of DNA Damage Response

1.6 Further insights from GWAS: distinct mechanisms across the SVD spectrum

1.7 Concluding remarks

1.7.1 Reaching a genetic diagnosis in familial SVD

1.7.2 Pathogenesis of SVD: what we understand from genetic studies

List of tables

Table 1-0 Monogenic disorders causing ischaemic stroke

Table 1-1 Monogenic disorders causing haemorrhagic strokes

Table 1-3 Monogenic disorders with strokes arising from SVD as a major clinical feature

Table 1-4 Monogenic disorders with strokes arising from both small and large artery disease as a clinical feature

Table 1-5 Clinical syndromes in COL4A1-related SVD

Table 1-6 Genes and key clinical features in hereditary cerebral amyloid angiopathy (CAA) and related amyloid disorders

Table 1-7 Matrisome proteins found to co-aggregate with NOTCH3 ^{ECD}

Table 1-8 Review of studies of the association between common or rare variants in Mendelian disease genes and sporadic SVD or its associated markers

Table 1-9 SNPs associated with WMH in community and stroke populations

List of figures

Figure 1-1 Radiological features of CADASIL on MRI

Figure 1-2 MRI appearances in two patients with COL4A1 mutations

Figure 1-3 The involvement of matrisome proteins in the pathogenesis of CADASIL and CARASIL

Figure 1-4 The convergence of disease pathways, particularly in the extracellular matrix (ECM), in the mechanisms underlying monogenic SVD

1. Introduction

1.1. Overview

Cerebral small vessel disease is an umbrella term which broadly describes the disease of small blood vessels in the brain. It encompasses two categories of disease: arteriosclerotic SVD arising mainly from hypertension and other cardiovascular risk factors, and cerebral amyloid angiopathy (CAA). Arteriosclerotic SVD, which is the focus of this thesis, is a major cause of ischaemic stroke and vascular dementia.

Among the subtypes of stroke, the mechanisms of SVD are perhaps the least understood. Genetic studies of SVD, as well as the study of inherited forms of SVD, are beginning to give us insights into possible mechanisms that underlie the disease in general. This thesis aims to contribute to our understanding of the underlying pathophysiology of SVD by exploring the genetic basis of the disease.

The following chapters will examine the genetics of SVD by considering the role of single genes. Beginning with a monogenic disease, Cerebral Autosomal Dominant Arteriopathy with Subcortical Infarcts and Leukoencephalopathy (CADASIL) as a 'model' of SVD, it goes on to examine rare variants in multiple candidate genes in patients with presumed 'sporadic' SVD. Finally, it explores the role of rare variants in candidate genes in families with suspected familial SVD, and investigates the presence of known monogenic disease mutations in the general population.

The chapters are briefly summarised as follows:

Chapter 1: Introduction. This chapter is a review of the monogenic causes of SVD and the overlapping mechanisms underlying these diseases. This review brings together recent data from studies in monogenic SVD and genetic studies in 'sporadic' SVD. It aims to show how these provide new insights into the pathogenesis of SVD, and highlights the possible convergence of disease mechanisms in monogenic and sporadic SVD.

Chapter 2: Materials and Methods

Chapter 3: Migraine, encephalopathy and stroke and their inter-relationships in CADASIL. This is a cross-sectional study of a cohort of patients diagnosed with CADASIL.

Chapter 4: A candidate gene study in presumed sporadic SVD. This chapter explores the hypothesis that monogenic disease genes also have a role in sporadic SVD strokes, using a next generation sequencing platform.

Chapter 5: Rare variants in suspected familial SVD: a candidate gene approach using whole genome sequencing. This chapter examines the prevalence of rare variants in known SVD genes, and other candidate genes associated with related phenotypes such as leukodystrophy, familial hemiplegic migraine, connective tissue disease and cerebral amyloid angiopathy in a cohort of patients who are suspected to have monogenic SVD.

Chapter 6: Frequency of known disease-causing mutations in control populations. In a population of unrelated controls, the prevalence of known disease-causing mutations is evaluated using whole genome sequencing. This section explores the question of whether known 'mutations' are truly fully penetrant disease-causing variants, or are part of a 'background carrier rate', and in fact require other genetic or environmental factors in order to cause disease.

Chapter 7: Summary and discussion of findings

1.2. Stroke and Cerebral SVD

A stroke is defined as the acute onset of focal neurological disturbance arising due to a cerebrovascular cause, either confirmed on histology or on imaging, where other causes have been excluded.[1] Although historically understood to be a single entity, it is now recognised that stroke is a broad term describing the acute culmination of multiple disease processes, each of which can lead to either ischaemic (80%) or haemorrhagic (20%) events.[2]

The three most common causes of ischaemic strokes are large vessel atherosclerotic disease (LVD), cerebral small vessel disease (SVD) and cardioembolic causes. Other rare causes include cervical artery dissection or vertebral artery dissection, while up to a quarter of ischaemic strokes may remain cryptogenic with no underlying cause identified[3] Haemorrhagic strokes are typically classified according to their location in the brain, with subcortical or deep haemorrhages arising due to hypertensive cerebral small vessel disease, and lobar or cortical bleeds arising due to cerebral amyloid angiopathy (see 1.2.2) or cerebrovascular abnormalities such as arteriovenous malformations.

1.2.1. Arteriosclerotic SVD

SVD arising from arteriosclerosis of the small perforating arteries supplying the white matter and deep grey matter of the brain accounts for up to a fifth of ischaemic strokes, and is also an important cause of haemorrhages occurring in the subcortical or deep regions of the brain (deep intracerebral haemorrhage, ICH). It is also the most common cause of vascular cognitive impairment and vascular dementia, which is the second most common cause of dementia.[4]

This type of SVD is characterised radiologically by the following features best seen on MRI: lacunar infarcts of presumed ischaemic origin, symptomatic subcortical haemorrhage, white matter hyperintensities (WMH) on T2-weighted/FLAIR MRI, which also corresponds to low signal on CT described as leukoaraiosis, cerebral microbleeds on gradient echo MRI, dilated perivascular spaces and brain atrophy.[5] On histopathological examination the deep perforating arteries in non-amyloid SVD show features of focal atherosclerosis both at the origin of and within the proximal perforating arteries, as well as diffuse abnormalities in the small perforating vessels. These include fibrinoid necrosis and lipohyalinosis, where the smooth muscle cells are replaced with collagen, amyloid and other hyaline material, leading to narrowing of the vessel lumen.[4,6]

The majority of SVD strokes are thought to be sporadic, occurring in the elderly population with hypertension being the major risk factor. Despite its prevalence, the pathogenesis of SVD is poorly understood in comparison with the other types of ischaemic stroke. It is hypothesized that the histopathological changes of the small blood vessels as described above result in reduced cerebral blood flow[7] and cerebral autoregulation,[8] which in turn cause hypoperfusion, and this has been demonstrated on imaging studies. Endothelial dysfunction has also been found to feature early in the disease, and this may contribute to the impairment of cerebral autoregulation.[9]

More recently it has been proposed that increased Blood Brain Barrier (BBB) permeability may play an important role in the development of SVD.[10] Neuropathological studies have shown the presence of plasma proteins such as fibrinogen in the brain parenchyma, suggesting that there had been extravasation of these proteins across a transiently open BBB.[11,12] Evidence of previous BBB disruption is also provided by cerebrospinal fluid (CSF) studies showing the presence of plasma proteins in the CSF.[13] Further support is provided by recent MRI studies demonstrating leakage of contrast agents such as gadolinium across the BBB.[14] It is likely that both hypoperfusion and increased BBB permeability interact, and endothelial dysfunction and activation could contribute to both.

1.2.2. Cerebral amyloid angiopathy

Cerebral amyloid angiopathy (CAA) is a term used to describe a number of cerebral vasculopathies which involve the deposition of amyloid fibrils in the walls of the small or medium cerebral blood vessels, as well as the capillaries of the brain parenchyma and leptomeninges.[15] These depositions are altered proteins which have adopted a β-pleated sheet conformation. CAA is classically characterised by large lobar (cortical) haemorrhages, but can also cause transient behavioural changes, seizures and cognitive impairment leading to dementia. On MRI, patients often have a distinctive distribution of classical SVD features, such as WMH with a predominant posterior distribution, and cerebral microbleeds in the lobar regions as visualised on gradient-echo MR imaging.[16] They may also have cortical superficial siderosis, a marker also used in the radiological diagnosis of CAA.[17] Possible or probable CAA can be diagnosed in life based on the clinical syndrome and imaging features with the aid of the modified Boston criteria, and a definitive diagnosis can be made with the

demonstration of amyloid fibril deposition on post-mortem histopathological analysis.[17] More recently, the Edinburgh criteria which uses CT and apolipoprotein Ɛ (APOE) genotyping to estimate the probability of CAA has also been proposed.[18]

1.3. Genetics of Stroke

The majority of strokes in general, and SVD strokes in particular, are thought to be 'sporadic', resulting from the interaction of multiple genetic and environmental factors. Early studies in animals,[19] twins, affected sibling-pairs [20,21] and of familial aggregation of strokes suggested that stroke is highly heritable. These findings were also supported by findings from epidemiological studies looking at the family history of stroke [22] and genome-wide complex trait analysis studies using genome-wide data, which also showed that younger-onset strokes have a higher heritability[23,24]. Most of the variants identified in genome-wide association studies (GWAS) of ischaemic stroke have also been found to be specific to one subtype of stroke, providing further evidence that each subtype of stroke has a different aetiology and genetic background.[23–26] For example, a locus at chromosome 7p21, in the region of the histone deacetylase 9 (HDAC9) gene was found to be associated with LVD stroke, and PITX2 and ZFHX3 were identified in cardioembolic stroke.[23] GWAS studies have also identified susceptibility loci for intracerebral haemorrhage (ICH), however these early studies in stroke did not identify genome-wide significant loci for SVD.[27]

A growing body of evidence now suggests that genetic susceptibility is particularly important in 'sporadic' SVD. This includes epidemiological data showing that family history of stroke is a risk factor for SVD,[28] and recent genome-wide association study (GWAS) data demonstrating a significant heritability for 'sporadic' SVD of the predominant lacunar ischaemic stroke sub-phenotype.[29] More recent GWAS findings in SVD are discussed in sections 1.4 and 1.6.

There are also both hereditary and sporadic forms of CAA. CAA also typically occurs as a sporadic disease in the elderly population, with the deposition of amyloid beta (Aβ) protein in the walls of the blood vessels being the most common form. It is thus commonly associated with parenchymal Aβ plaques and Alzheimer's disease.[30] Like Alzheimer's disease, the APOE genotype is a risk factor for sporadic CAA and CAA-related intracerebral haemorrhage (ICH). The Ɛ2 genotype is associated with the presence of CAA, and both the Ɛ2 and Ɛ4 genotypes

are associated with an increased risk of lobar ICH, recurrent lobar ICH and an earlier age at onset of lobar ICH.[31]

1.3.1. Monogenic causes of stroke

Strokes may also present as a feature of monogenic disease. Most inherited forms of stroke cause a specific stroke syndrome due to a single mechanism, however some syndromes may lead to strokes through more than one mechanism. Tables 1-1 and 1-2 summarise the known monogenic forms of stroke for ischaemic and haemorrhagic strokes respectively.

***Table 1-1**: Monogenic disorders causing ischaemic stroke, classified according to the stroke subtype they result in. The list is not exhaustive but includes examples of the major stroke subtypes.*

Stroke subtype	Monogenic disorders	OMIM	
Small vessel disease	CADASIL	#125310	
	CARASIL / HTRA1-related autosomal dominant SVD	#600142	
	CARASAL	-	
	Retinal Vasculopathy with Cerebral Leukodystrophy and Systemic Manifestations (RVCL-S)	#192315	
	FOXC1-related SVD	-	
	COL4A1/A2 –related small vessel arteriopathy with haemorrhage and intracerebral aneurysms	#607595 #614519	#611773 #175780
	Cerebral microangiopathy, leukoencephalopathy with calcifications and cysts (LCC)	#614561	
	Deficiency of ADA2	#615688	
Large artery atherosclerosis and other arteriopathies	Familial hyperlipidaemias	#144250 #602491 #604499	#145750 #143890
	Moya-moya disease	#252350 #607151 #608796	#300845 #614042
	Pseudoxanthoma elasticum	#264800	
	Neurofibromatosis type I	#162200	
Large artery disease – dissection	Ehlers Danlos Syndrome Type IV	#130050	
	Marfan syndrome	#154700	
	Fibromuscular dysplasia	#135580	
	Arterial Tortuosity Syndrome	#208050	
Disorders affecting both small and large arteries	Fabry disease	#301500	
	Homocystinuria	#236200	
	Sickle cell disease	#603903	

Stroke subtype	Monogenic disorders	OMIM	
Cardioembolism	Familial cardiomyopathies Familial arrhythmias	>40 different types	
	Hereditary Haemorrhagic Telangiectasia	#187300	
Prothrombotic disorders	Factor V Leiden	#612309	
	Prothrombin (F2), Protein S (PROS1), Protein C (PROC), Antithrombin III (AT3) deficiencies	#613679 #612336	#613118 #176860
Mitochondrial disorders	Mitochondrial myopathy, Encephalopathy, Lactic Acidosis and Stroke (MELAS)	#540000	

Table 1-2: *Monogenic disorders causing haemorrhagic strokes.*

Stroke subtype	Monogenic disorders	OMIM	
Small vessel disease	COL4A1/A2* (subcortical haemorrhages)	#607595 #614519	#611773 #175780
	Hereditary cerebral amyloid angiopathy (lobar haemorrhages)	#605714 #105150 #176500 #117300	#105210 #105120 #137440
Large artery disease – rupture of cerebrovascular malformations	Cerebral aneurysms • Familial intracranial aneurysm	#105800 #609122 #612586	#300870
	• Autosomal dominant polycystic kidney disease	#173900 #613095	#600666
	• Loeys-Dietz Syndrome	#613795	
	• COL4A1/A2	#607595 #614519	#611773 #175780
	Arteriovenous malformation • Hereditary haemorrhagic telangiectasia	#187300	
	• Capillary malformation – arteriovenous malformation	#608354	
	Venous malformations Familial cerebral cavernous malformation	#116860 #603284 #603285	#108010

Syndromes marked with a * predominantly cause haemorrhagic strokes, but may also cause ischaemic strokes.

1.3.2. Monogenic causes of SVD

1.3.2.1. Monogenic diseases which cause ischaemic stroke

SVD is the subtype of stroke that is most likely to present as a familial disease.[4] There are several monogenic forms of SVD, which not only recapitulate the clinical, radiological and histopathological features of classical sporadic SVD, but may also have other distinctive neurological or non-neurological features. (Table 1-3 and Table 1-4) Although rare, these extremes in phenotype share both clinical and radiological features with sporadic SVD, and the study of these provide important insights into the mechanisms of SVD.

Table 1-3: Monogenic disorders with strokes arising from SVD as a major clinical feature

Disease (OMIM #)	Gene(s) and chromosome	Gene products and function	Key clinical features	Inheritance mode, mutations	Key diagnostic or radiological features
CADASIL: Cerebral Autosomal Dominant Arteriopathy with Subcortical Infarcts and Leukoencephalopathy (#125310)	NOTCH3 Chr 19	Notch3 transmembrane receptor. Has roles in angiogenesis, vascular smooth muscle cell remodelling [32]	• Migraine with aura • Subcortical ischaemic stroke • Psychiatric disturbance • Subcortical cognitive impairment and dementia • Seizures • Encephalopathy	Autosomal dominant, Cys-altering mutations in exons 2 – 24	• WMH involve anterior temporal poles and external capsules • Granular Osmiophilic Material (GOM) surrounding smooth muscle cells of systemic small vessels • Diagnosis with skin biopsy or molecular genetic testing
CARASIL: Cerebral Autosomal Recessive Arteriopathy with Subcortical Infarcts and Leukoencephalopathy (#600142)	HTRA1 Chr 10	High temperature requirement serine protease A1. Switches off transforming growth factor β pathway[33]	• Subcortical ischaemic strokes • Cognitive impairment • Early-onset diffuse alopecia • Degenerative disc disease	Autosomal recessive and dominant missense and nonsense mutations	Diagnosis on molecular genetic testing
CARASAL: Cathepsin A related arteriopathy with strokes and Leukoencephalopathy	CTSA Chr 20	Cathepsin A. Has roles in the lysosomal transport, activation and stabilisation of β-galactosidase and neuraminidase-1. Inactivates selected neuropeptides and regulates a lysosomal pathway of protein degradation.[34]	• Subcortical ischaemic and haemorrhagic strokes • Dry mouth, difficulty swallowing, dry eyes, muscle cramps • Cognitive impairment • Treatment-resistant hypertension	Autosomal dominant c.973C>T	Diagnosis on molecular genetic testing

Disease (OMIM #)	Gene(s) and chromosome	Gene products and function	Key clinical features	Inheritance mode, mutations	Key diagnostic or radiological features
COL4A1/A2 related SVD (#611773)	COL4A1, COL4A2 Chr 13	Alpha 1 and Alpha 2 chains of collagen IV. Encode α1 and α2 collagen chains, which are the most abundant components of the extracellular matrix [35]	Paediatric: • Porencephaly • Infantile hemiparesis • Developmental delay Adult: • Subcortical ischaemic and haemorrhagic strokes • Cognitive impairment • Retinal small vessel abnormalities Systemic: • Renal cysts, haematuria • Muscle cramps with raised creatine kinase • Cerebral aneurysms	Autosomal dominant and recessive missense and nonsense mutations, most commonly affecting Gly in Gly-X-Y repeats	• Significant subcortical microbleeds on MRI • Periventricular cysts • Diagnosis on molecular genetic testing
RVCL-S Retinal vasculopathy with cerebral leukodystrophy and systemic manifestations (#192315)	TREX1 Chr 3	Three prime repair exonuclease. Encodes DNase III (Three prime repair exonuclease), which has roles in DNA repair [36]	• Visual loss • Migraines • Subcortical ischaemic and haemorrhagic strokes • Cognitive impairment • [Hereditary systemic angiopathy in some patients: Raynaud's phenomenon, hepatic cirrhosis, renal dysfunction, osteonecrosis]	Autosomal dominant frameshift mutations in C-terminus	• Subcortical 'pseudotumours' – contrast-enhancing mass lesions with surrounding oedema • Diagnosis on molecular genetic testing
FOXC1/PITX2-related SVD	FOXC1 PITX2 Chr 6	Forkhead box transcription factor C1. Has roles in blood vessel development [37] / Paired-like homeodomain transcription factor 2. Determines left-right asymmetry of internal organs [38]	• Subcortical infarcts • Axenfeld Rieger Syndrome • Cerebellar malformations • Hearing impairment	Autosomal dominant or de novo 6p25 deletions or duplications	Diagnosis on molecular genetic testing

Disease (OMIM #)	Gene(s) and chromosome	Gene products and function	Key clinical features	Inheritance mode, mutations	Key diagnostic or radiological features
Cerebral microangiopathy, leukoencephalopathy with calcifications and cysts (LCC) and Coats plus syndrome (Coats disease + LCC)	SNORD118 in LCC, CTC1 in Coats disease/ Coats plus (and not in LCC)	Box C/D snoRNA U8. Has function in methylation of ribosomal RNA Conserved telomere maintenance component 1. Has roles in maintaining telomere ends.	LCC- exclusively neurological: • Seizures • Dystonia, spasticity, ataxia • Hemiplegia • Cognitive delay Coats plus– multisystem: • Features of LCC • Sparse pale hair • Skin depigmentation • Dysplastic nails • Intrauterine growth retardation • Retinal telangiectasias and exudates • Vascular ectasias • Osteopenia	Autosomal recessive (homozygosity or compound heterozygosity). Missense mutations, and complete gene deletion	Leukoencephalopathy, calcification and cysts seen in LCC and Coats plus, but not Coats disease. Cysts not seen in all LCC patients Angiomatous blood vessels, gliosis and Rosenthal fibres in white matter
Deficiency of ADA2 (DADA2)	CECR1	Adenosine Deaminase 2 (ADA2). Roles in downregulation of extracellular adenosine, and cellular proliferation and differentiation.	• Intermittent fevers • Small subcortical ischaemic and haemorrhagic strokes • Raised acute phase proteins • Livedoid rash • Hepatosplenomegaly • Hypogammaglobulinaemia	Autosomal recessive (homozygosity or compound heterozygosity). Missense mutations	White matter relatively spared. Skin biopsies show neutrophils and macrophages in interstitium with perivascular T lymphocytes.

Table 1-4: *Monogenic disorders with strokes arising from both small and large artery disease as a clinical feature*

Disease (OMIM #)	Gene(s) and chromosome	Gene products	Key clinical features	Inheritance mode, mutations	Key diagnostic or radiological features
Fabry disease (#301500)	GLA X	Alpha galactosidase A enzyme	• Acute pain crises (small fibre peripheral neuropathy) • Ischaemic and haemorrhagic strokes • Renal failure • Angiokeratomas • Tortuous retinal vessels, whorl keratopathy on slit-lamp exam • Dolichoectasia of intracranial arteries • Cardiomyopathy, hypertension • Lipid staining may show Gb3 'zebra bodies' in lysosomes of kidney biopsy	X-linked missense mutations	• Biochemical diagnosis with low α-galactosidase in plasma and peripheral leukocytes (males) • Genetic testing in females with low-normal α-galactosidase
Pseudoxanthoma elasticum (#264800)	ABCC6 16	Multidrug resistance-associated protein 6	• Angina, hypertension, intermittent claudication, restrictive cardiomyopathy, mitral valve prolapse • GIT bleeding • Ischaemic strokes • Yellow popular skin lesions • Visual impairment, peau d'orange changes, angioid streaks on retina, neovascularisation and retinal haemorrhages • Calcified, fragmented elastic fibres in skin, eyes, arteries	Autosomal dominant/ autosomal recessive missense mutations	Diagnosis only by molecular genetic testing, not skin biopsy.
Neurofibromatosis 1 (#162200)	NF1 17	Neurofibromin	• Neurofibromas, café-au-lait spots, freckling • Lisch nodules • Neurofibromas, Schwannomas • Dural ectasia • Scoliosis, dysplasia • Phaeochromocytoma • Strokes	Autosomal dominant missense mutations and deletions	

CADASIL

The most common familial form of SVD is Cerebral Autosomal Dominant Arteriopathy with Subcortical Infarcts and Leukoencephalopathy (CADASIL), caused by mutations in the NOTCH3 gene.[39] These are cysteine-altering mutations (either a gain or loss of a cysteine residue) in the epidermal growth factor-like repeat (EGFr) domains encoded by exons 2 to 24 of the NOTCH3 gene.[39] CADASIL is the most common inherited form of stroke, with an estimated population prevalence in the UK of about 2 to 4 per 100 000.[40,41] A study of 994 apparently sporadic lacunar stroke patients aged ≤70 years found CADASIL mutations in 0.5%; this was higher (1.5%) in patients with confluent leukoaraiosis on MRI.[42]

CADASIL is typically characterised by exclusively neurological features, the most prominent being migraine, subcortical ischaemic lacunar infarcts and cognitive impairment. Migraine, usually with aura, is the first symptom in 60-75% of patients, and usually arises in the second to third decade. The aura is typically visual or sensory, but can also be dysphasic. Confusion is not an uncommon aura, and may form one end of a spectrum with the encephalopathic presentation (further discussed in Chapter 3). The subcortical strokes tend to be ischaemic lacunar infarcts, with a mean age at onset of 47 years in a recent study, although the first stroke can occur as late as in the 70s.[43]

Psychiatric disturbances, particularly depression and apathy, but also anxiety, are common, and depression often precedes other symptoms. Cognitive impairment, with early involvement of selective executive dysfunction and impaired processing speed is common in the fourth and fifth decades, and may progress to dementia.[44] Some patients may present with dementia and without symptomatic strokes.[45] Seizures, both focal and generalised, are rare but well described.[44]

Patients may experience encephalopathic episodes (also known as a 'CADASIL coma'), a reversible acute confusion or coma often with seizures usually developing from a migraine and lasting as long as 7 – 14 days.[43] These may be susceptible to hormonal influences, occurring frequently around the puerperium.[46] Spinal cord involvement is very rare but a few cases have been reported.[47] A number of reports have also documented the occurrence of ocular involvement defined by retinal arteriolar narrowing on fundoscopy or fluorescence angiography, with or without patchy visual field loss [48–50] .

The earliest radiological features are patchy T2 white matter hyperintensities (WMH) up to 15 years prior to the onset of symptoms, with patients typically having an abnormal MRI by age 35.[51] As symptoms develop, this progresses to confluent WMH with characteristic involvement of the anterior temporal pole. (Figure 1-1) This feature is rare in sporadic SVD, but has been shown to have high sensitivity and specificity in CADASIL.[52] Other MRI markers are seen in sporadic SVD, such as subcortical lacunar infarcts, microbleeds and dilated perivascular spaces, and cerebral atrophy.[53]

Although commonly described as an exclusively neurological disease, a few case reports have documented co-existing renal abnormalities.[54,55] The systemic small blood vessels also show distinctive histopathological features seen electron microscopy, with granular osmiophilic material (GOM) accumulating around the smooth muscle cells of blood vessels. GOM deposition gradually replaces the vascular smooth muscle cells, and occurs in both peripheral and CNS arteries, primarily in the leptomeningeal and small penetrating arteries.[56] Electron microscopy for GOM in skin and muscle biopsies has been used in the diagnosis of CADASIL, in particular where patients have not tested positive on a screen for NOTCH3 mutation.[56]

Patients can vary markedly in terms of the type of clinical features, their severity and the ages at onset, both between and within families.[44,45,57] Disease severity is not thought to relate to mutation site,[58] although a recent study in Dutch CADASIL patients has suggested the possibility of a more severe phenotype arising with mutations in EGFr domains 1 – 6, and a milder phenotype with mutations in other EGFr domains.[59] Family-based studies suggest that other genetic factors may influence disease severity. Both hypertension and smoking have been associated with an earlier age at onset of stroke in CADASIL, while higher blood pressure has been related to more rapid progression of MRI lesion volume.[60]

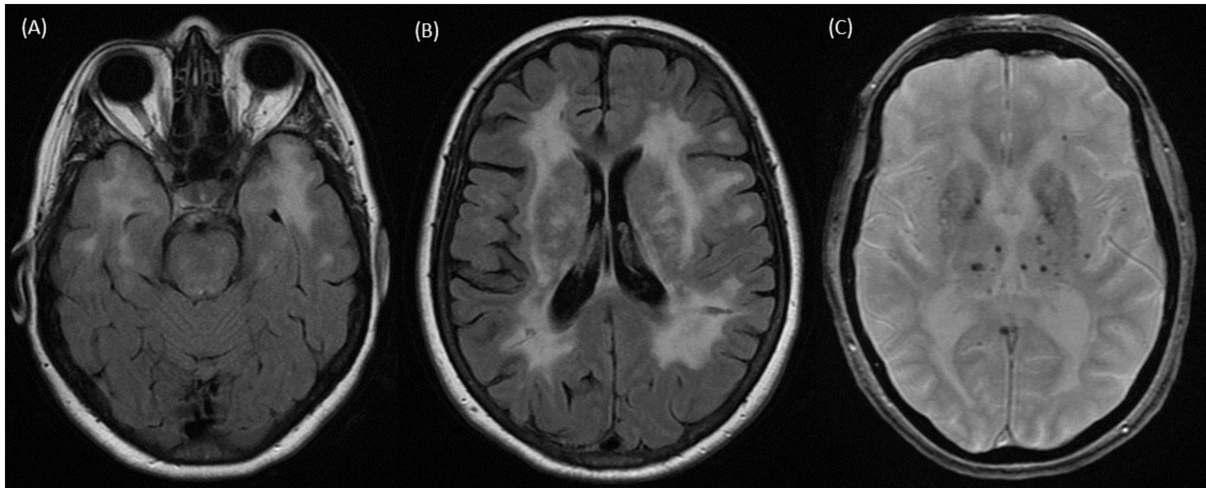


Figure 1-1*: Radiological features of CADASIL on MRI. White matter hyperintensities (WMH) involving the anterior temporal pole (A) and external capsules (B) are seen on T2/FLAIR MRI sequences. (C) Subcortical cerebral microbleeds can be seen in some patients on T2* Gradient Recalled Echo (T2* GRE) sequences.*

CARASIL

An autosomal recessive form of familial SVD has also been described in consanguineous Japanese and Chinese families, and has been attributed to mutations in the HTRA1 gene.[61] This disease is known as Cerebral Autosomal Recessive Arteriopathy with Subcortical Infarcts and Leukoencephalopathy (CARASIL), and is caused by the inheritance of a missense or nonsense mutation in each copy of the HTRA1 allele.[62] It has been described in around 50 individuals in Japanese and Chinese pedigrees showing consanguinity,[62–64] and recently cases have also been described in Europeans with compound heterozygous or homozygous HTRA1 mutations.[65,66] A less severe form of the disease due to heterozygous mutations in the same gene has been described in Caucasian and Japanese populations.[67,68]

Patients with CARASIL also develop SVD manifesting as recurrent subcortical ischaemic lacunar stroke, progressive cognitive and motor impairment, seizures and psychiatric disturbances including personality changes, emotional lability, abulia and apallic syndrome.[62] Unlike CADASIL, these patients also develop non-neurological symptoms including early-onset diffuse alopecia and degenerative disc disease resulting in acute middle to lower back pain. Patients with CARASIL have a more rapid progression of symptoms, developing dementia and becoming bedridden by around 40 years.[62]

On MRI, confluent hyperintensities in the deep white matter (WMH) and periventricular regions are seen, and these changes are thought to precede the onset of symptoms.[69] The

involvement of the anterior temporal poles and external capsules has also been documented in some patients.[63] Histopathological features of CARASIL are similar to those seen in sporadic SVD, such as arteriopathy (loss of vascular muscle cells) arteriosclerosis (hyaline degeneration of the media), non-occlusive fibrous thickening of the intima, and fragmentation of the internal elastic lamina. No GOM deposits have been found.[63,70–72]

RVCL-S

Retinal vasculopathy with cerebral leukodystrophy and systemic manifestations (RVCL-S) is an autosomal dominant disease of the cerebral small vessels caused by frameshift mutations in the C-terminus of the TREX1 gene.[73] RVCL-S was previously described as a number of discrete syndromes, which are now understood to have a common genetic aetiology and overlapping phenotypes: cerebroretinal vasculopathy (CRV),[74,75] hereditary endotheliopathy, retinopathy, nephropathy and stroke (HERNS)[75,76] and hereditary vascular retinopathy (HVR).[73,75,77]

RVCL-S typically presents with progressive visual impairment in the fourth to fifth decades, secondary to retinal vasculopathy, neovascularization of the optic disc, retinal haemorrhages, macular oedema, microaneurysms and capillary obliteration. Fluorescein angiograms show telangiectatic capillaries and avascular areas in the retina.[75]

Patients subsequently develop neurological features in the form of subcortical ischaemic strokes, TIAs, migraine, cognitive impairment, psychiatric abnormalities (such as personality disorders, depression and anxiety) and seizures, with a progressive decline to death around 5 to 10 years after the onset of symptoms.[75]

Some patients may also have systemic features, and this disease variant was initially described as hereditary systemic angiopathy (HSA). These patients develop systemic small vessel vasculopathy resulting in the premature infarction and necrosis of tissue, manifesting as Raynaud’s phenomenon, hepatic micronodular cirrhosis, renal dysfunction and osteonecrosis.[78,79]

In addition to typical MRI features of SVD, some of these patients were also reported to have contrast-enhancing mass lesions in the deep white matter of the cerebrum and cerebellum. Also known as pseudotumours, these masses are associated with surrounding vasogenic oedema and can regress in size over several months.[36] Although unique to RVCL-S, this feature

is not sensitive for RVCL.[80] Histopathologically these tumours are found to have areas of coagulative necrosis secondary to obliterative vasculopathy.[73] These areas are surrounded by oedematous white matter with prominent reactive astrogliosis similar to that seen in radiation necrosis. The small vessels are often occluded by fibrin thrombi, with thickened fibrotic walls and a distinctive multilamellar subendothelial basement membrane.[73]

COL4A1/A2 related arteriopathy

A spectrum of conditions with both paediatric and adult onset of neurological and systemic features has previously been described as four separate syndromes: autosomal dominant Type I Porencephaly brain SVD with haemorrhage, brain SVD with Axenfeld-Rieger Anomaly and Hereditary Angiopathy with Nephropathy, Aneurysms and muscle Cramps (HANAC syndrome). (Table 1-5) These were later found to be attributable to mutations in the same gene, COL4A1, which encodes the Type IV collagen α1 chain.[81]

COL4A1 arteriopathy was first described in infants with porencephaly, infantile hemiparesis and developmental delay, as well as intracerebral haemorrhages associated with trauma and anticoagulation.[82] A mouse model of COL4A1 mutations developed intracerebral haemorrhages during birth, which was avoided by caesarean section, lending support to the suggested association between haemorrhage and head trauma in the perinatal period.[83]

More recently, it has been recognised that COL4A1 mutation carriers may present for the first time in adulthood with subcortical intracerebral haemorrhages and ischaemic lacunar infarcts, as well as WMH on MRI and microbleeds on T2* sequences.[82] (Figure 1-2) Other neurological features include seizures, cognitive impairment progressing to dementia and visual loss. Patients with visual loss also show marked tortuosity of the medium and small retinal arterioles and venules alongside retinal ischaemic changes.[84] Systemic involvement has also been described, such as renal agenesis, muscle and cardiac symptoms as well as Raynaud's phenomenon.[81] The phenotypes seen in COL4A1 angiopathy vary widely, and the age at onset and phenotype can vary markedly between members of the same family.[81]

On neuroimaging, patients may have periventricular fluid-filled cysts involving subcortical structures in the brain (porencephaly).[82] Aneurysms may also be seen at the intracranial portions of the internal carotid artery.[82] WMH are also found in the posterior periventricular, frontal and parietal areas, sparing arcuate fibres and temporal poles. Dilated perivascular

spaces are mainly seen in the basal ganglia, while microbleeds are prominent and found both in the subcortical region and cortical-subcortical junction.[82] On post-mortem examination these patients had thickening and focal disruptions of capillary basement membranes, and dissociated smooth muscle fibres with abnormal spreading of the basement membrane at the dermo-epithelial junction and kidney tubules.[85]

Mutations in the COL4A2 gene at the same chromosomal locus have also been found to cause a similar phenotype of cerebral SVD manifesting as recurrent intracerebral haemorrhage, infantile porencephaly, congenital hemiplegia, intracranial aneurysms, nephropathy and myopathy,[86] with similar histopathological features.[87]

Table 1-5: *Clinical syndromes in COL4A1-related SVD.*

	Autosomal Dominant Type I Porencephaly	HANAC Syndrome	Brain SVD with haemorrhage	Brain SVD with Axenfeld Rieger anomaly	Walker Warburg Syndrome
Inheritance	Autosomal dominant	Autosomal dominant	Autosomal dominant	Autosomal dominant	Autosomal recessive
Clinical features	• Perinatal haemorrhage with porencephaly	• Renal – microscopic or gross haematuria, renal cysts, mild renal failure • Muscle cramps –raised CK • Visual loss – retinal artery tortuosity or haemorrhage • Raynaud's phenomenon • Supraventricular arrhythmia	• Infantile hemiparesis • Adult-onset haemorrhage • Seizures • Intellectual disability leading to dementia • Dystonia • Strokes and TIAs • Migraine	• Ocular involvement: retinal arteriolar tortuosity, cataracts, glaucoma, ocular anterior segment dysgenesis (Axenfeld Rieger anomaly) • Muscle cramps • Raynaud's phenomenon • Kidney defects • Cardiac arrhythmia	• Ocular dysgenesis • Neuronal migration defects • Congenital muscular dystrophy – hypotonia or severe myopathy leading to fatal respiratory insufficiency • Mental retardation and epilepsy
Stroke mechanism	Small vessel disease	• Small vessel disease (usually asymptomatic) • Large artery disease – unruptured intracranial aneurysms of carotid siphon (usually asymptomatic)	Small vessel disease	• Large vessel disease – aneurysms of intracranial segment of internal carotid artery • Small vessel disease	(Strokes are not a typical feature)
Radiological findings	• Large fluid-filled cavities appearing as paraventricular cysts involving subcortical structures, often asymptomatic • Diffuse white matter hyperintensities	• White matter abnormalities sparing temporal lobe and arcuate fibres • Single or multiple aneurysms localized on carotid siphon	• Diffuse white matter hyperintensities • Intracerebral haemorrhage	Diffuse white matter hyperintensities	Hydrocephalus Callosal hypogenesis
Pathological findings	• Thickening of basement membrane of capillaries	• Abnormal thickening, multilamination and focal disruption of basement membranes in kidney and skin • Dissociation of vascular smooth muscle cells			Cerebral cortical malformation – cobblestone lissencephaly, cerebellar hypoplasia, hydrocephalus, encephalocele.

These have now been found to share both genetic aetiology and overlapping phenotypes.

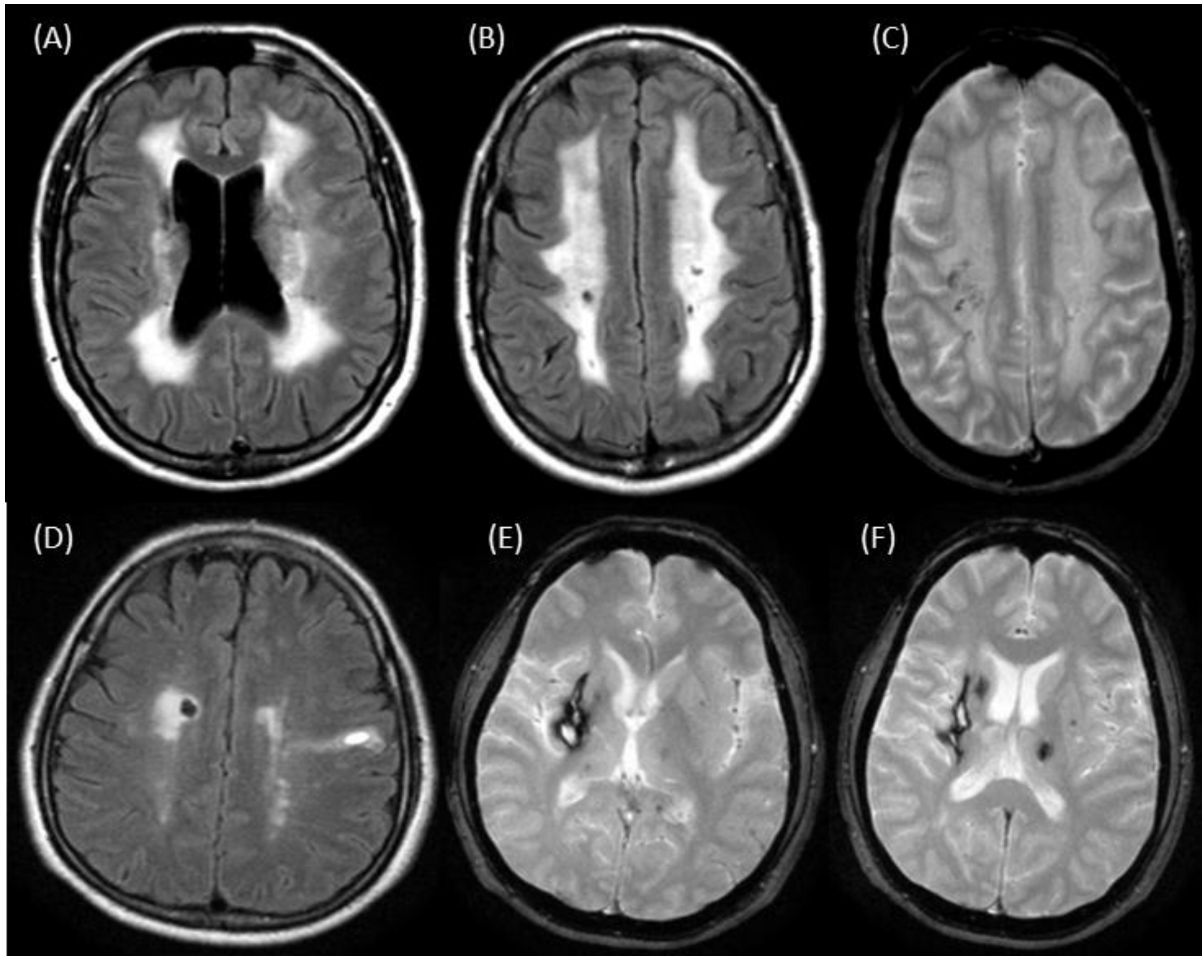


Figure 1-2: *MRI appearances in two patients (A-C and D-F) with COL4A1 mutations. T2/FLAIR sequences show confluent white matter hyperintensities (A,B,D). T2* GRE sequences show cortical-subcortical microbleeds (C,F) and a previous haemorrhage (E,F)*

FOXC1-PITX2 deletion or -duplication related SVD

The forkhead box transcription factor 1 (FOXC1) gene on chromosome 6p25 encodes a member of the winged/helix forkhead family of transcription factors. FOXC1 was initially identified as the causative gene for Axenfeld-Rieger Syndrome (ARS) and cerebellar malformations in patients with chromosome 6p25 copy number variations.[88] Patients with ARS present with ocular abnormalities such as anterior segment dysgenesis and early-onset glaucoma, as well as non-ocular features such as systemic dysmorphisms and dental and umbilical abnormalities.[88] This often overlaps with another phenotype known as the Dandy Walker malformation, characterised by cerebellar vermis hypoplasia and mega cisterna magna.[89]

In multiple case reports of patients with 6p25 deletions, individuals with ARS and other developmental abnormalities were also found to have WMH from as early as 18 months of age.[90] A meta-analysis and study of expression quantitative trait loci (eQTL) in GWAS data from the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium later demonstrated that three single nucleotide polymorphisms (SNPs) associated with WMH strongly influenced FOXC1 transcript levels, and that 18 out of 18 patients with FOXC1-related ARS also showed MRI evidence of SVD.[91] Experimental overexpression and suppression of the FOXC1 gene in zebrafish also led to cerebral haemorrhage, lending further support for the gene's independent involvement in cerebrovascular disease.[91]

FOXC1 interacts with paired-like homeodomain transcription factor 2 or pituitary homeobox 2 (PITX2) a developmental transcription factor expressed in the neural crest. Mutant forms of PITX2 also cause ARS, and ARS patients with PITX2 mutations have also shown features of SVD on brain imaging.[91]

Deficiency of ADA2 (DADA2)

Deficiency of adenosine deaminase 2 (DADA2) is an autosomal recessive disease first described in children aged 3 to 6 years old. These children presented with intermittent fevers, recurrent strokes, elevated levels of acute phase reactants, livedoid rash, hepatosplenomegaly and hypogammaglobulinaemia – a syndrome not unlike that seen in the sporadic autoimmune vasculitic disease described as Sneddon syndrome.[92]

These patients all suffered small subcortical ischaemic or haemorrhagic strokes before the age of 5 during acute episodes of inflammation, with or without fever. On MRI these strokes were found to involve the subcortical white matter, brain stem, basal ganglia and thalamus, however there was no evidence of WMH. Skin biopsies showed a predominance of neutrophils and macrophages in the interstitium with perivascular T lymphocytes and no overt vasculitis.

Whole exome sequencing on 9 patients and their unaffected parents was performed and a candidate gene was identified under a recessive model of inheritance – the Cats Eye Syndrome Chromosome Region Candidate 1 (CECR1) gene encoding adenosine deaminase 2. Patients were either homozygous or compound heterozygous for missense mutations in the

CECR1 gene, and were demonstrated to have a deficiency of the ADA2 enzyme.[92] Haematopoietic stem cell transplantation has been suggested as a possible therapeutic option, showing success in at least one patient.[93]

Cerebral microangiopathy, leukoencephalopathy with calcifications and cysts (LCC)

Cerebral microangiopathy, leukoencephalopathy with calcifications and cysts (LCC) was first described in children presenting with neurological features such as dystonia, ataxia, spasticity, hemiplegia and cognitive delay. These patients were found to have cerebral white matter disease, intracranial calcifications in the white matter, basal ganglia and thalamus, as well as cysts within the hemispheres, basal ganglia or thalamus, some of which resulted in obstructive hydrocephalus.[94] Subsequently, similar features were reported in adults.[95,96]

Histopathological features of LCC include the small, tortuous 'angiomatous' appearance of small cerebral blood vessels, together with glial changes, Rosenthal fibres, microcalcifications and microhaemorrhages.[97]

A study of 40 patients with LCC, mostly of European ancestry, found homozygous or compound heterozygous mutations in the SNORD118 gene in 33 of these pedigrees. The SNORD118 gene encodes a non-protein-coding RNA involved in ribosome biogenesis, Box C/D small nuclear RNA (snoRNA) U8.

There is significant phenotypic overlap between LCC and Coats disease, a multisystem disorder which typically presents with non-neurological features such as skin depigmentation, dysplastic nails, sparse pale hair and intrauterine growth retardation. Classic features of Coats disease are retinal telangiectasias and exudates. Patients with features of both Coats disease and LCC have been described as having 'Coats plus' syndrome or Cerebroretinal Microangiopathy with Calcifications and Cysts 1 (CRMCC1),[98] although LCC and CRMCC1 are considered distinct diseases with different genetic aetiologies,[99] with mutations in the CST Telomere Replication Complex Component 1 (CTC1) gene identified in Coats disease and Coats plus.[100]

1.3.2.2. Monogenic diseases which cause haemorrhagic stroke

COL4A1/A2-related arteriopathy (as described above), and cerebral amyloid angiopathy (CAA) may present with intracerebral haemorrhage. Haemorrhagic strokes in COL4A1/A2-related disease typically occur in the subcortical areas (deep intracerebral haemorrhage), while haemorrhages arising due to CAA tend to be in the cortex ('lobar').

Hereditary cerebral amyloid angiopathy

Several large founder families have been identified as having a hereditary form of CAA, with an autosomal dominant pattern of inheritance. The majority of cases are due to the deposition of amyloid beta (Aβ), derived from the amyloid precursor protein (APP), which is encoded on chromosome 21.[101] Aβ- and non-Aβ-type hereditary CAA may also co-occur in familial Alzheimer's disease caused by mutations in APP, PSEN1 (Presenilin1) and PSEN2 (Presenilin2) genes.[102]

Different forms of hereditary CAA have been described, and these have overlapping clinical features. A common feature of most hereditary CAA cases is cerebral SVD resulting in small cortical infarcts and recurrent lobar intracerebral haemorrhages. Patients also tend to develop cognitive decline progressing to dementia. Some patients may also experience movement disorders. The key clinical features described, and genes implicated in different pedigrees with familial CAA are described in Table 1-6.

Table 1-6: *Genes and key clinical features in hereditary cerebral amyloid angiopathy (CAA) and related amyloid disorders*

Gene	CAA type	OMIM #	Mutation(s) identified	Amyloid protein product	Clinical and pathological features	References
APP	HCHWA-Dutch	605714	c.693G>C	Amyloid beta (Aβ)	• Haemorrhagic strokes • Ischaemic strokes • Expressive dysphasia • Personality changes • Myoclonic jerks • Cognitive impairment leading to dementia • Seizures • Parenchymal plaques • Neurofibrillary degeneration	[103]
	HCHWA-Flemish		c.692C>G			[104]
	HCHWA-Italian		c.693G>A			[105]
	HCHWA-Piedmont		c.705G>C			[106]
	HCHWA-Iowa		c.694G>A			[107]
	HCHWA-Arctic		c.693A>G			[108]
CST3	HCHWA-Icelandic	105150	c.68A>T	N-terminal degradation products of Cystatin-C protein (ACys)	• Haemorrhagic strokes • Dementia • Involvement of leptomeninges, cerebral cortex, basal ganglia, brainstem, cerebellum • Peripheral tissue involvement: lymphoid organs, skin	[109]
ITM2B (BRI2)	Familial British Dementia	176500	c.799T>A	C-terminal proteolytic fragment of mutated BRI protein (ABri)	• Ischaemic and haemorrhagic strokes • Spastic tetraparesis • Dementia • Parenchymal amyloid and preamyloid plaques • Involvement of leptomeninges, cortical, subcortical regions, brainstem, cerebellum and spinal cord.	[110]
	Familial Danish Dementia	117300	c.795-796insTTTAATTTGT	C-terminal proteolytic fragment of mutated BRI protein (Adan)	• Ischaemic and haemorrhagic strokes • Cerebellar ataxia • Paranoid psychosis leading to dementia • Cataracts • Deafness • Parenchymal plaques with co-deposition of Aβ • Involvement of leptomeninges, cortical and subcortical regions, brainstem, cerebellum and spinal cord.	[111]

Gene	CAA type	OMIM #	Mutation(s) identified	Amyloid protein product	Clinical and pathological features	References
TTR	Familial transthyretin related amyloidosis	105210	c.113A>G (Hungarian) c.149T>G (Ohio)	Variants of transthyretin (ATTR)	• Sensorimotor and autonomic polyneuropathy • Intracerebral haemorrhage • Dementia • Psychosis • Visual impairment • Motor paresis or ataxia • Deposition of ATTR in vitreous, leptomeninges, meningeal blood vessels	112
Gelsolin	Familial amyloidosis – Finnish type (FAF) (gelsolin amyloidosis)	105120	Multiple mutations: c.654G>A (Finnish) c.654G>T (Danish/Czech)	Fragments of domain 2 of mutant plasma gelsolin (AGel)	• Corneal lattice dystrophy • Polyneuropathy (cranial and peripheral nerves) • Autonomic dysfunction • Dermatologic features e.g. cutis laxa • Renal involvement • AGel deposition in basement membranes, systemic organs and CNS	113
PRPN	Prion-related cerebral amyloidosis	137440	Premature stop codon, such as c.435T>G	N- and C-terminally truncated PrP^{SC} peptide	• Parenchymal perivascular PrP deposition • Neurofibrillary tangle pathology	114

1.4. The role of the extracellular matrix in the pathogenesis of SVD: insights from monogenic SVD

Recent data from genetic studies in both monogenic and 'sporadic' SVD are beginning to shed light on possible shared molecular mechanisms that may be important in the disease, including a key role for abnormalities in the extracellular matrix (ECM), or 'matrisome' proteins in the pathogenesis of the disease. The ECM is the non-cellular component of tissues made up of water, proteins and polysaccharides. It provides scaffolding for cellular components by producing fibrous proteins such as collagen, laminin and elastin, and is also biochemically active, providing signals which contribute to tissue function and homeostasis. The ECM binds to and serves as a reservoir for many other biochemically active molecules.[115] The matrisome is thus defined as the ensemble of nearly 300 proteins which make up the ECM (core matrisome), or are associated with the ECM (matrisome-associated proteins), and have been characterised by bioinformatics and proteomic methods. [116]

In the blood vessels, the ECM interacts with other vascular cells to influence vascular development and remodelling. The blood vessels have ECM components in each of its three layers. The innermost layer (tunica intima) is lined with endothelial cells on a basement membrane comprising of matrisome proteins such as type IV collagen; the tunica media contains sheets of smooth muscle cells separated by ECM, while the outermost tunica adventitia contains myofibroblast cells and is rich in type I and III collagen in addition to many other matrisome proteins.[117] A recent review by Joutel et al discusses the role of the matrisome in the small vessels, and the evidence for the alteration of matrisome function in SVD in greater depth.[118]

1.4.1. Matrisome involvement in CADASIL

The involvement of the matrisome in different monogenic forms of SVD suggests that the ECM may be the basis of shared molecular pathways in SVD. This has been illustrated in CADASIL, where the basis of ECM involvement has now been characterised at several stages - from histopathological studies in post-mortem analysis,[119] to a direct link to cerebral vasoreactivity in animal models of CADASIL. [120–122]

Post-mortem studies in patients with CADASIL have shown a possible aggregation cascade of matrisome proteins. The basis of this arose from the fact that a pathognomonic feature of CADASIL is the deposition of granular osmiophilic material (GOM) in the extracellular space of the small blood vessels systemically,[123] and the accumulation of deposits of the NOTCH3 ectodomain (NOTCH3^{ECD}) cleaved from the mutant NOTCH3 receptor.[124] In transgenic mice expressing the human NOTCH3 p.R90C mutation, NOTCH3^{ECD} accumulation and GOM deposits are often the earliest pathological features of the disease. [125] This is followed by a potassium channelopathy which precedes and results in the onset of impaired cerebral vasoreactivity,[126] eventually leading to the development of white matter lesions. [127]

Studies in post-mortem specimens from CADASIL patients and transgenic CADASIL mouse models suggest that increased levels of NOTCH3^{ECD} promote the formation of disulphide cross-linked aggregates in a cascade. These aggregates sequester key matrisome proteins, which have roles in maintaining the integrity and function of the ECM, in the walls of the blood vessels.[119,128] A summary of these proteins and their functions is provided in Table 1-7.

This protein aggregation cascade demonstrated by proteomic studies shows parallels with the progression of features in animal models of CADASIL, suggesting that each protein's involvement may contribute to different features in the disease pathway. Decreased baseline cerebral blood flow and cerebrovascular reactivity have been demonstrated in CADASIL patients, and the latter has been shown to be associated with the progression of white matter lesions.[129] A transgenic mouse model of CADASIL has recapitulated these features, showing dysfunctional cerebral vasoreactivity early in disease. This was characterised by the impaired cerebral blood flow autoregulation in response to vasodilator stimuli (hypercapnia and acetazolamide), and increased vessel resistance in the context of hypertension. [130]

***Table 1-7**: Matrisome proteins found to co-aggregate with NOTCH3 ECD.*

Matrisome protein	Function in ECM	Involvement in CADASIL / CARASIL	Models studied
Thrombospondin-2 (TSP2)	• Interacts with NOTCH3[131] • Regulates ECM assembly processes, such as collagen fibrillogenesis • Regulates angiogenesis[132,133]	• NOTCH3 ECD deposits found to co-aggregate with thrombospondin-2.[134]	Post-mortem human CADASIL specimens
Latent TGFβ-binding protein (LTBP-1)	• TGFβ is secreted as an inactive complex with LTBP-1 and latency associated peptide (LAP) • LTBP-1 regulates bioavailability of active TGFβ in the ECM[135]	• NOTCH3 ECD deposits found to co-aggregate with LTBP-1.[128] • CARASIL mutations preclude physiological cleavage of LTBP-1 by HtrA1[136]	Mouse brain tissue, embryonic and patient skin fibroblasts
Tissue inhibitor of metalloproteinase 3 (TIMP3)	• Regulatory function in ECM remodelling – inhibits a disintegrin and metalloproteinase 17 (ADAM17), a metalloprotease which degrades ECM. [137]	• NOTCH3 ECD forms complexes with TIMP3[119] • Increased TIMP3 activity contributes to vessel fibrosis, dysfunctional cerebral blood flow and myogenic responses to changes in neural activity, but is not associated with white matter lesion load[121]	Post-mortem human CADASIL specimens and transgenic mouse models
Vitronectin	• Glycoprotein in blood plasma and ECM • Roles in cell attachment, aggregation, atherosclerosis and thrombus formation[138]	• NOTCH3 ECD aggregation promotes the sequestration of TIMP3, which then promotes the co-aggregation of vitronectin.[119] • Reduced vitronectin levels associated with lower white matter burden in mouse model but not cerebral blood flow or GOM load[122]	Post-mortem human specimens and transgenic mouse models

The impaired cerebral vasoreactivity in a transgenic CADASIL mouse model was later suggested to be due to an increase in the number of voltage-gated potassium channels in the membranes of the smooth muscle cells. [126] These channels oppose depolarisation due to pressure, and downregulation of these channels restores normal myogenic responses to pressure.[126]

In biochemical and proteomic studies of cerebral vessels from CADASIL patients, $NOTCH3^{ECD}$ aggregation was found to induce the co-aggregation of Tissue inhibitor of metalloproteinase 3 (TIMP3), which then promotes the sequestration of another matrisome protein, vitronectin, in these aggregates.

A potential mechanistic link between increased TIMP3 activity and impaired cerebral blood flow regulation has recently been demonstrated. Increased TIMP3 expression in transgenic mice was shown to promote the upregulation of potassium channel current density in the cerebral arterial myocytes, and thus the reduction of myogenic tone and cerebral autoregulation.[121,122] This process is thought to be mediated by TIMP3/ a disintegrin and metalloproteinase 17 (ADAM17) interactions.

ADAM proteases cleave off the extracellular domains in the activation of membrane-bound proteins. In particular, ligands of the EGFR family, such as heparin-binding EGF-like growth factor are substrates of ADAM17.[139] The ADAM17/Heparin-binding EGF-like growth factor (HB-EGF)/EGFR (ErbB1/ErbB4) signalling axis regulates cerebral arterial tone and cerebral blood flow.[140] TIMP3 inhibits this signalling axis, and restoration of this axis with the delivery of exogenous ADAM17 or HB-EGF restores cerebral blood flow autoregulation in transgenic mice.[122]

TIMP3 is associated with potassium channelopathy and impaired cerebrovascular reactivity but not white matter lesion load, while the subsequent involvement of vitronectin is associated with the presence of white matter lesions but not cerebral vasoreactivity impairment.[119] The stepwise involvement of each protein in the cascade thus shows direct parallels with each stage of disease progression in the animal model. (**Figure 1-3**)

The cascade of sequential recruitment and aggregation of matrisome proteins triggered by an altered $NOTCH3^{ECD}$ is also reminiscent of the 'prion hypothesis' in other neurodegenerative diseases such as Alzheimer's Disease and Parkinson's Disease, where a misfolded protein

molecule acts as a 'seed' and triggers further misfolding and protein aggregation.[141] In these diseases, proteins such as Aβ, tau and α-synuclein adopt β-sheet-rich conformations and self-propagate.[142] Although mutant proteins in the aggregatory process in CADASIL may not necessarily act in a prion-like manner – but instead promote the aggregation of different proteins - the similarities between these processes may eventually point us toward common targetable pathways.

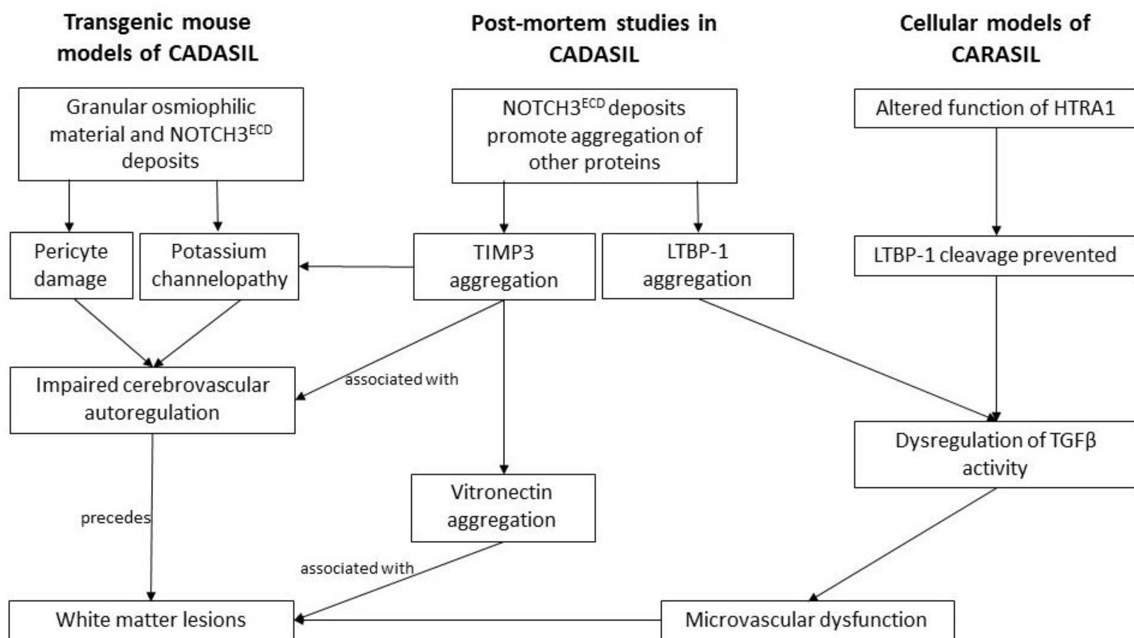


***Figure 1-3**: The involvement of matrisome proteins in the pathogenesis of CADASIL and CARASIL. Vitronectin and TIMP3 may serve as molecular correlates of clinical features and terminal pathways in the disease. Vitronectin levels are associated with white matter lesion load, while TIMP3 levels are associated with cerebral vasoreactivity in a transgenic CADASIL mouse model. LTBP-1, which co-aggregates with NOTCH3^{ECD} in CADASIL, has also been identified as the proteolytic target of HtrA1 protease, the enzyme altered as a result of CARASIL mutations.*

1.4.2. Convergence of disease mechanisms in CADASIL and CARASIL: relevance to sporadic SVD

The molecular pathways in CADASIL may be particularly relevant to our understanding of the pathogenesis of sporadic SVD. Evidence from both CADASIL and population-based genetic studies suggest that the same pathways may also contribute to sporadic disease.

While GOM deposits are pathognomonic of CADASIL, other histopathological features of CADASIL recapitulate those seen in sporadic disease. These include the fibrosis of the adventitia, and the loss of endothelial and smooth muscle cells of the perforating arteries. Similar features are also seen in post-mortem studies of cerebral vasculature in CARASIL, with fibrous intimal proliferation, hyaline degeneration of the media, loss of arterial smooth muscle cells and splitting of the internal elastic lamina contributing to the narrowing of the vessel lumen. [70,71]

Genetic studies in the population also suggest the possible involvement of monogenic disease genes in the pathogenesis of sporadic SVD. In a study of 888 population-based stroke- and dementia-free individuals in the Austrian Stroke Prevention Study, the association between common single nucleotide polymorphisms (SNPs) in the NOTCH3 gene region and white matter hyperintensities and lacunes was investigated. Four common variants, rs1043994, rs10404382, rs10423702 and rs1043997, which are in strong linkage disequilibrium, were found to be significantly associated with both the presence and progression of WMH, with this effect only being present in hypertensives. This suggests that the minor alterations in Notch3 receptor function may act together with, or augment the effects of hypertension to cause this association. These results were replicated in a sample of 8545 individuals from the CHARGE Consortium.[143] However, the association with WMH was not replicated in meta-analyses of GWAS data sets from ischaemic stroke cohorts in 3670 cases and 7397 controls, and no association was found between NOTCH3 SNPs and lacunar stroke or with WMH in stroke patients.[144] Similarly, the largest trans-ancestral meta-analysis to date, of 67,162 stroke cases and 454,450 controls, did not show any association between NOTCH3 and WMH.[145] Evidence for the involvement of monogenic disease genes in sporadic SVD is summarised in Table 1-8.

Table 1-8: *Review of studies of the association between common or rare variants in Mendelian disease genes and sporadic SVD or its associated markers*

Gene	Phenotype studied	Approach	No. of patients/controls	Variants identified*	Strength of evidence	Ref
NOTCH3 **HTRA1** **TREX1** **GLA** **COL4A1/A2** **CECR1** **APP**	Small vessel stroke (ischaemic and haemorrhagic)	Exome (n=368) and targeted gene (n=2599) gene sequencing	1251 small vessel stroke cases <70 years 1716 controls	3 ischaemic and 3 haemorrhagic patients had CADASIL mutations 1 small vessel ischaemic stroke patient had a Fabry mutation No significant association between rare variants and stroke	• Multiple genes evaluated • Study powered only to detect associations with mutations of low penetrance	[146]
NOTCH3	WMH in community-based population	Direct sequencing of all 33 exons, promoter and 3'UTR of NOTCH3	Sequenced: 195 community-based Caucasians, 82 controls with no WMH Genotyped: 888 participants from Austrian Stroke Prevention Study	4 SNPs associated with WMH presence and progression in hypertensives • rs1043994(i) • rs10404382(i) • rs10423702(i) • rs1043997(e)	• Only candidate gene studied • Replication of rs10404382 in GWAS data from hypertensive stroke-free elderly individuals in CHARGE consortium (n=8545)	[143]
	Ischaemic stroke	Direct sequencing of all 33 exons	269 Caucasians with ischaemic stroke, 95 controls	1 SNP associated with ischaemic stroke: • rs78501403(e) 1 CADASIL mutation • p.R558C	• Only candidate gene studied • Insufficient power to demonstrate association with stroke subtypes	[147]
	Symptomatic lacunar stroke or WMH in stroke patients	Meta-analysis of GWAS data sets	1350 European patients with MRI-confirmed lacunar stroke 3670 patients with ischaemic stroke and WMH 7397 controls	No association between NOTCH3 variants and lacunar stroke or WMH volume	• Only candidate gene studied	[144]

Gene	Phenotype studied	Approach	No. of patients/controls	Variants identified*	Strength of evidence	Ref
	Leukoaraiosis (Fazekas scale 3)	Screen of exons 3-6 by PCR single-stranded conformational polymorphism analysis	218 patients with lacunar stroke (48 with leukoaraiosis)	No association identified between common polymorphisms and leukoaraiosis 1 CADASIL mutation • p.C697T	• Limited screen of NOTCH3 gene	148
	Symptomatic ischaemic cerebrovascular disease	PCR analysis of T6746C polymorphism	235 Japanese patients with CT/MRI defined ischaemic stroke/TIA (142 with lacunar stroke) 315 controls	No association found between polymorphism and cerebrovascular disease or lacunar stroke	• Only one polymorphism studied	149
	Ischaemic stroke and dementia	Novel diagnostic array for known mutations and polymorphisms in exons 3 and 4 of NOTCH3	70 patients with CT/MRI confirmed ischaemic stroke and 77 patients with dementia 117 controls	No association between known polymorphisms and stroke or dementia.	• Only 5 previously identified polymorphisms in 2 exons studied	150
NOTCH3 **GLA**	Lacunar stroke	Genome-wide genotyping and direct sequencing of GLA and NOTCH3	994 patients with lacunar stroke with age at onset ≤70 years	5 patients with CADASIL mutations • p.Arg169Cys • p.Arg207Cys • p.Arg587Cys • p.Cys1222Gly • p.Cys323Ser No classical pathogenic Fabry mutations identified • One patient had missense p.R118C mutation associated with late-onset Fabry disease.	• Both GLA and NOTCH3 genes examined	151

Gene	Phenotype studied	Approach	No. of patients/controls	Variants identified*	Strength of evidence	Ref
GLA	Cryptogenic ischaemic or haemorrhagic strokes, periventricular WMH	Sequencing of gene in females. Plasma α-galactosidase activity in males, followed by sequencing if enzyme activity if low	721 German patients with cryptogenic ischaemic stroke aged 18 to 55	Fabry mutations identified in 21 males (4.9%) and 7 females (2.4%) • 11 males, 3 females with periventricular WMH • 18 males, 7 females with ischaemic stroke • 3 males, 0 females with haemorrhagic stroke	• Only candidate gene studied • Aetiology of stroke uncertain	152
	Ischaemic and haemorrhagic stroke	Sequencing of gene and plasma α-galactosidase activity	493 Portuguese patients with young onset stroke age 18 to 55	12 patients (2.4%) had GLA mutations • 9 ischaemic stroke, including 2 SVD • 2 ICH • 1 cerebral venous thrombosis	• Only candidate gene studied • Aetiology of stroke uncertain	153
	Stroke and white matter lesions	Sequencing of gene in females. Plasma α-galactosidase activity in males.	1000 patients with ischaemic stroke, TIA, ICH, white matter lesions, vertebrobasilar dolichoectasia	19 males (3.5%) and 8 females (1.8%) had missense mutations.	• Only candidate gene studied • All stroke subtypes, including cryptogenic, included.	154
	Ischaemic stroke	Sequencing of gene in females. Plasma α-galactosidase activity in males, followed by sequencing if enzyme activity if low	475 Japanese males with ischaemic strokes	10 males had decreased α-Gal A activity – all had multiple small vessel infarcts 5 had p.Glu66Gln mutation	• Only candidate gene studied	155

Gene	Phenotype studied	Approach	No. of patients/controls	Variants identified*	Strength of evidence	Ref
COL4A1	Presumed sporadic ICH	Direct sequencing of coding regions of COL4A1 and flanking intronic regions	48 patients with presumed hypertension-related deep ICH 48 patients with probably CAA-related ICH 145 controls	2 rare variants associated with ICH: • c.C1055T (p.Pro352Leu) (e) • c.C1612G (p.Arg538Gly) (e)	• Only candidate gene studied • Only rare variants studied – not common variants • Cellular assay of variants demonstrated impaired secretion of α1 chain	[156]
COL4A1 **COL4A2**	• ICH (deep/lobar) • Ischaemic stroke (cardioembolic, LVD, SVD) • WMH (ischaemic stroke and population based)	Meta-analysis of GWAS data sets	1545 patients with ICH, 1485 controls 1854 patients with lacunar stroke, 2733 with ischaemic stroke and WMH, 9361 controls	3 COL4A2 SNPs associated with deep ICH • rs9521732 (i) • rs9521733 (i) • rs9515199 (i) SNPs did not reach significance threshold for association with lacunar stroke or with WMH volume	• Both COL4A1/A2 studied • No significant eQTLs with 3 SNPs or 5 other SNPs in high linkage disequilibrium with these three. • SNPs located in regions with possible regulatory roles	[157]
COL4A2	WMH in stroke patients	Meta-analysis of GWAS data	3670 stroke patients	4 novel SNPs associated with WMH, one of which is in COL4A2 • rs9515201 (i)	• SNP in strong LD with those previously identified (above) • SNP may have regulatory function	[158]
COL4A2	Presumed sporadic ICH	Direct sequencing of coding regions of COL4A2 and flanking intronic regions	48 patients with presumed hypertension-related deep ICH 48 patients with probable CAA-related ICH 145 controls	3 rare coding variants associated with ICH • c.3368A>G (p.Glu1123Gly) (e) • c.3448C>A (p.Gln1150Lys) (e) • c. 5068G>A (p.Ala1690Thr) (e)	• Only candidate gene studied • Only rare variants studied, not common variants • Cellular assay demonstrated impaired secretion of a1 and a2 chains	[159]

Gene	Phenotype studied	Approach	No. of patients/controls	Variants identified*	Strength of evidence	Ref
TREX1	Severe WMH and recurrent strokes	Sequencing of TREX1 gene	100 patients with severe WMH and young-onset strokes suspicious of CADASIL	2 patients had missense mutations • p.Tyr305Cys • p.Arg169His	• Only candidate gene studied	[160]
FOXC1 **PITX2**	WMH in community-based dementia- and stroke-free populations	Meta-analysis of GWAS data and study of patients with FOXC1-related Axenfeld-Rieger Syndrome (ARS)	9361 patients in GWAS 18 patients with FOXC1-related ARS	10 SNPs located in GMDS gene adjacent to FOXC1 • rs12206258 (i) • rs12203614 (i) • rs12199578 (i) • rs12193217 (i) • rs10458129 (i) • rs12206340 (i) • rs12189662 (i) • rs6936881 (i) • rs7765344 (i) • rs7765461 (i) 9 SNPs near PITX2 • rs2129979 (ds) • rs11931959 (ds) • rs13121924 (ds) • rs3866831 (ig) • rs6533531 (ig) • rs13141190 (ig) • rs6533530 (ig) • rs7697491 (ig) • rs723363 (ig)	• Only candidate gene region studied • 3 SNPs strongly modify FOXC1 transcript levels: rs12206258, rs6936881, rs7765344 • 18 of 18 patients with FOXC1-related ARS have features of SVD	[91]

Gene	Phenotype studied	Approach	No. of patients/controls	Variants identified*	Strength of evidence	Ref
FOXF2 (on 6p25 near FOXC1)	Stroke and stroke subtypes (ischaemic, cardioembolic, non-cardioembolic) WMH in stroke-free adults	Meta-analysis of GWAS data	84961 Europeans: 4348 with stroke, of which 1770 were non-cardioembolic ischaemic strokes.	• rs12204590 near FOXF2 associated with all stroke and WMH burden in stroke-free adults	• rs12204590 replicated in validation samples (stroke patients) associated with risk of all stroke • rs12200309 in complete LD with rs12204590, associated with SVD ischaemic stroke in validation samples • Region includes enhancers with 2 SNPs in high LD with rs12204590 having probable roles in regulating gene expression	[161]
CECR1	Ischaemic lacunar stroke	Examination of individual-level whole exome data	94 patients with late-onset stroke in the Siblings with Ischemic Stroke Study (SWISS) subcohort	Two brothers with late-onset small vessel ischaemic strokes heterozygous for p.Tyr453Cys mutation.	• Only candidate gene studied	[92]

Table 3.3 footnotes: *i: intronic, e: exonic, ds: downstream, ig: intergenic

1.4.3. Collagen: involvement of the most abundant matrisome protein

Collagen is the most abundant protein in the ECM, and has a characteristic triple-stranded helical structure known as tropocollagen. Tropocollagen is made up of polypeptide chains with highly conserved repetitive three-residue sequences (Gly-X-Y). As glycine is the amino acid with the smallest side chain, it allows the tight assembly of each collagen strand in a helix, with glycine forming the core of each helix. Multiple tropocollagen molecules polymerise to form collagen fibrils which provide tensile strength to tissues.[35]

Type IV collagen in the basement membrane is formed with α1 and α2 collagen chains in a 2:1 ratio. These chains, encoded by the COL4A1 and COL4A2 genes respectively, are the most abundant proteins in basement membranes and surrounding smooth muscle cells in the tunica media of blood vessels.[117] The relationship between collagen mutations and vessel fragility is well described in diseases such as osteogenesis imperfecta (COL1A1 or COL1A2 mutations) and Ehlers-Danlos syndrome (COL3A1 mutations).

While a precise molecular pathway in COL4A1 and A2-associated SVD has yet to be identified, it is likely that the impaired function of collagen in the ECM contributes to the disease process. The majority of reported mutations in COL4A1/A2-related SVD affect the glycine residue of the Gly-X-Y residues, disrupting the three-dimensional conformation of each α1 or α2 strand and thus impairing the formation of the resulting tropocollagen molecule.[162] This is thought to result in the impaired synthesis of the basement membrane and thus lead to blood vessel fragility.[83] Both human and mouse mutations in COL4A2 have been shown to cause the impaired secretion of both α1 and α2 chains, thus resulting in the retention of mutant α1 and α2 chains in the endoplasmic reticulum.[159] The accumulation of mutant α1 or α2 chains results in endoplasmic reticulum stress, which is also thought contribute to disease.[163,164]

A genotype-phenotype correlation in the subtype of COL4A1-related SVD known as HANAC syndrome has been suggested.[165] Mutations in these patients were found to aggregate in the 31-residue cyanogen bromide-derived fragment (CB3[IV]) region. This region of the cell contains integrin-binding domains, and is thought to be instrumental in integrin-mediated cell adhesion to the collagen molecules, [165] suggesting that abnormal interactions between type IV collagen and cells may result in a systemic form of the disease [81,165].

1.4.3.1. Role of monogenic disease genes in sporadic SVD

Recent evidence suggests that common variants in COL4A2 are associated with sporadic SVD and ICH. A meta-analysis of genotype data from large GWAS studies in stroke in individuals of European ancestry identified three common variants (rs9521732, rs9521733, rs9515199) in intronic regions in COL4A2 which were significantly associated with deep ICH. There was a similar trend for lacunar stroke and WMH although the associations did not achieve the stringent significance levels set to account for the multiple comparison made.[157] (Table 1-8)

A multi-ethnic genome-wide meta-analysis of dementia- and stroke-free cohorts found a SNP located in an intron of the COL4A2 gene, rs9515201 which was associated with WMH in community populations; this SNP [158] is in strong linkage disequilibrium with SNPs that were previously identified to be associated with sporadic ICH.[157] (Table 1-8) A further meta-analysis of genotype data from 19569 European stroke patients and 37853 controls suggested an association between a locus in COL4A2 and both lacunar ischaemic stroke (rs9515201)and deep ICH (rs4771674).[166] In this same study, an association was found between a locus in HTRA1 (rs79043147)and lacunar ischaemic stroke. [166]

1.5. Other possible mechanisms of disease – insights from genetics

1.5.1. Blood Brain Barrier development and integrity

The FOXC1 gene codes for the forkhead box transcription factor C1 (Foxc1), which is critical in the development of blood vessels.[167] Foxc1 originates from the neural crest and is expressed by brain pericytes, which are integral components of the BBB, and regulates vascular morphogenesis and maturation during embryological brain development. While Foxc1 deletion does not preclude angiogenesis and may not affect BBB formation and permeability, it results in altered brain pericyte and endothelial cell proliferation, impairing blood vessel stability and thus predisposing these vessels to haemorrhage. [167] Hence, although the precise mechanisms behind FOXC1-associated SVD are not known, the theoretical basis of this disease suggests an impairment of the BBB.

FOXC1 interacts with Paired-like homeodomain transcription factor 2 or Pituitary Homeobox 2 (PITX2), a developmental transcription factor expressed in the neural crest. Mutant forms

of PITX2 also cause ARS.[88] ARS patients with PITX2 mutations also had features of SVD on brain imaging.[91] The similar phenotype seen with PITX2 mutations lends further support for the involvement of the FOXC1 pathway in the development of SVD.

Studies in FOXC1 knockout models have led to speculation that matrisome proteins may mediate disease mechanisms in FOXC1-related SVD.[118] The expression of matrix metalloproteinases (MMPs), which regulate the ECM, is increased in the cornea of global and neural crest-deleted Foxc1$^{-/-}$ mice. These MMPs regulate the bioavailability of vascular endothelial growth factors sequestered in the ECM. Upregulation of MMP expression leads to disorganisation of the ECM and excessive growth of vessels in the cornea of mutant mice.[37] Suppression of Foxc1 in zebrafish also reduced expression of platelet-derived growth factor (PDGF), a matrisome-associated protein integral to the development of vasculature. Consistent with evidence in humans that alterations in Foxc1 dosage were associated with SVD, zebrafish with either Foxc1 knockdown or overexpression also exhibited cerebral haemorrhage.[91]

Adjacent to the FOXC1 gene on chromosome 6p25 is FOXF2, a gene that encodes the Foxf2 transcription factor. Foxf2 is expressed specifically in CNS pericytes and is required for pericyte differentiation and BBB development. [161] FOXF2 knockout mouse embryos develop defects in the BBB, and FOXF2 inactivation in adult mice lead to BBB breakdown, cerebral infarction and microhaemorrhage. [168]

FOXF2 mutations and copy number variations have been implicated in Anterior Segment Dysgenesis, an ocular condition which also occurs in ARS. [169] In ARS, patients with FOXC1 and FOXF2 both deleted have more extensive WMH than those with deletion of only FOXC1, suggesting that the loss of interactions between FOXC1 and FOXF2 contribute to a shared disease pathway. [25]

The same forkhead box protein loci have also been implicated in sporadic SVD. (Table 1-8) A meta-analysis of GWA data of the FOXC1 and PITX2 gene locus identified 10 WMH-associated SNPs which lie in an intron of the GDP-mannose 4,6-dehydratase gene (GMDS) adjacent to FOXC1. Three of these SNPs have effects on FOXC1 transcript levels.[91] In the PITX2 gene locus, nine SNPs were found to be significantly associated with WMH.[91] Another recent large-scale GWAS meta-analysis in ischaemic stroke identified a novel locus close to FOXF2.[161] The same

SNP was also associated with WMH, suggesting that the mechanism by which disease risk is conferred is though SVD. [161] These converging results from sporadic and monogenic disease lend support to the possible roles of the FOXC1-PITX2-FOXF2 interactions and their roles in maintaining BBB integrity via proteins in the matrisome.

Evidence for the involvement of the BBB in SVD is also seen in CADASIL. Histopathological examinations of cerebral vessels from CADASIL patients and transgenic mouse models of CADASIL have shown damaged pericytes.[170–172] The TGFβ pathway, which has been implicated in both CADASIL and CARASIL (Figure 1-3), may provide a possible explanation for this process. In vitro studies of TGFβ1, which is the most extensively studied form of TGFβ, have shown that TGFβ1 reduces pericyte proliferation and elevates the expression of MMPs and other proinflammatory cytokines which may disrupt BBB function.[173]

1.5.2. Impairment of DNA Damage Response

RVCL-S is an autosomal dominant form of SVD caused by frameshift mutations in the TREX1 gene.[73] TREX1 codes for the most abundant DNA exonuclease in mammals, known as DNase III or Three prime Repair Exonuclease.[73] DNase III has a role in the repair of DNA damage, and serves this function by translocating from the endoplasmic reticulum to the nucleus during oxidative DNA damage.[78] DNase III enzymatically digests cytosolic single-stranded DNA to prevent the cell from responding to immunostimulatory DNA, such as those arising from pathogenic viruses.

Dysfunctional DNase III arising from loss-of-function TREX1 mutations may thus result in the erroneous recognition and clearance of self-nucleic acids, resulting in autoimmune and inflammatory diseases such as systemic lupus erythematosus (SLE), an inherited form of SLE known as familial chilblain lupus, and an inflammatory early-onset encephalopathy known as Aicardi-Goutiéres Syndrome.[174]

Functional analyses have shown that RVCL-S-causing TREX1 mutations cause cellular mislocalization,[78] rather than a loss of enzymatic activity. The expression of DNase III with a truncated C-terminus and disrupted transmembrane domain impairs its cellular localisation in the endoplasmic reticulum.[73] The nuclear target of DNase III was recently identified as poly(ADP-ribose) polymerase-1 (PARP1), an enzyme which repairs single stranded DNA breaks through a process of base excision repair, and is integral to the cell's DNA damage

response.[175] The disrupted localisation of DNase III may theoretically have a toxic gain-of-function effect, or attenuate normal DNA damage responses, however the precise mechanisms underlying RVCL-S remain to be characterised.

1.6. Further insights from GWAS: distinct mechanisms across the SVD spectrum

Significant progress has been made in our understanding of the genetic architecture of sporadic SVD following recent GWA studies. Rather than SVD pathology being homogeneous across all individuals, it is possible that there are distinct pathological pathways leading to SVD in different groups. In a population of stroke patients,[176] genetic factors underlying WMH were distinct in hypertensive individuals compared to non-hypertensives, with only a very low correlation between the genetic components (r^2=0.15).[176] A recent study using GWAS data from a young onset population with MRI-confirmed lacunar stroke showed that genetic factors are an important contributor to risk in this population,[29] with higher heritability seen in this group than in previous populations where most phenotyping was done using CT.[177] Much of the heritability arose from regions of the genome influencing expression of genes, or in DNase I Hypersensitivity sites, suggesting that the genetic risk of sporadic SVD is conferred through subtle changes to gene expression and regulation. In addition, when categorising lacunar stroke cases into those with extensive WMH and those without, analysis suggested that disease in the two groups was attributable to different rare genetic variants not thought to be in linkage disequilibrium, again highlighting that multiple distinct pathways lead to different manifestations SVD in different groups of patients.[29]

GWAS studies in MRI-confirmed WMH have identified 13 loci strongly associated with the trait, as summarised in Table 1-9.[158,178,179] Four of the loci arise from an extended region containing NEURL1, PDCD11, and SH3PXD2A. NEURL, a highly conserved E3 ubiquitin ligase. is particularly notable as it inhibits the Notch pathway through decreasing expression of the Notch ligand, JAG1.[180,181] Of interest, five of these loci are found in genes which have been implicated in malignant brain tumours of the white matter involving glial cells, suggesting the importance of these cells in pathogenesis of SVD. As well as influencing WMH in both

community and stroke patient populations,[178] 12 of the identified WMH loci also confer risk of lacunar stroke,[182] and one of the loci, in 1q22, is also associated with ICH.[27]

Taken together, GWAS to date emphasise that there are likely multiple distinct pathways leading to SVD. Some of these pathways are shared across different manifestations of the disease, but some are likely to be specific to disease groups; and in some cases are likely to act through interactions with environmental or non-genetic risk factors such as hypertension.

Table 1-9: *SNPs associated with WMH in community and stroke populations*

SNP	Chromosome	Nearest Gene	Phenotype Association	References
rs7214628	17	TRIM65	WMH in community and stroke populations	[158,178]
rs72848980	10	NEURL	WMH in community populations	[158,178]
rs7894407	10	PDCD11	WMH in community populations	[158,178]
rs12357919	10	SH3PXD2A	WMH in community populations	[158,178]
rs7909791	10	SH3PXD2A	WMH in community populations	[158,178]
rs78857879	2	EFEMP1	WMH in community and stroke populations	[158,178]
rs2984613	1	PMF1-BGLAP	WMH in community populations, intracerebral haemorrhage	[27,158,178]
rs11679640	2	HAAO	WMH in community populations	[158,178]
rs72934505	2	NBEAL1	WMH in community and stroke populations	[158,178]
rs941898	14	EVL	WMH in community and stroke populations	[158,178]
rs962888	17	C1QL1	WMH in community and stroke populations	[158,178]
rs9515201	13	COL4A2	WMH in community and stroke populations, intracerebral haemorrhage	[157,158,178]
rs12445022	16	ZCCHC14	Small vessel stroke, WMH in stroke populations	[179]

1.7. Concluding remarks

1.7.1. Reaching a genetic diagnosis in familial SVD

Cerebral small vessel disease is a complex disease that most commonly occurs due to interactions between genetic and environmental factors in the elderly, hypertensive population. SVD is the subtype of stroke most likely to present as a familial disorder, the most common of which is CADASIL. In recent years, mutations in several other genes such as HTRA1, COL4A1, COL4A2, FOXC1 and TREX1 have been identified in other families with CADASIL-like or other familial SVD syndromes.

Some of these diseases have distinctive features which may point towards a specific genetic diagnosis. For example, a patient with a clinical phenotype compatible with CADASIL may have characteristic anterior temporal pole involvement on MRI. However, many of the other monogenic forms of SVD such as CARASIL do not have characteristic MRI features. The decision of which candidate genes to test for mutations in is also seldom straightforward, due to the overlapping phenotypes between each disorder and the heterogeneity of phenotypes between families. Mutation screening tests previously looked at one gene at a time, with the testing of multiple genes one after another being prohibitively expensive. Testing for some genes may not be available under conventional Sanger sequencing, and there remain many families with early-onset SVD with non-specific clinical or radiological features, or where no mutations in known genes have been identified. The increasing use of next generation sequencing (NGS) technology in clinical diagnosis may serve to identify novel genetic causes, as described in Chapters 4 and 5.

1.7.2. Pathogenesis of SVD: what we understand from genetic studies

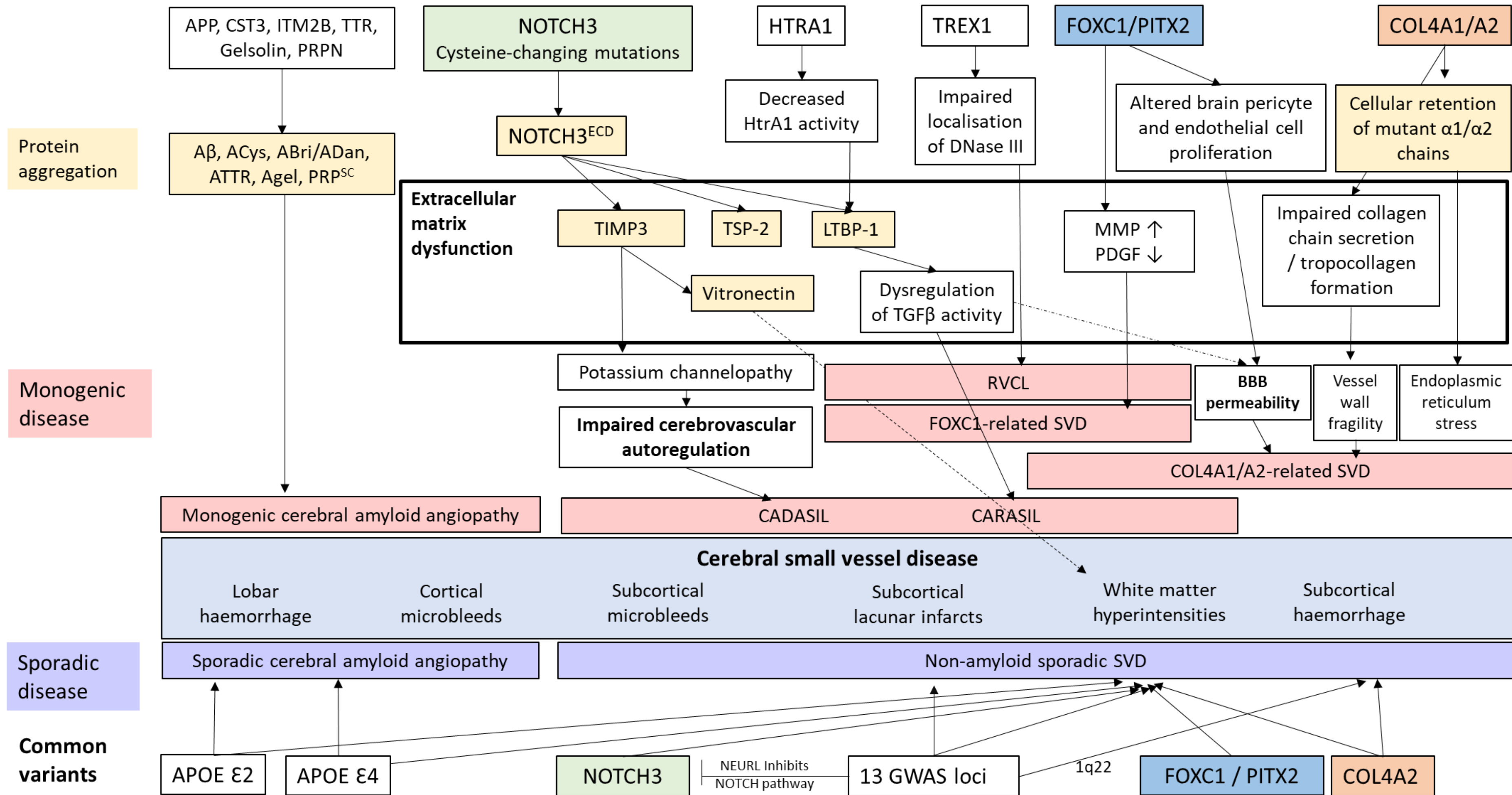
Studies in both monogenic forms of SVD and 'sporadic' SVD are now beginning to fill in the blank edges in the map of the disease processes in SVD. (Figure 1-4) Shared pathways affecting the integrity and function of the ECM appear to play an integral role in the development of this disease. It is likely that there are multiple shared pathways, each being involved to varying degrees in different manifestations or subtypes of SVD. These genetic mechanisms, as well as their interactions with environmental factors, may provide explanations as to why different patients in the sporadic disease population exhibit each feature of SVD to different extents.

In addition, there is now accumulating evidence of a protein aggregation cascade seen in CADASIL, suggesting that the convergence of pathways may extend beyond SVD, and that there may be a convergence of pathogenic pathways as seen in neurodegenerative diseases in general.

Elucidating the disease pathways in SVD may allow us to identify therapeutic targets. An example is seen in a monogenic large vessel vasculopathy, Marfan syndrome, which can be caused by mutations in the fibrillin-1 gene. Fibrillin-1 is a key ECM component and binds to the latent TGFβ complex. Antagonists of TGFβ signalling such as losartan have now been shown to reduce the development of aortic aneurysms in a mouse model of Marfan syndrome.[183] The involvement of the TGFβ pathway in SVD may lead us towards the use of TGFβ antagonists to halt disease progression, while the protein aggregation cascade suggests the potential utility of therapeutic agents currently being developed in the treatment of other neurodegenerative diseases with similar mechanisms.

Further genetic studies in SVD will likely provide more conclusive evidence of the overlap of molecular pathways involved in both monogenic and sporadic disease. Although an improved understanding of the processes in monogenic or 'sporadic' disease may aid the development of treatment options for each specific disease, it is likely that the distinction between the two is blurred, and that convergent pathways between the two will eventually act as therapeutic targets.

Figure 1-4 *:The convergence of disease pathways, particularly in the extracellular matrix (ECM), in the mechanisms underlying monogenic SVD. These pathways may also be biological correlates for clinical and other disease features identified in post-mortem and transgenic animal studies, as seen in the example of CADASIL. Pathological and clinical features are also shared between monogenic and sporadic disease, lending support to the possibility of these shared pathways also being involved in sporadic SVD.*



Chapter 2: Methodology

Outline

2.1 Background and introduction to genetic analysis in disease

- 2.1.1 Linkage analysis
- 2.1.2 Genome-wide association studies (GWAS)
- 2.1.3 High-throughput sequencing (HTS) in the context of monogenic SVD
- 2.1.4 Whole genome sequencing
- 2.1.5 Targeted gene sequencing
- 2.1.6 High throughput sequencing: quality control and analysis pipeline
- 2.1.7 Current challenges in using HTS in the clinic
- 2.1.8 Control databases
- 2.1.9 Predicting pathogenicity at a gene or protein level

2.2 Patient cohorts studied

- 2.2.1 CADASIL
- 2.2.2 Suspected familial SVD
- 2.2.3 Patients with early-onset, presumed sporadic SVD

2.3 Conclusions

List of tables

Table 2-1 Comparison of high-throughput sequencing techniques

Table 2-2 Rare diseases in the BRIDGE Whole Genome Sequencing study

Table 2-3 Sites and total recruitment over a 24-month period for the BRIDGE-SVD Whole Genome Sequencing study

Table 2-4 Selection criteria for patients with suspected monogenic SVD

List of figures

Figure 2-1 Bridge amplification in the process of high throughput sequencing by Illumina

Figure 2-2 Generic processing pipeline for high throughput sequencing data.

2. Methodology

2.1 Background and introduction to genetic analysis in disease

High throughput sequencing (HTS) platforms have become widely used since 2005, with the rapidly falling cost of DNA sequencing relative to Sanger sequencing.[184] Prior to the availability of such technology, progress was made through the use of linkage analysis followed by Sanger sequencing of specific haplotypes. The introduction of genome-wide arrays revolutionised the approach to genetic studies, shifting the strategy from being primarily pedigree-based to population based. With the arrival of HTS, several of these techniques have now been combined, as detailed in this chapter.

2.1.1 Linkage analysis

Genetic linkage refers to the phenomenon whereby two alleles at different loci are inherited together from parents to offspring more often than expected by chance. Linkage analysis is a tool that allows the detection of the loci of disease genes, by identifying regions of the chromosomes that co-segregate with the disease, and has its basis in the recombination between the two loci.[185] After an initial disease-causing mutation occurs, recombination events or 'crossing over' occurs through multiple generations, with marker loci occurring within shorter and shorter distances of the causative variant.

Linkage analysis was formerly the mainstay of genetic studies, and was used to identify loci of interest in large pedigrees. These analyses would typically identify a number of genetic regions that could be explored in greater depth using Sanger sequencing. Some of the successes of linkage analysis include the identification of the Cystic Fibrosis Transmembrane conductance Regulator (CFTR) gene in Cystic Fibrosis[186], and the huntingtin gene in Huntington's disease.[187] Linkage analyses were thus useful in the testing hypotheses in patients thought to have a 'one gene, one disease' inheritance pattern, but less helpful in the study of complex, 'common disease, common variant' hypotheses.

2.1.2 Genome-wide association studies (GWAS)

With the availability of genome-wide association (GWA) arrays, genetic linkage took a back seat and population-based studies were brought to the fore. The genetic basis of complex,

non-Mendelian diseases continues to be explored using association studies between large numbers of unrelated individuals. GWAS associations were more powerful in identifying genotype-phenotype associations, and circumvented the limitation of small families with limited numbers of generations available for study. The aim of GWAS is primarily to identify the genetic basis of traits or disease susceptibility using common variants (SNPs) in the population.

GWAS utilises common variations across the human genome, known as single nucleotide polymorphisms or SNPs, to identify genetic risk factors for disease. SNPs are single base-pair variations in the sequence that occur at high frequency in the population, and serve as markers throughout the genome. The underlying assumption is that each SNP has minimal impact on function and disease, either through its role in alteration of protein availability or function via amino acid changes, mRNA transcript stability or transcription. Hence, a single SNP is unlikely to be associated with disease on its own, but disease risk is likely to be shared across multiple genetic risk factors interacting with the environment.[188]

2.1.3 High throughput sequencing in the context of monogenic SVD

High throughput sequencing (HTS), or next generation sequencing, refers to highly parallel DNA sequencing technologies that can produce hundreds of thousands to millions of reads (of length 25 – 500bp) for a low cost, and with a rapid turnover. These technologies can deliver complete genome sequences (whole genome sequencing, WGS, as utilised in Chapter 5), or the coding sequences only (whole exome sequencing, WES), which covers about 1% or 30 million base pairs of the genome. Other forms of HTS include sub-exome panels of specific genes (as described in Chapter 4), RNA sequencing and methylation sequencing. A comparison of the strengths and weaknesses of different HTS techniques is shown in **Table 2-1**.

__Table 2-1__: Comparison of high-throughput sequencing techniques

Techniques	Description	Advantages	Disadvantages
Whole genome sequencing (WGS) [189]	Sequencing of whole genome using fragmentation, followed by paired-end sequencing and mapping to a reference genome (or de novo alignment if reference unavailable)	• Can identify non-coding variants, translocations, copy number variants • Can provide phased information • Moderate depth of sequencing	• No information about active transcription • Significant cost with storage and interpretation • Challenges with assembly and interpretation
Whole exome sequencing (WES)[190]	Targeted sequencing of exome (protein-coding regions of genome) using probe-based capture technology	• Moderate-high depth of sequencing • Cheaper than WGS • Easier to analyse and store • Single universal panel used for capture	• Does not cover other important non-coding sequences beyond exome • No information on transcriptional changes
Sub-exome/ small gene panels[190]	Targeted sequencing of selected regions using probe-based capture technology	• High depth of sequencing • Less to interpret and analyse	• Can be as expensive as WES • Need to design specific panels for each region covered instead of using single universal panel • No information on CNVs outside what is captured • Can only detect what is on panel
RNA Sequencing[191]	Sequencing of all transcribed sequences (transcriptome) or only polyadenylated transcript	• Can complement WGS or WES to provide transcriptional level data • Can estimate transcript abundance • Can detect rare transcripts, isoforms and mutations in abundant transcripts • Not limited to transcripts with corresponding genomic sequence e.g. viral transcripts	• Challenging to produce cDNA libraries • More sequencing needed for low-abundance transcripts • Difficult to integrate with WGS/WES

Techniques	Description	Advantages	Disadvantages
Methylation sequencing[192]	Quantifying of DNA methylation patterns and differences in methylation	• Can provide epigenetic level information • Complementary with RNAseq, WGS, WES	• Role of methylation poorly understood • Low quantitative accuracy

These high throughput techniques were first used to provide variant-level information on susceptibility loci identified by GWAS, by providing a means of targeted resequencing of the region. As common variants could only explain part of the genetic architecture of common diseases, it was thought that rare variants may account for much of this 'missing heritability'. These rare variants could be identified using HTS techniques.

HTS has also been used at the exome- or genome-wide level to identify disease-causing variants in patients suspected of having a rare inherited disorder. These studies rely on comparisons between unaffected and affected related individuals, or between affected unrelated individuals, to discover novel alleles.[184] Genetic linkage analysis performed with a view to identifying individuals in a pedigree suitable for whole genome sequencing, or performed using whole genome sequencing data, can provide statistical evidence for the role of a rare variant in a rare disease. It can also serve as a form of verifying purported relationships in the pedigree. However, the union of these two techniques still does not overcome the possibility that a variant in linkage disequilibrium is falsely associated with the disease.[193]

HTS is thus likely to improve our understanding of cerebral small vessel disease (SVD) as a growing number of novel genes are discovered in the context of disease. The impact of HTS in the diagnosis of rare neurological disease has already been seen, with a growing number of monogenic causes being described with the use of this technology.[194]

With the rapidly declining cost of sequencing, HTS is now no longer predominantly a research tool, but is also being used routinely in diagnostic gene panels in the clinic in place of conventional Sanger sequencing,[195] and already has a growing impact on the diagnosis of rare disease.[194] Patients may now be routinely tested on targeted gene panels instead of being tested for one gene after another, which is costly and time-consuming.

2.1.4 Whole genome sequencing

In Chapter 5 we describe our study on suspected familial SVD where no known genetic diagnosis has been identified. We performed whole genome sequencing in patients and their relatives.

In this process genomic DNA is fragmented into lengths of 150bp, and adaptors are ligated to the ends of each fragment. These single stranded fragments are then added to a flow cell, where they bind randomly on the surface containing bound primers. The free ends of the fragments (containing adaptors) bind to these primers, causing the fragment to arch over to form a bridge, as shown in Figure 2-1. Nucleotides are incorporated extending from the primer, using the fragment as a template. This forms a double-stranded bridge. These are then denatured, leaving twice the number of single-stranded templates anchored to the substrate. The process is repeated until millions of dense clusters are formed. Up to a thousand identical copes of a single template molecule are created in a cluster.

Sequencing is performed by adding a single labelled deoxynucleoside triphosphate (dNTP), so that it is incorporated using DNA polymerase. This dNTP terminates polymerization, so that it can be imaged using a laser, where emitted fluorescence from each cluster is captured to identify each base in turn. The label is then enzymatically cleaved to allow incorporation of the next dNTP. This cycle is repeated until the sequence of bases in a fragment are all determined. The data is subject to quality control at multiple stages, aligned and compared to a reference genome (human genome build GRCh37), and differences (‘variants’) are called based on this reference genome.[196] In our study using WGS by Illumina, a coverage depth of at least 15X was achieved for at least 95% of the genome.

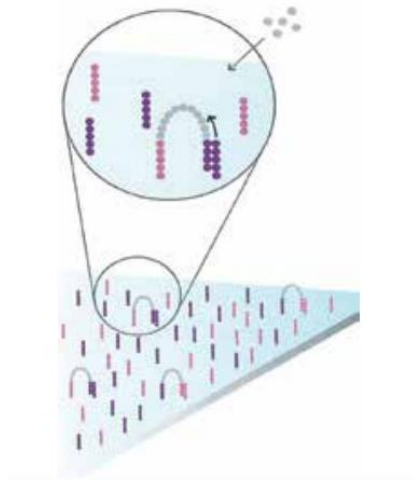


***Figure 2-1**: Bridge amplification in the process of high throughput sequencing by Illumina. Unlabelled nucleotides and enzymes are added to initiate amplification. Image from Illumina, Inc.*[196]

2.1.5 Targeted gene sequencing

Targeted sequencing of genes of interest, as used in our study of presumed-sporadic early-onset SVD cases in Chapter 4, involves the sequencing of selected regions of the genome using capture probes. This often allows a much higher depth of sequencing of more than 10x that achieved in whole genome sequencing.

Genomic libraries are first prepared and amplified by polymerase chain reaction (PCR). These are then hybridised to specifically-designed capture probes which are complementary to target sequences in the gene of interest. These probes are then enzymatically extended, and the post-capture library is amplified using PCR, and sequenced using methods described in the previous section.

In our study of presumed sporadic early-onset SVD stroke, we used a multi-gene sequencing platform known as ThromboGenomics,[197] which is described in detail by Simeoni and colleagues, and briefly summarised in this chapter.

2.1.6 High throughput sequencing: quality control and analysis pipeline

Whole genome sequencing and sub-exome panel sequencing data are run typically run through a bioinformatics pipeline for quality control, alignment to a reference genome, variant calling and annotation. A typical workflow and the resulting file formats is shown in **Figure 2-2**.

Nucleotide sequences generated by sequencers such as the HiSeq X used for whole genome data, and HiSeq 2000 used for multi-gene sequencing in this work, are stored in text-based .fastq files together with quality scores. These scores measure the probability of an incorrect base call. The raw data is subjected to quality control steps during which the read lengths and number of reads are assessed, and adaptors or contaminants are trimmed from the reads.

The millions of reads generated are then aligned to a reference genome (or transcriptome in the case of RNA-seq data). The quality of mapping is assessed, and duplicates are removed. The resulting nucleotide sequence alignment files are in the .sam (sequence alignment/map) format, and the equivalent compressed binary format of this file is .bam.

Variants are called on .bam files by comparing the aligned files to the reference genome. This produces variant call format (.vcf) files, which detail information about positions in the genome, and quality and mapping scores. Variants are annotated using information from predictive tools such as Genomic Evolutionary Rate Profiling (GERP), MutationTaster, Sorting Intolerant from Tolerant (SIFT) and Polyphen to provide data on the expected impact and deleteriousness of each variant. Further downstream analysis can then be performed, and subsets of these resultant files can be created during the analysis process to produce .vcf files with different combinations of annotations.

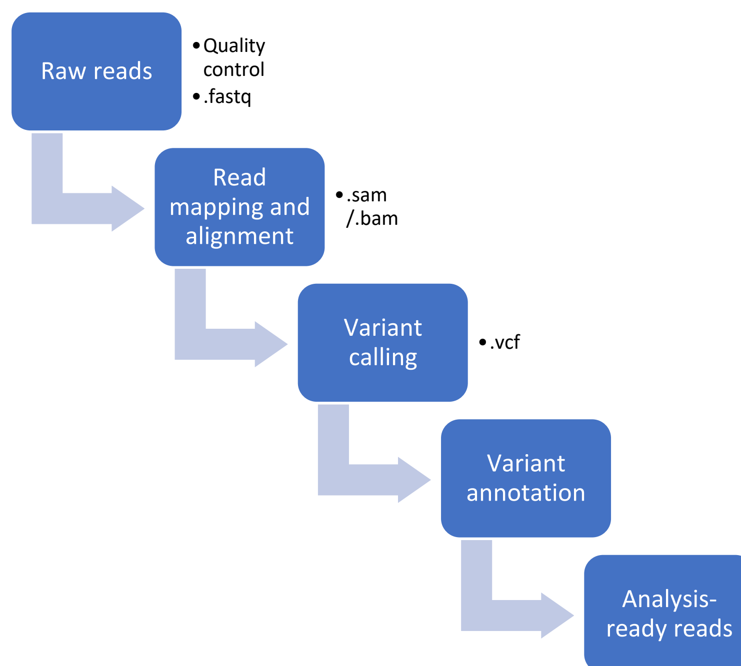


***Figure 2-2**: Generic processing pipeline for high throughput sequencing data.*

2.1.7 Current challenges in using HTS in the clinic

2.1.7.1 Interpretation of variants

The interpretation of sequencing results remains the greatest challenge in the use of HTS for diagnosis in the clinic. This is largely due to the vast amount of data produced in the context of an as-yet poorly understood genome. A single assay can produce hundreds of thousands of variants in a single individual suspected of having a single variant causative of disease – and the challenge lies in identifying which variant is pathogenic.

While intensive filtering against the predicted impact of variants and their frequencies in the general population may help to narrow down candidates, there are likely to also be pleiotropic variants where the same mutation in the same gene may have different phenotypic effects. For example, hexanucleotide intronic expansions in C9ORF72 have been found to lead to the presence of TDP43 positive inclusions in the CNS in both frontotemporal dementia (FTD) and amyotrophic lateral sclerosis (ALS). Despite this similar genotype and pathology, the clinical presentations of ALS and FTD are vastly different and would not conventionally be classed together in the same phenotype for exploratory analysis.[198] The lack of a one-gene, one-disease correlation can pose a significant challenge to identifying a single causative variant.

In addition, reference genomes and databases are by no means a representation of what is normal, but only a reflection of what has already been sequenced. Filtering is often based on the assumption that the potentially disease-causing variants are not carried by any of the unaffected individuals, however it is more likely that there are a small number of pathogenic alleles that do segregate into the general population at low frequencies. The inheritance pattern of the disease further complicates this – in recessive disorders where the carrier status does not result in disease, carriers may be erroneously included in the control set and identification of causative variants becomes complex.[199]

Further challenges of using HTS in the context of the work in this thesis are explored in Chapter 6.

2.1.8 Control databases

Established control databases are useful in genetic studies as they aid in the clinical interpretation of the pathogenicity of variants, and provide an additional dataset to that used in each study. These databases are, however, limited by the ethnicities and populations covered, with European populations taking predominance. Two of the largest sequencing databases are used in this thesis.

The Exome Aggregation Consortium (ExAC) database is a catalogue of exome sequencing data of 60,706 unrelated individuals across different populations, allowing the estimation of allele frequency for protein-coding genetic variants (http://exac.broadinstitute.org/downloads). The database is curated by the Broad Institute, and is derived from disease-specific and population-based studies.[200]

From the ExAC database grew the genome aggregation database or gnomAD (http://gnomad.broadinstitute.org/), which comprises the exome sequences of 123,136 unrelated individuals and whole genome sequences of 15,496 unrelated individuals. This expanded the available dataset significantly to include non-coding regions, and is now able to provide a more robust estimate of allele frequencies in the population. Although derived from both disease and population-based cohorts, in the latest release (February 2017) severe paediatric onset disease-afflicted individuals and their first-degree relatives have been removed.[201]

2.1.9 Predicting pathogenicity at a gene or protein level

In-silico bioinformatic tools can be used to predict the impact of a variant, such as if it results in a loss or gain of protein function. These can be useful in narrowing down potentially deleterious variants, but are not clinically useful when used in isolation. Three illustrative examples are discussed here.

2.1.9.1 SIFT

SIFT, or sorting intolerant from tolerant (http://sift.jcvi.org/), is an algorithm which predicts the deleteriousness of an amino acid substitution. Curated by the Craig Venter Institute, the algorithm is based on the assumption that residues which are important in the function of the

protein would be highly conserved across species or generations of a species, whereas less conserved positions are less likely to be critical to protein structure or function and thus tolerate substitution. SIFT predicts if a variant is 'tolerated' or 'damaging' using the degree of alteration from closely related sequences.[202]

2.1.9.2 Polyphen-2

Polyphen-2 (Polymorphism Phenotyping v2, http://genetics.bwh.harvard.edu/pph2/), is a tool which predicts the impact of an amino acid substitution on the structure and function of the resulting protein. It uses physical and comparative factors to compute this, based on sequence (i.e. the specific region or domain of the protein), phylogenetic (conservation of the protein sequence based on its alignment with homologues), and structural information (the impact on protein folding and the three-dimensional conformation). Each variant is classified as 'benign', 'possibly damaging' or 'probably damaging'.[203] Polyphen and Polyphen-2 were developed by the Brigham and Women's Hospital.

SIFT and Polyphen are both protein-sequence based, and are thus not useful for the interpretation of non-coding variation. As SIFT uses only data on conservation, it does not take into account any functional information. The Polyphen algorithm is also trained on known pathogenic variants, which may not necessarily be predictive of other novel pathogenic variants. It is also worth noting that both SIFT and Polyphen have been suggested to have high sensitivity but low specificity, thus limiting their utility in the interpretation of variants of uncertain significance.[204]

2.1.9.3 CADD

In both our sequencing studies in Chapters 4 and 5 we have used Combined Annotation Dependent Depletion (CADD) scores (http://cadd.gs.washington.edu) as a filter for variants identified. This tool, developed by the University of Washington and Hudson-Alpha Institute for Biotechnology, integrates 63 distinct annotations including SIFT and Polyphen, functional genomic data and transcription factor binding information to provide an estimate of deleteriousness (i.e. the degree to which fitness is reduced) and pathogenicity. This tool scores both SNVs and short insertions or deletions, and prioritizes known pathogenic variants.[205] Although it is an improvement on earlier protein-level tools such as SIFT and

Polyphen it remains an aid to narrowing down variants, and not a clinical tool. The cut-offs for potential pathogenicity are also arbitrarily set – in our studies we have used a score of 15 as a threshold for data produced from both whole genome sequencing and sub-exome panel sequencing. This value represents the median score for all theoretical splice site changes and variants which result in an alteration of an amino acid.[205]

2.2 Patient cohorts studied

Cerebral Autosomal Dominant Arteriopathy with Subcortical Infarcts and Leukoencephalopathy (CADASIL), caused by mutations in the NOTCH3 gene, is the most common inherited form of stroke, with an estimated prevalence of 2 to 4 per 100 000 in the UK population.[40,41] Multiple families have now been identified with CADASIL-like syndromes, but with no mutations in the NOTCH3 gene.[206,207] While the genetic basis of a few of these have subsequently been resolved with the aid of linkage analysis in the context of whole exome sequencing,[208,209] many other families remain without a genetic diagnosis.

This thesis explores familial cerebral SVD from by moving from a narrow to broad focus, looking first at the phenotype of CADASIL (Chapter 3) as a prototype of familial SVD, and then examining multiple monogenic disease genes in a cohort of 'sporadic' SVD stroke patients (Chapter 4), followed by exploring a larger number of candidate genes in a cohort of NOTCH3-negative familial SVD patients (Chapter 5). Finally, it considers the prevalence of known monogenic SVD-causing mutations in the general population (Chapter 6).

2.2.1 CADASIL

Data from 300 symptomatic CADASIL patients (170 females, 130 males) seen in the UK CADASIL National Referral Service was collected prospectively over a nineteen-year period, from 1996 to 2015. All patients had a confirmed diagnosis of CADASIL, either by direct sequencing of the NOTCH3 gene to identify a pathogenic cysteine-changing mutation (n=296), by electron microscopy of a skin biopsy demonstrating granular osmiophilic material (n=2), or by characteristic imaging features of CADASIL in the context of a family history of genetically-confirmed CADASIL (n=2). Pre-symptomatic patients diagnosed with CADASIL on predictive genetic testing were not included in this study. All patients were evaluated and examined by consultant neurologists. Written and informed consent was obtained, and data was collected using a standardised questionnaire at first review and at follow-up encounters.

2.2.2 Suspected familial SVD

Patients with a suspected familial SVD, but who did not have a disease-causing NOTCH3 mutation, were recruited to the National Institute for Health Research (NIHR) BioResource Rare Disease (BRIDGE) study. This was a multicentre pilot study of the UK's 100 000 Genomes Project, and collectively involved the recruitment and whole genome sequencing of approximately 10 000 individuals across 14 rare diseases. A summary of these diseases is included in **Table 2-2**.

Table 2-2: *Rare diseases in the BRIDGE Whole Genome Sequencing study*

Rare disease category	Number of patients sequenced
Cerebral small vessel disease	265 (data from 246 available at time of writing)
Bleeding and Platelet Disorders	1110
Pulmonary Arterial Hypertension	1136
Primary Immune Disorders	1320
Specialist Pathology: Evaluating Exomes in Diagnostics	1388
Steroid Resistant Nephrotic Syndrome	250
Primary Membranoproliferative Glomerulonephritis	1360
Ehlers-Danlos and Ehlers-Danlos-like syndromes	94 (WES only)
Multiple Primary Malignant Tumours	559
Intrahepatic Cholestasis of Pregnancy	262
Hypertrophic Cardiomyopathy	241
Neuropathic Pain Disorders	170
Stem Cell and Myeloid Disorders	221
Leber Hereditary Optic Neuropathy	72
[Genomics England Pilot]	2000

Participants were recruited across 12 sites in the UK, as listed in Table 2-3. Patients presenting to acute stroke units and those referred outpatient TIA and stroke clinics were screened for eligibility for the BRIDGE study. Potential participants were screened both retrospectively and prospectively. Patients who tested negative for NOTCH3 mutations at the South West Thames Regional Genetics Service from 2010 to 2014 were also retrospectively screened for eligibility

(n=1135), while all clinicians receiving negative NOTCH3 mutation testing reports issued by the Royal Devon and Exeter Clinical Genetics Service were informed of the study. A statement was included on all negative reports for NOTCH3 mutation screening generated by the clinical genetics laboratory to alert clinicians of the possibility of whole genome sequencing under the BRIDGE project should their patients be eligible.

Clinical records for potential participants were reviewed by a consultant neurologist or consultant stroke physician for all cases except those reviewed at the Leicester University Teaching Hospitals and Royal Devon and Exeter, where records were reviewed by both a consultant neurologist and a consultant clinical geneticist. Patients were selected if they were suspected to have familial SVD, based on criteria including their clinical presentation, with the onset of MRI-confirmed SVD, and clinical features such as lacunar stroke or cognitive impairment, at an early age (typically before the age of 60, although this arbitrary cut-off served as a guideline rather than an absolute criterion). Other neurological features typical of CADASIL such as complicated migraines, seizures and psychiatric disturbance counted towards the suspicion of familial disease. These clinical factors were interpreted in the context of the patients' risk factors for sporadic SVD, such as diabetes, hypertension, obesity and smoking. Further guidelines for the selection of patients for whole genome sequencing are summarised in **Table 2-4**. As described above, clinical information and MR images for all recruited patients underwent assessment by a consultant neurologist, with phenotypic data collected using a standardised questionnaire at recruitment and at follow-up encounters.

In addition to patients testing negative for NOTCH3 mutations on screening by clinical genetics departments, patients in the Oxford Vascular study (OXVASC) were also reviewed. The OXVASC (https://www.ndcn.ox.ac.uk/research/oxvasc) is a longitudinal study of patients with stroke, TIAs and heart attacks in Oxfordshire. Patients presenting with stroke are all phenotyped according to the TOAST classification, and are followed up at varying intervals for up to 10 years. Medical records for patients with strokes of the SVD subtype at or before the age of 65, with a family history of stroke were reviewed. Forty-two records were reviewed, three patients were found to be eligible and two were successfully recruited.

A total of 280 individuals were recruited to the project, and of these 265 samples were sent for whole genome sequencing. Nine samples had failed initial stages of quality control (QC) prior to sequencing, 4 were saliva samples and thus were not selected for sequencing, 1

sample was not sequenced as it belonged to a pedigree where the proband's sample had failed QC. Of the 265 samples sent for sequencing, seven were further excluded due to other diagnoses being made after recruitment (such as CADASIL on clinical genetic testing, or neurosarcoidosis), and due to probands in the same pedigree not having been successfully sequenced. At the time of writing data was available from 246 individuals (118 index, 121 relatives, 143 affected). The final set of data will comprise 129 index cases, and a total of 157 affected individuals.

Table 2-3: *Sites and total recruitment over a 24-month period for the BRIDGE-SVD Whole Genome Sequencing study*

Site	Approach to participant screening	Number of index cases	Number of relatives	Total number recruited
Cambridge University Hospitals	• CADASIL national referral service • Neurology and stroke outpatients, inpatients • Direct referrals	77	44	121
St George's Hospital, London	• (Former) CADASIL national referral service • Neurology and stroke outpatients, inpatients • All patients testing negative for NOTCH3 mutations	16	25	41
University College London Hospitals	• Neurology and stroke outpatients, inpatients • Patients testing negative for NOTCH3 mutations	3	9	12
Leeds Teaching Hospitals		11	14	25
Sheffield Teaching Hospitals		13	30	43
Southern General Hospital (Glasgow)		12	6	18
Oxford University Hospitals	• OXVASC database[210] • Patients testing negative for NOTCH3 mutations	2	1	3
Newcastle-upon-Tyne Hospitals	Patients testing negative for NOTCH3 mutations	7	4	11
Leicester University Teaching Hospitals		1	0	1
King's College Hospitals (London)		2	0	2
Imperial College Hospitals (London)		1	0	1
Royal Devon and Exeter Hospital	Neurology and stroke outpatients, inpatients	1	1	2
Total		146	134	280

Table 2-4: *Selection criteria for patients with suspected monogenic SVD*

Diagnosis of SVD	• MRI-confirmed features of SVD • Exclusion of other aetiology for stroke, such as cardioembolic or large vessel atherosclerosis

Clinical presentation	• Early-onset stroke or other features of SVD such as cognitive impairment (typically before age 60).
	• Syndromic disease: history of other clinical features which fit with recognised monogenic stroke syndromes
	    o Other neurology such as complicated migraines, seizures, psychiatric disturbance
	    o Non-neurological features such as skeletal, facial or ocular abnormalities

Risk factors	• Degree of severity of SVD in context of patient's risk factors, such as diabetes, hypertension and smoking
Family history	• Family history of early-onset stroke or dementia, especially if occurring in a Mendelian pattern of inheritance • Family history of neurological disease which may, in retrospect, have been SVD, such as familial hemiplegic migraine, multiple sclerosis, Alzheimer's dementia and Parkinson's disease dementia.

Atypical radiological features	• Presence of atypical features on imaging, such as
	    o Evidence of SVD beyond what is expected for age and risk factors
	    o Atypical distribution of white matter hyperintensities on T2/FLAIR MRI, such as anterior temporal pole and external capsule involvement (suggestive of CADASIL), or predominant posterior subcortical white matter involvement (suggestive of cerebral amyloid angiopathy)
	• Extensive microbleeds in the cortical (suggestive of CAA) or subcortical (suggestive of COL4A1/A2) regions
	• Presence of pseudotumours (seen in RVCL-S)
	• Presence of vascular malformations such as aneurysms (suggestive of COL4A1/A2), or dolichoectasia (suggestive of Fabry disease)

Exclusion	• Exclusion of other causes of white matter disease such as metabolic disorders, such as through urine organic acid and very long chain fatty acid (VLCFA) analyses. • Known Cysteine-changing variant within exon 2 to 24 of NOTCH3 gene • Any other known SVD-causing mutation

2.2.2.1 Phenotypes of patients recruited for Whole Genome Sequencing

The 265 individuals sequenced under the BRIDGE-SVD study were phenotyped using Human Phenotype Ontology (HPO) terms based on the information collected at all initial and follow-up encounters. All patients with evidence of amyloid or non-amyloid SVD on MRI were classified as 'affected', and relatives who had not had an MRI were classified as having an 'unknown' disease status. Information and ages at onset were also collected on the features summarised in **Table 2-5**.

Table 2-5: Phenotypic data collected on participants recruited to BRIDGE-SVD

	Feature and HPO terms	HPO terms
Imaging	• Presence or absence of features of cerebral small vessel disease on MRI • SVD subtype (amyloid or non-amyloid)	
Neurology *denotes 'major' events	• Migraine o with aura o without aura o Aura types: sensory, motor, sensorimotor, confusion, dysphasia	HP:0002076 HP:0002077 HP:0002083
	• Encephalopathy*, defined as acute reversible encephalopathy with evidence of reduced consciousness in the absence of any other organic cause, where symptoms lasted for longer than 24 hours, and were sufficient to warrant hospital admission.	HP:0004372
	• Transient Ischaemic Attack*	HP:0002326
	• Stroke* o Haemorrhagic o Ischaemic	HP:0001297 HP:0001342 HP:0002140
	• Seizures*	HP:0001250
	• Psychiatric disturbance	HP:0000708
	• Non-dementia cognitive impairment as diagnosed on the Brief Memory and Executive Test (BMET) and/or Montreal Cognitive Assessment (MoCA)	HP:0100543
	• Dementia*	HP:0000726
	• Gait and movement disorders*	HP:0001288
	• Fatigue	

	Feature and HPO terms	HPO terms
Other neurology	• Hearing impairment	HP:0000365
	• Visual impairment	HP:0012373
	• Pseudobulbar signs	HP:0002200
	• Postural instability	HP:0002172
	• Vertigo	HP:0002321
Cardiovascular risk factors prior to first major event or at presentation	• Smoking • Alcohol excess • Height and weight	
	• Hypercholesterolaemia	HP:0003124
	• Type 2 Diabetes Mellitus	HP:0005978
	• Essential hypertension	HP:0000822
	• Ischaemic heart disease	HP:0001677
	• Peripheral vascular disease	HP:0004950
Family history	• Stroke • Cognitive impairment or dementia • Migraine (with or without aura) • Psychiatric problems • Epilepsy • Any other neurological problems • Cause of death and age at death	
Baseline blood results	• Haemoglobin • White cell count • Platelet count • Glucose • HbA1C • Lipid panel (total cholesterol, LDL, HDL, Triglycerides) • Creatinine • Homocysteine	

2.2.2.2 Recruitment, sample collection and processing

Patients identified as suitable for whole genome sequencing were contacted via telephone or post. These patients and their relatives were either recruited during a home visit, in an outpatient research clinic, or through the post. 10-15ml of whole venous blood in ethylenediaminetetraacetic acid (EDTA) was collected from participants, either in the research clinic, or at their local primary care practice or hospital. Samples not collected locally were sent through regular post in sealed specimen bags and padded envelopes, together with the consent form. Legacy frozen blood samples from deceased or anonymised patients were retrieved and sent directly to the lab. In other deceased patients not local to Cambridge, stored DNA samples at room temperature were requested from clinical genetics laboratories. Where patients were unable to provide a blood sample, a saliva sample was collected in the Oragene DNA (OG-500) kit by DNA Genotek.

All samples were processed by the Cambridge Translational Genomics (CATGO) laboratory at the Cambridge Biomedical Research Centre. These were barcoded, and DNA was extracted from whole blood and checked for genomic/self-reported sex and ethnicity matches. Where extracted DNA was provided, DNA was checked for integrity and quantity. Samples passing initial quality control measures were then forwarded on to Illumina for whole genome sequencing. All saliva samples were processed, but not used for sequencing.

No quality control issues were encountered with fresh blood samples sent through the post, or thawed frozen samples. Six out of ten stored DNA samples had insufficient quantity for sequencing.

2.2.3 Patients with early-onset, presumed sporadic SVD (UK Young Lacunar Stroke DNA Study)

In Chapter 4 we investigate the role of monogenic disease genes in the context of presumed sporadic SVD, using a targeted multi-gene HTS panel. One-thousand and twenty-nine Caucasian patients with MRI-confirmed lacunar stroke occurring at or before the age of 70 were recruited to the UK Young Lacunar Stroke DNA Study from 2002 to 2012, as previously described.[29] Patients underwent full investigation for the aetiology of their stroke, and were categorised according to the Trial of Org 10172 in Acute Stroke Treatment (TOAST) classification, and only those with strokes of presumed SVD origin were included in this study.

White matter hyperintensities were graded on MRI using the semi-quantitative Fazekas scale.[211]

DNA extracted from whole venous blood from 993 participants of the UK Lacunar Stroke Study were processed at the Cambridge Biomedical Research Centre. Stored DNA samples were quantified using a Qubit Fluorometer and the GloMax system.

The quality of the DNA was then assessed by agarose gel electrophoresis. Samples were sequenced on a high-throughput sequencing multi-gene platform known as ThromboGenomics.[197] In this process, DNA was then fragmented into 200bp fragments and the designed probes were used to capture each DNA library. We designed probes to capture the consensus coding sequence (CCDS), 5' and 3' UTRs, as specified by Ensembl and RefSeq, as well as 1000 bases upstream of the start of the transcript, and variants reported on the Human Genome Mutation Database (HGMD) of candidate genes.

The resulting libraries were amplified, then sequenced using Illumina sequencing as described above. Reads were subject to quality control, and aligned to a reference genome (build GRCh37), before variants were called.

The platform called single nucleotide variants, short insertions and deletions, and large copy number variants in 16 candidate SVD-associated genes. An automated filtering procedure prioritized the variants based on their presence in the Human Gene Mutation Database (HGMD), allele frequency and predicted impact on translation, transcription, splicing and protein structure.

2.3 Conclusions

In this thesis we consider the genetic basis of cerebral small vessel disease from the perspective of the 'one-gene, one-disease' hypothesis. There are two key approaches taken in sequence in this thesis: the recruitment and phenotyping of patients, and the sequencing and interpretation of genetic variants.

We have recruited patients with familial SVD, including both 'solved' cases where the genetic cause is known (in particular, CADASIL cases only), as well as 'unsolved' cases where there is no known genetic mutation identified. The selection of cases is a clinical decision made by a consultant neurologist, guided by criteria such as the age at onset of symptoms being at or

before the age of 60, family history of stroke in first-degree relatives and syndromic features such as complicated migraine with aura. Cases were first tested for NOTCH3 mutations, and if these were negative they fed into our whole genome sequencing study.

We have also used samples from presumed-sporadic young-onset SVD stroke patients. As the age at onset of familial SVD is relatively late in life compared to other genetic diseases, it is challenging to delineate whether an unsolved case is 'sporadic' or 'familial'. We thus attempt to explore the possibility that there are disease-causing mutations in Mendelian disease genes in this cohort of individuals

Using a variety of filter thresholds similar to those previously described,[212,213] we narrow down the number of variants identified, and interpret these in the context of published clinical and functional data. In the age of falling costs of high throughput sequencing technologies, we attempt to explore the utility of HTS technologies to provide a genetic diagnosis for unsolved familial SVD cases, as well as to provide an estimate of the prevalence of these genetic diseases in a cohort of patients with young-onset presumed-sporadic SVD stroke.

Chapter 3: Migraine, Encephalopathy and Stroke in CADASIL

Outline

3.1 Introduction

3.2 Methods

3.2.1 Statistical methodology

3.3 Results

3.3.1 The pattern of migraine in CADASIL

3.3.2 Response to therapies

3.3.3 The relationship between migraine and encephalopathy

3.3.4 The relationship between migraine and lacunar stroke

3.4 Discussion

3.5 Appendix

List of tables

Table 3-1	Features of migraine and stroke in 300 symptomatic CADASIL patients.
Table 3-2	Treatment and responses to acute and prophylactic management of migraine in CADASIL
Table 3-3	Features of CADASIL coma or encephalopathy in 33 patients with CADASIL

List of figures

Figure 3-1	Distribution of ages at onset of migraine, first episode of encephalopathy and first stroke
Figure 3-2	Ages at onset of migraine, encephalopathy and stroke for all symptomatic patients, females and males
Figure 3-3	Cumulative incidence of stroke in males and females, and migraineurs and non-migraineurs
Figure 3-4 (Appendix)	Cumulative incidence of stroke in males and females, and migraineurs and non-migraineurs, in an analysis with an equal number of males and females

3. Migraine, Encephalopathy and Stroke in CADASIL

3.1 Introduction

Cerebral Autosomal Dominant Arteriopathy with Subcortical Infarcts and Leukoencephalopathy (CADASIL) is the most common monogenic form of stroke, and is caused by mutations in the NOTCH3 gene.[51] CADASIL serves as an excellent model of SVD, as patients develop similar SVD features on imaging and histopathology at an earlier age before the development of other age-related pathology such as tau or Aβ deposits. This disease is thus a 'pure' prototype of SVD that is relatively uncontaminated by other confounding disease processes. This chapter explores CADASIL as an example of monogenic SVD, by examining the possible relationships between some of the key clinical features of the disease.

As described in Chapter 1, the clinical features of CADASIL tend to be exclusively neurological, causing subcortical lacunar infarcts and cognitive impairment progressing to vascular dementia.[43] However, the pathological changes in the form of granular osmiophilic material (GOM) are seen in both systemic and cerebral vessels,

Migraines are often the earliest feature of disease and have been reported in up to 75% of cases.[43–45] The pattern of migraines differs from that seen in the general population with a predominance of migraine with aura, and frequent reports of complicated and prolonged aura.[43]

Despite its prevalence there have been few detailed studies of migraine in CADASIL.[214–217] Furthermore, little is known about how it responds to therapy and whether the management of migraine in CADASIL should differ from that of migraine in the general population.[218–221] Epidemiological studies have demonstrated that migraine, in particular migraine with aura, is an independent risk factor for stroke in the general population.[222–225] However, whether the presence of migraine with aura in CADASIL is predictive of an earlier onset of stroke and a more severe phenotype is unknown. Another related clinical phenotype of CADASIL is encephalopathy or "CADASIL coma". This can be preceded by an acute migraine headache,[43,226] but the associations between migraine and encephalopathy are poorly understood.

In a cohort of 300 symptomatic patients with CADASIL we performed a cross-sectional study examining the prevalence and characteristics of migraine. We recorded responses to migraine medication, and analysed the relationship between migraine, and both stroke and encephalopathy.

3.2 Methods

Data from 300 symptomatic CADASIL patients (170 females, 130 males) seen in a UK CADASIL National Referral Service was collected prospectively over a nineteen-year period, from 1996 to 2015. All patients had a confirmed diagnosis of CADASIL, either by direct sequencing of the NOTCH3 gene to identify a pathogenic cysteine-changing mutation (n=296), by electron microscopy of a skin biopsy demonstrating GOM (n=2) or by characteristic imaging features of CADASIL, with a family history of genetically-confirmed CADASIL (n=2). Patients diagnosed with CADASIL on pre-symptomatic genetic testing were excluded from this study. Patients were evaluated and examined by consultant neurologists, written and informed consent was obtained, and data was collected using a standardised questionnaire at first review and at follow-up encounters.

Patients were diagnosed with migraine at the point of care according to the International Classification of Headache Disorders (ICHD). The first edition (ICHD-1) was used from 1996 to 2004, followed by ICHD-2 from 2004 to 2013, and ICHD-3 (beta version)[227] from 2013 to 2015. Details of migraine episodes such as the duration, frequency and auras were collected.

A cross-sectional analysis was carried out in 2015. The data was reviewed and all episodes of migraine were classified according to the ICHD-3 beta, where migraine with aura was defined as episodes where the 'aura is accompanied, or followed within 60 minutes, by headache'. Aura types were classified according to typical aura (visual, sensory and/or speech or language symptoms with no motor weakness or monocular field defect), hemiplegic (visual, sensory and/or speech or language symptoms together with motor weakness) and brainstem aura (at least two of dysarthria, vertigo, tinnitus, hyperacusis, diplopia, ataxia and a decreased level of consciousness).

Data was also collected on confusional migraine aura, defined as disorientation and anterograde amnesia, with or without a decreased level of consciousness,[228] preceding the headache and not requiring hospital admission.

CADASIL encephalopathy or ‘coma’ was diagnosed clinically as an acute reversible encephalopathy with evidence of reduced consciousness in the absence of any other organic cause, where symptoms lasted for longer than 24 hours,[226] and were sufficient to warrant hospital admission.

Stroke was defined as a clinical stroke syndrome with MRI confirmation of a subcortical lacunar infarct at a site corresponding anatomically with the symptoms. Haemorrhagic strokes or cortical infarcts were excluded from analyses.

Patients were asked about their responses to migraine treatment, and these were divided into two categories: (a) response: a consistent or partial reduction in severity or frequency of migraine attacks; or (b) no response: no response or worsening of migraine attacks. Prophylactic therapy was defined as medication taken specifically for the prevention of migraines, and not in response to an acute attack.

All patients have given written informed consent and the study was approved by the South Thames Research Ethics Committee.

3.2.1 Statistical methodology

Quantile–quantile plots and the Shapiro-Wilk test were used to estimate distributions of ages at onset of migraine, encephalopathy and stroke. Where distributions of ages at onset were non-normal, the Mann-Whitney U test was used to calculate differences between groups. Odds ratios were calculated using the 2x2 Fisher’s exact test, or a logistic regression analysis where there were multiple categorical variables. Statistical analyses were performed using the R statistical software (version 3.2.2). Probability values of $p<0.05$ were considered statistically significant.

To assess the association between migraine and lacunar stroke risk a competing risks regression analysis was used to estimate the difference in cumulative incidence of lacunar stroke between migraineurs and non-migraineurs and between the sexes. In competing risks analyses, patients are subject to multiple outcomes (‘failures’), one of which may preclude

the other, such as stroke (in this case, the outcome of interest) and death (a competing risk). A competing risks regression analysis can be used to demonstrate the cause-specific hazard of an event, as well as the cumulative incidence of the event while taking into account the influence of competing risks. It is thus superior to conventional survival analyses, which assume that competing events do not occur.

3.3 Results

3.3.1 The pattern of migraine in CADASIL

Two hundred and twenty-six of 300 cases (75.3%) had a history of migraine. Migraine was the first feature of CADASIL in 203 (67.7%) (Figure 3-1). Migraine was more common among females (139 of 170 women, 81.8%) than males (87 of 130 men, 66.9%).

The age at onset of migraine was highly variable, with a median of 28 years (interquartile range 20, range 5 – 61, mean ± SD 29.0 ± 13.1). The age at onset was earlier in females (median 25 years, interquartile range 20.5, mean ± SD 26.9 ± 12.7) than in males (median 31 years, interquartile range 18, mean ± SD 32.3 ± 13.2) ($p=0.004$, $w=7422$) and followed a left-skewed distribution in both sexes. (Table 3-1, Figure 3-2)

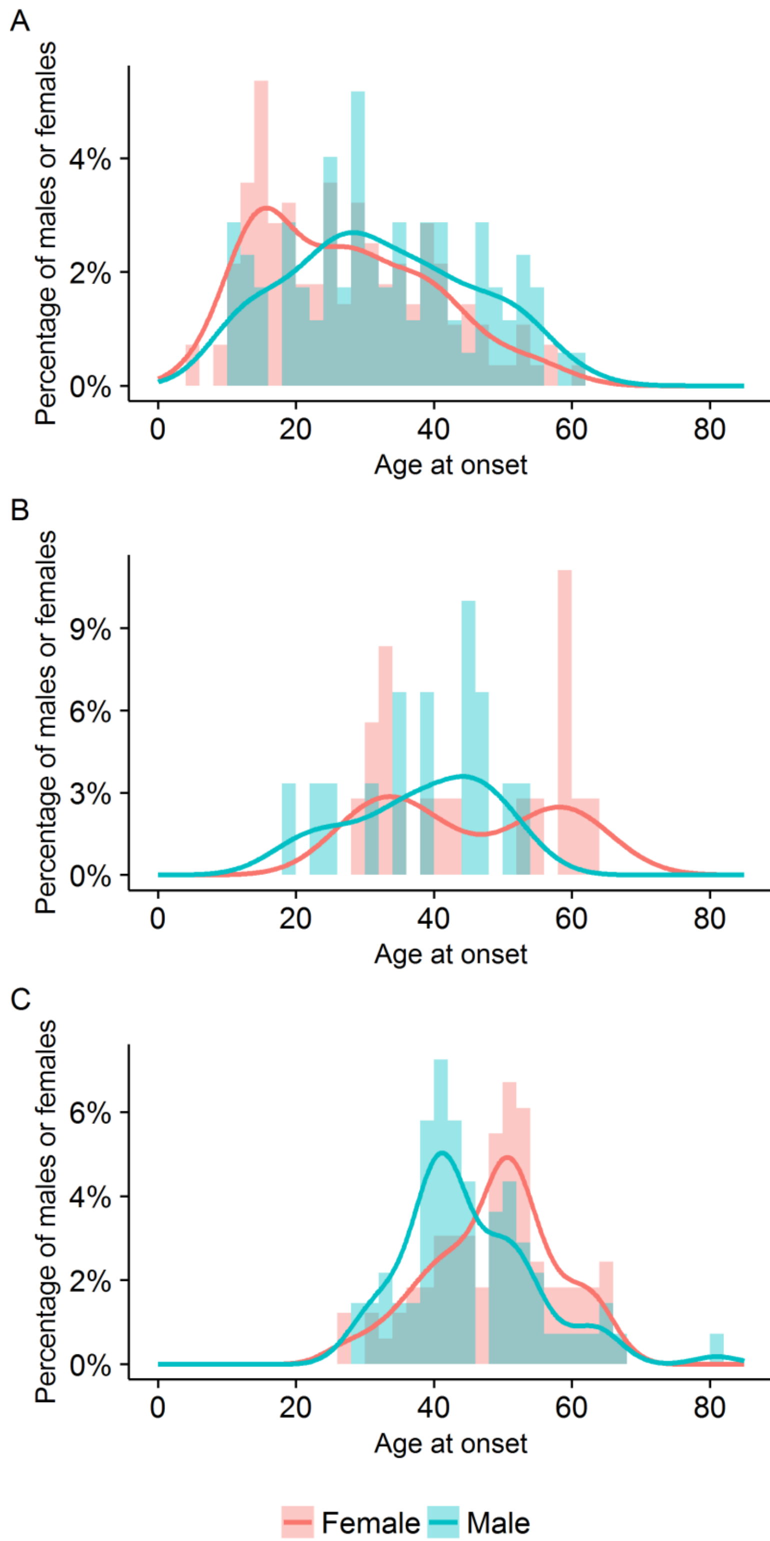


Figure 3-1*: Distribution of ages at onset of (A) migraine, (B) first episode of encephalopathy and (C) first stroke.*

Table 3-1: *Features of migraine and stroke in 300 symptomatic CADASIL patients.*

		Male	Female	Total	p-values
All symptomatic patients		130	170	300	
Migraine		87 (66.9%)	139 (81.8%)	226 (75.3%)	
Migraine as first feature		75 (57.7%)	128 (75.3%)	203 (67.7%)	
Migraine age at onset (years)	Median (interquartile range)	31 (18)	25 (20.5)	28 (20)	Mann-Whitney U test: Males vs. females: p=0.004, w=7422
	Mean (SD)	32.3 (13.2)	26.9 (12.7)	29.0 (13.1)	
	Range	10 - 61	5 – 60	5 – 61	
Migraine without aura		9 (6.9%)	19 (11.2%)	28 (9.3%)	
Migraine with aura		81 (62.3%)	122 (71.7%)	203 (67.7%)	
Migraine with and without aura		3 (2.3%)	2 (1.2%)	5 (1.7%)	
Migraine with aura (Percentage of those with aura)	Typical aura	80 (98.8%)	119 (97.5%)	199 (98.0%)	
	Typical aura with headache	75 (92.5%)	111 (91.0%)	186 (91.6%)	
	Typical aura without headache	10 (12.3%)	19 (15.6%)	29 (14.3%)	
	Hemiplegic migraine	14 (17.3%)	19 (15.6%)	33 (16.3%)	
	Confusional aura	17 (21.0%)	23 (18.9%)	40 (19.7%)	
Number of auras* experienced (% of those with aura)	1	51 (63.0%)	83 (68.0%)	134 (66.0%)	
	2	29 (35.8%)	36 (29.5%)	65 (32.0%)	
	3	1(1.2%)	3(2.5%)	4 (1.9%)	

		Male	Female	Total	p-values
Stroke (all) [†]		69 (53.1%)	82(48.2%)	151 (50.3%)	
	Lacunar stroke	68 (52.3%)	81 (47.6%)	149 (49.7%)	
	Recurrent strokes (percentage of patients with lacunar stroke)	34 (50.0%)	30 (37.0%)	64 (43.0%)	
Lacunar stroke age at onset (years)	Median (interquartile range)	43 (11)	50 (11)	48 (13)	Mann-Whitney U test: Males vs. females: p=0.003, w=1976.5
	Mean (S.D.)	44.9 (9.6)	48.8 (9.2)	47.0 (9.5)	
	Range	28 – 81	26 – 67	26 – 81	
Lacunar stroke – past medical history of migraine (% of those with lacunar stroke)		26 (38.2%)	57 (70.4%)	83 (55.7%)	

Table 3-1 footnotes: Asymptomatic CADASIL patients diagnosed on pre-symptomatic genetic testing were not included in this study. Migraine was classified according to the ICHD-3 beta, with aura classified according to typical aura (visual, sensory and/or speech or language symptoms and no motor weakness or monocular field defect) and hemiplegic migraine (visual, sensory and/or speech or language symptoms, as well as motor weakness). Data was also collected for patients who experienced confusional aura. *Aura types include typical, hemiplegic or confusional aura. [†]All strokes were subcortical lacunar infarcts, apart from one case of fatal brainstem haemorrhage in a patient on warfarin, and one patient with a cerebellar vermis haemorrhage. Haemorrhagic strokes were not included in subsequent analyses.

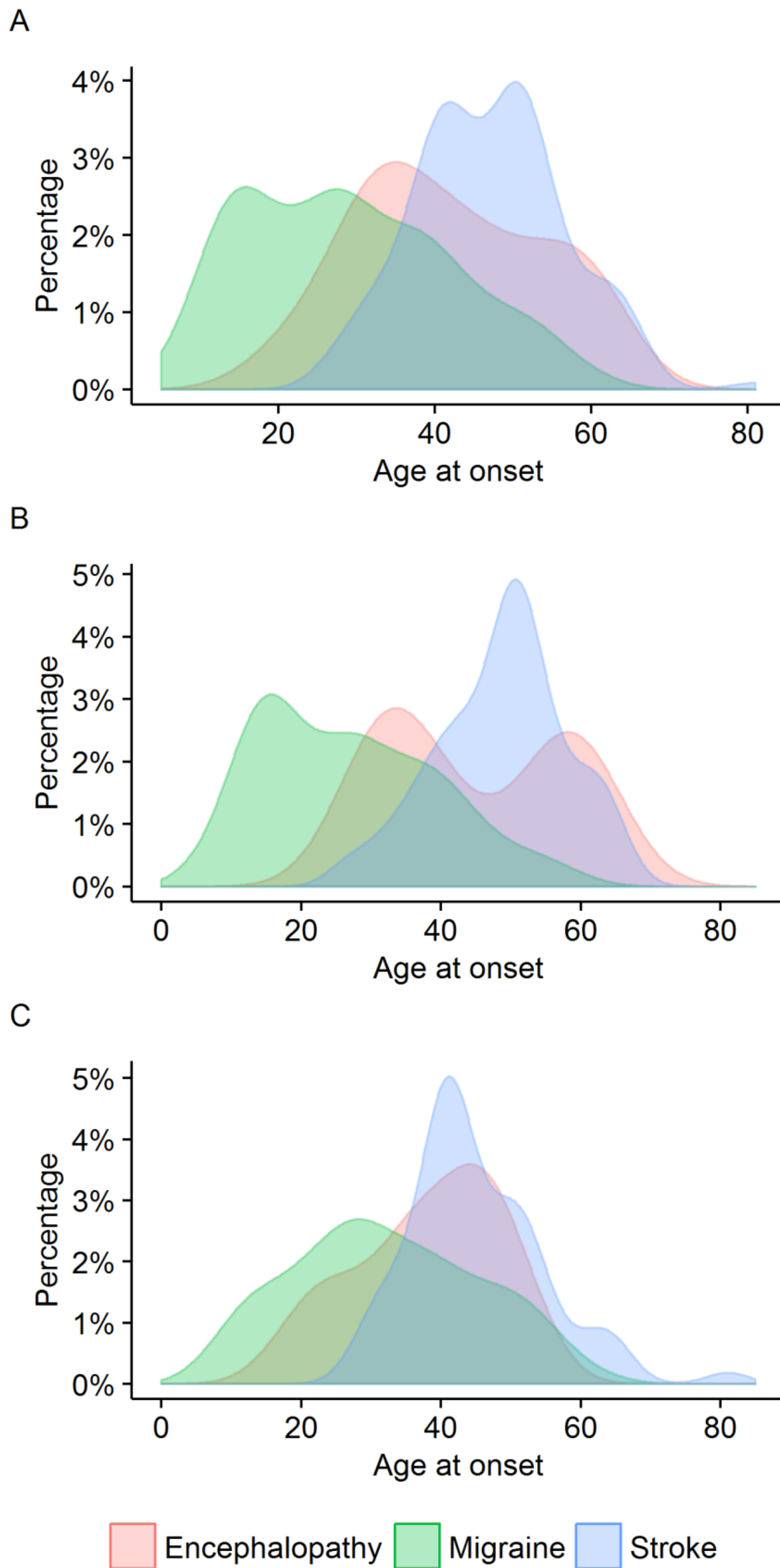


Figure 3-2: *Ages at onset of migraine, encephalopathy and stroke for (A) all symptomatic patients, (B) females and (C) males. Migraine was often the first feature of CADASIL, and preceded 75.8% of encephalopathic episodes.*

Two hundred and three (67.7%) patients with migraine experienced aura, with the majority of patients reporting more than one type of aura (Table 1). Of these 203 patients, 199 (98.0%) experienced typical aura, 33 (16.3%) fulfilled criteria for hemiplegic migraine, and 42 (20.7%) reported confusional aura (Table 3-1). None of these patients reporting confusional aura fulfilled criteria for brainstem aura.

3.3.2 Response to therapies

A variety of acute and preventive drugs had been used in treatment of the migraine. Of 226 patients with migraine, information on the treatment of migraine was available in 213 (94.2%). One hundred and three (48.4%) were taking at least one medication for their migraine. Seventy-four (34.7%) had taken medication for acute relief of migraine, and 48 (22.5%) had taken medication as migraine prophylaxis at some point in time. The treatments used and their responses are shown in Table 3-2.

Table 3-2: *Treatment and responses to acute and prophylactic management of migraine in CADASIL*

Drug		No. of patients (n=213)	Response	No response	Effect unknown
No drugs used in migraine treatment		110 (51.6%)	-	-	-
Drugs used in migraine treatment		103 (48.4%)			
	For prophylaxis	48 (22.5%)	31	10	7
	For acute relief	74 (34.7%)	43	14	17
Drugs used for acute treatment of migraine					
Simple analgesics*		31	20	4	7
	Paracetamol†	11	9	1	1
	Ibuprofen	8	5	2	1
	Aspirin†	8	7	0	1
	(Unspecified)	6	1	1	4
Paracetamol and codeine		30	20	5	5
	Co-codamol	3	1	1	1
	Co-dydramol	2	1	0	1
	Migraleve†	25	18	4	3
Co-proxamol (paracetamol + dextropropoxyphene)		1	1	0	0
Other NSAIDs	Indomethacin	1	1	0	0
Other opioid analgesics		3	2	0	1
	Tramadol	1	0	0	1
	Morphine	2	2	0	0

Drug		No. of patients (n=213)	Response	No response	Effect unknown
Triptans*†		24	10	12	2
	Sumatriptan§	18	9	9	1
	Naratriptan	2	0	1	1
	Zolmitriptan	3	1	2	0
	Lyophilisate rizatriptan	2	0	2	0
	(Unspecified)	2	1	1	0
Primidone		1	0	1	0
Domperidone		1	1	0	0
Diazepam		1	0	1	0
Drugs used as regular migraine prophylaxis					
β blockers*		25	10	8	10
	Propranolol†‡	22	8	8	6
	Atenolol	1	0	1	0
	Metoprolol	1	0	0	1
	(Unspecified)	2	1	0	1
Calcium channel blockers		4	3	1	0
	Flunarizine	1	1	0	0
	Verapamil	2	1	1	0
	Nimodipine	1	1	0	0
Pizotifen‡		16	8	5	3
Gabapentin		4	2	2	0

Drug		No. of patients (n=213)	Response	No response	Effect unknown
Amitriptyline‡		15	6	5	5
Anticonvulsants		6	4	2	0
	Topiramate	3	2	1	0
	Sodium valproate	3	2	1	0
Ergolines		2	0	1	1
	Methysergide	1	0	0	1
	(Unspecified)	1	0	1	0
Acetazolamide		1	0	0	1
Fluoxetine		1	0	1	0
Combinations	Pizotifen + Amitriptyline	2	2	0	0
	Amitriptyline + Propranolol	1	1	0	0
	Pizotifen + Amitriptyline + Propranolol	1	1	0	0

Table 3-2 footnotes:

*Some patients have tried more than one type of drug in this class.

†Drugs used both as regular migraine prophylaxis and pro re nata in acute treatment.

‡Drugs which failed in isolation or in some combinations but showed good responses in other combinations.

§Drugs with responses that later wore off.

For acute management, simple analgesics such as paracetamol, ibuprofen and aspirin were most commonly used (in 31 of 74 patients, 41.8%), and were reported to have a good or partial response in 20 patients (64.5 %). A combination of paracetamol and codeine (the most common formulation used was Migraleve) had been used in 30 patients (40.5%), with a good or partial response in 20 patients (66.7%).

Twenty-four patients had taken triptans with data on responses available in 22; of these they were associated with a response in 10 (45.5%) and lack of response in 12 (54.5%). One patient reported that the effectiveness of sumatriptan wore off after repeated use. Patients did not report any worsening of neurological status on triptans (Table 3-2).

The most common drugs used for migraine prophylaxis were beta-blockers. With these 10 out of 25 (32%) patients reported a response (40%) and 8 (32%) reported no benefit. Responses to other prophylactic agents are shown in Table 3-2.

3.3.3 The relationship between migraine and encephalopathy

Thirty-three (11.0%) patients had a history of CADASIL coma or encephalopathy and this was the presenting feature leading to a diagnosis of CADASIL in 26 (8.7%), and the first feature of CADASIL in 2 (0.67%) (Fig 2). Often initially diagnosed as a viral encephalitis with nonspecific results on analysis of cerebrospinal fluid, these episodes recurred in 10 (30.3%) patients. The age at onset of encephalopathy followed a bimodal distribution in females and a right-skewed distribution in males (Fig 1(B)). The median age at onset of encephalopathy was 40 years (interquartile range 19, mean ± SD 41.8 ± 12.1, and range 19-63). The age at onset of encephalopathy did not differ between the sexes ($p=1.0$) (**Table 3-3**). Four patients developed encephalopathy during the perinatal or postpartum period.

__Table 3-3__: Features of CADASIL coma or encephalopathy in 33 patients with CADASIL

Feature		Male	Female	Total	p-values
Encephalopathy (%)		15 (11.7%)	18 (10.8%)	33 (11.0%)	
	Encephalopathy as a presenting feature	11 (8.5%)	15 (8.8%)	26 (8.7%)	
	Encephalopathy as first feature	2 (1.5%)	0	2 (0.67%)	
Age at onset of encephalopathy (years)	Median (interquartile range)	38 (13)	41 (25.8)	40 (19)	**Mann-Whitney U test: Males vs. Females**: p=0.2, w = 100.5
	Mean (SD)	38.1 (10.2)	44.9 (12.9)	41.8 (12.1)	
	Range	19 – 52	29 – 63	19 – 63	
Encephalopathy – past medical history of migraine		13 (13 with aura)	18 (17 with aura)	31 (30 with aura)	
Duration of encephalopathy (days)	Median (interquartile range)	7 (4.5)	8 (4.8)	8 (5)	**Mann-Whitney U test: Males vs. Females**: p = 1.0, w = 136.5
	Mean (SD)	8.2 (3.8)	7.9 (3.3)	8.1 (3.4)	
	Range	3 – 17	3 – 14	3 – 17	
Recurrence of encephalopathy		4	8	12 (36%)	
	Preceded by migraine	13	12	25 (75.8%)	

Feature		Male	Female	Total	p-values
Associated features of encephalopathy	Fever	2	2	4 (12.1%)	
	Hallucinations	5	8	13 (39.4%)	
	Meningism	0	1	1 (0.03%)	
	Vomiting	6	3	9 (27.3%)	
	Seizures	3	7	10 (30.3%)	

Encephalopathic episodes were characterised by a reduced conscious level, or confusion, lasting up to 3 to 17 days, with a median of 8 days (interquartile range 5). In 25 patients (75.8%) the coma evolved from an acute migraine headache. Common accompanying features were seizures (n=10), fever (n=4), hallucinations (n=13), nausea and/or vomiting (n=9) and meningism (n=2) (Table 3). Patients also reported associated focal neurological symptoms typical of migraine aura, such as speech and language disturbances or sensorimotor deficit. Although the majority of encephalopathic episodes resolved completely, five patients reported the persistence of associated symptoms lasting at least 6 weeks after the acute episode.

Data on brain imaging (with diffusion-weighted MRI) during the episodes was available in 8 patients, and there were no acute DWI positive lesions. Cerebrospinal fluid (CSF) analysis results were available in 7 patients, of which 6 showed raised protein levels and two showed leucocytosis. All seven had normal glucose levels. Electroencephalograms (EEGs) were performed in five patients and showed generalized slowing.

Patients with a past history of migraine with aura had a higher odds ratio of developing encephalopathy (30 of 33 vs. 173 of 267; OR=5.4, 95% CI 1.6 – 28.4, p=0.002). Migraine sufferers with a history of confusional aura also had an increased risk of developing encephalopathy (OR=2.5, 95% CI 1.0– 5.8, p=0.04), but there was no increase in risk with other aura types, or with sex or age at onset of migraine.

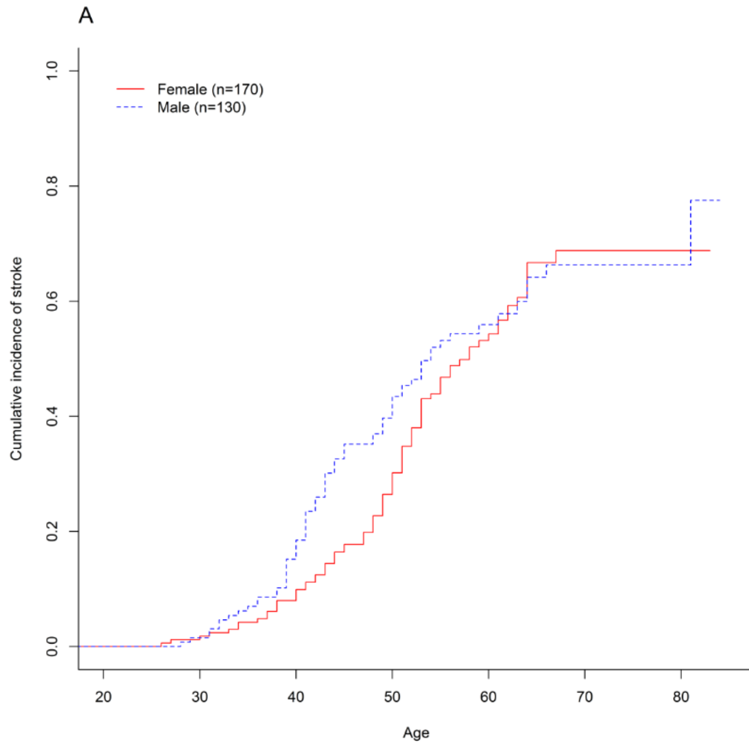
3.3.4 The relationship between migraine and lacunar stroke

One hundred and fifty-one patients (82 females, 48.2% and 69 males, 53.1%) had experienced stroke, all of which were subcortical lacunar infarcts, apart from one case of fatal brainstem haemorrhage in a patient on warfarin, and one patient with a cerebellar vermis haemorrhage, both of which were excluded from subsequent analyses. None of the lacunar infarcts were fatal. Sixty-four patients suffered more than one stroke.

The distribution of ages at onset of lacunar stroke followed a symmetrical and unimodal distribution in both sexes, with a median age at onset of first stroke of 48 years (interquartile range 13, mean 47.0, SD 9.5, range 26 – 81). The age at onset was higher in females (median

50 years, interquartile range 11, mean ± SD 48.8 ± 9.2) than in males (median 43 years, interquartile range 11, mean ± SD 44.9 ± 9.6), ($p=0.003$, $w=1976.5$) (Figure 3-1, Table 3-1).

Eighty-three (55.7%) patients (57 females, 70.4% and 26 males, 38.2%) had a past medical history of migraine before their first stroke. A history of migraine prior to stroke onset was associated with a lower cumulative incidence of stroke, compared to individuals without migraine or who developed migraine only after their first stroke. On competing risks analysis of the time to first stroke, the hazard ratio for the presence of a history of migraine prior to stroke was 0.5 (95% CI 0.3 – 0.6, $p = 2.1x\ 10^{-6}$) (Figure 3-3). Despite the difference in the ages at onset of stroke between the sexes, the hazard ratio for sex was not significant (HR for females: 0.9, 95% CI = 0.7 – 1.3, $p=0.57$) (Figure 3-2).



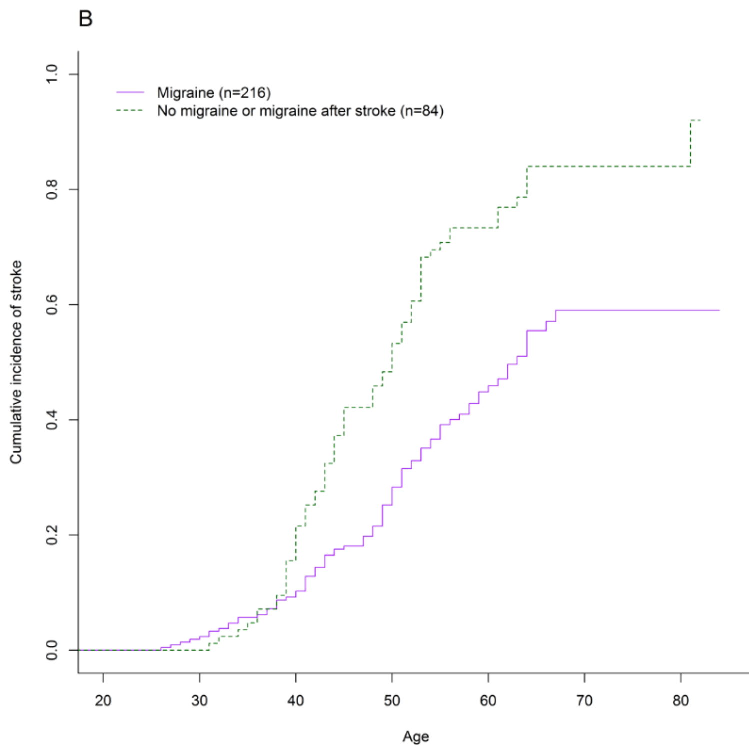


Figure 3-3: *Cumulative incidence of stroke in (A) males and females, and (B) migraineurs and non-migraineurs. As demonstrated on competing risks analysis, migraineurs had a lower hazard ratio than non-migraineurs (HR = 0.5, 95% CI 0.3-0.6, p = 2.1 x10*$^{-6}$*), while the hazard ratio for female sex was not significant (HR=0.9, 95% CI 0.7 – 1.3, p = 0.57).*

3.4 Discussion

In this study in 300 symptomatic CADASIL patients, migraine was the presenting symptom in two thirds, and present in three quarters. The nature of migraine differed from that in the general population, with 90% of CADASIL patients experiencing migraine with aura, and prolonged and complex auras being common. This is in contrast to the predominance of migraine without aura in the general population.[229–231] The age at onset of migraine in CADASIL was also later than in the general population.[232] However, consistent with migraine in the general population,[233] as well as previous studies of European CADASIL patients,[234] migraine in CADASIL was more common in females, with a mean age at onset of around 4 years earlier.

There is little published data on treatment responses to migraine in CADASIL.[218–221] In this cohort, a significant proportion of patients with migraine did not take medication due to the migraine episodes being infrequent. However in those that did, treatment responses seemed to be similar to that seen in the general population.[235,236] Of interest, 24 patients had taken triptans and none had reported serious side effects. Drug information sheets tend to indicate that triptans are contraindicated in patients with stroke or TIA, or at high risk of cardiovascular disease, because of the theoretical risk that they might exacerbate cerebral ischaemia.[237] In this retrospective study we found no evidence that this occurred. This is consistent with data from the general population which found no association of triptans with stroke risk,[238] however it is important to note that since some of these purported risks are similar to the natural progression of the disease it would be challenging to tease out if subsequent TIAs, strokes or worsening cognitive impairment could be attributed to the triptans. A fifth of CADASIL migraineurs required regular prophylactic treatment for migraines, with a variable response to these preventive medications.

Encephalopathy in CADASIL was first described as being completely reversible episodes of confusion, coma, fever and seizures.[226,239] In this study we found that 11% of patients developed such episodes warranting hospitalization. Encephalopathy was also the presenting feature leading to a diagnosis of CADASIL in majority of these 33 patients, hence CADASIL should be considered in the differential of a patient presenting with encephalopathy with inconclusive results on imaging, EEG and CSF analysis. Four patients experienced

encephalopathic episodes around the puerperium as previously reported.[46,240] The predilection for the puerperium suggests a possible hormonal contribution, which might also contribute to the increased prevalence of migraine in females.

Encephalopathy episodes have been previously reported to often develop from a migraine attack, and we hypothesised that they might share an underlying aetiology. Consistent with this, 76% evolved from a migraine attack, and we found that encephalopathy was more common in CADASIL cases who had previously experienced migraine with aura. Furthermore, individuals with migraine with aura who had experienced confusional migraine aura were more likely to develop encephalopathy, compared with those with other aura types. Our data is consistent with a continuum of symptoms in CADASIL, ranging from acute confusional migraine episodes lasting up to three days,[217,228] status migranosus with persisting aura lasting up to 8 days,[217] and encephalopathic episodes lasting up to 14 days.[43,226]

Migraine with aura is an independent risk factor for stroke, and is associated with white matter changes and silent infarct-like lesions on MRI.[223] Migraine and stroke also co-occur in other monogenic forms of small vessel disease. We hypothesised that migraine might also contribute to a more severe phenotype in CADASIL; however, we found no evidence of this and indeed non-migraineurs with CADASIL had a higher cumulative incidence of stroke. The explanation for this association is uncertain and the finding needs replicating in other cohorts, but it is reassuring that migraine with aura is not associated with a worse phenotype. The electrophysiological basis of migraine in CADASIL is believed to be Cortical Spreading Depression (CSD) in which waves of synchronised depolarisation spread across the brain cortex,[241] being associated first with an increase in cortical blood flow, followed by a period of hypoperfusion.[242] It is possible that such episodes could be associated with protection against subsequent ischaemia, perhaps by mechanisms related to ischaemic preconditioning.[243]

While this is one of the largest studies of migraine, encephalopathy and stroke in CADASIL, it is not without limitations. As this was a cross-sectional analysis of prospectively collected data, it is probable that more severe cases of CADASIL are detected earlier, while there may be many NOTCH3 mutation-carrying individuals with few or no symptoms that remain undiagnosed. There is currently no clinically accepted definition of encephalopathy in CADASIL. The distinction between complicated migraine with confusional aura and an

encephalopathic episode may be difficult – in this study we have set an arbitrary measure of severity (the need for hospitalisation) and a temporal definition (encephalopathic symptoms lasting at least 24 hours).

Our evaluation of treatment responses is also limited by the design of this study. As this was a cross-sectional study, patients were not randomised to different types of migraine therapy. It is thus challenging to estimate the complication rate from therapy in such a cohort, although none of the ischaemic events in this cohort were attributed to triptan use.

In conclusion our results provide data from a large population of prospectively recruited CADASIL patients on the frequency, characteristics and treatment responses of migraine. Our data shows that while many patients do not need treatment for their migraine, about half do, and in this group similar treatment responses are seen to that in the general population with migraine. Our series further characterises the CADASIL coma or acute reversible encephalopathy which may lie on the spectrum of migraine auras, sharing similar features with confusional migraine episodes.

3.5 Appendix

3.5.1 Correction for ascertainment bias

There was a disproportionate female-to-male ratio (170 females, 130 males) in this cohort, where females had a higher proportion of migraineurs (81.8%) compared to males (66.9%). Females also had a later age at onset of stroke (mean 48.4 ± 9.1 years) compared to males (45.1 ± 9.7 years). The competing risk regression analysis of all 300 individuals showed a reduced cumulative incidence of stroke in migraineurs than non-migraineurs, and no difference in the risk of stroke between the sexes.

In view of the disproportionate sex ratio and differing migraine prevalence within each sex, the competing risk regression analysis was re-run using two different approaches to confirm the results found in all 300 patients.

First, equal numbers of each sex were selected – the first 130 females recruited to the study, together with all 130 males. This included 183 migraineurs (61.5% of males and 79.2% of females), and 77 non-migraineurs. This again showed a reduced cumulative incidence of stroke in migraineurs, with a hazard ratio of 0.45 (95% CI 0.32 - 0.62, $p = 1.7 \times 10^{-6}$) (**Figure 3-4**) (A, B). There was no difference in the cumulative incidence of stroke between the sexes.

Second, patients were selected such that there were equal proportions of migraineurs for each sex. All 130 males (66.9% migraineurs) and 94 females were selected. The first 63 females with migraine recruited to the study were selected, together with the first 31 non-migraineur females (66.9% migraineurs). This approach produced similar results, with migraineurs having a lower cumulative incidence of stroke (hazard ratio 0.46, 95% CI 0.32 - 0.65, $p = 1.4 \times 10^{-5}$). Again, there was no difference in the cumulative incidence of stroke between the sexes. (**Figure 3-4**) (C, D)

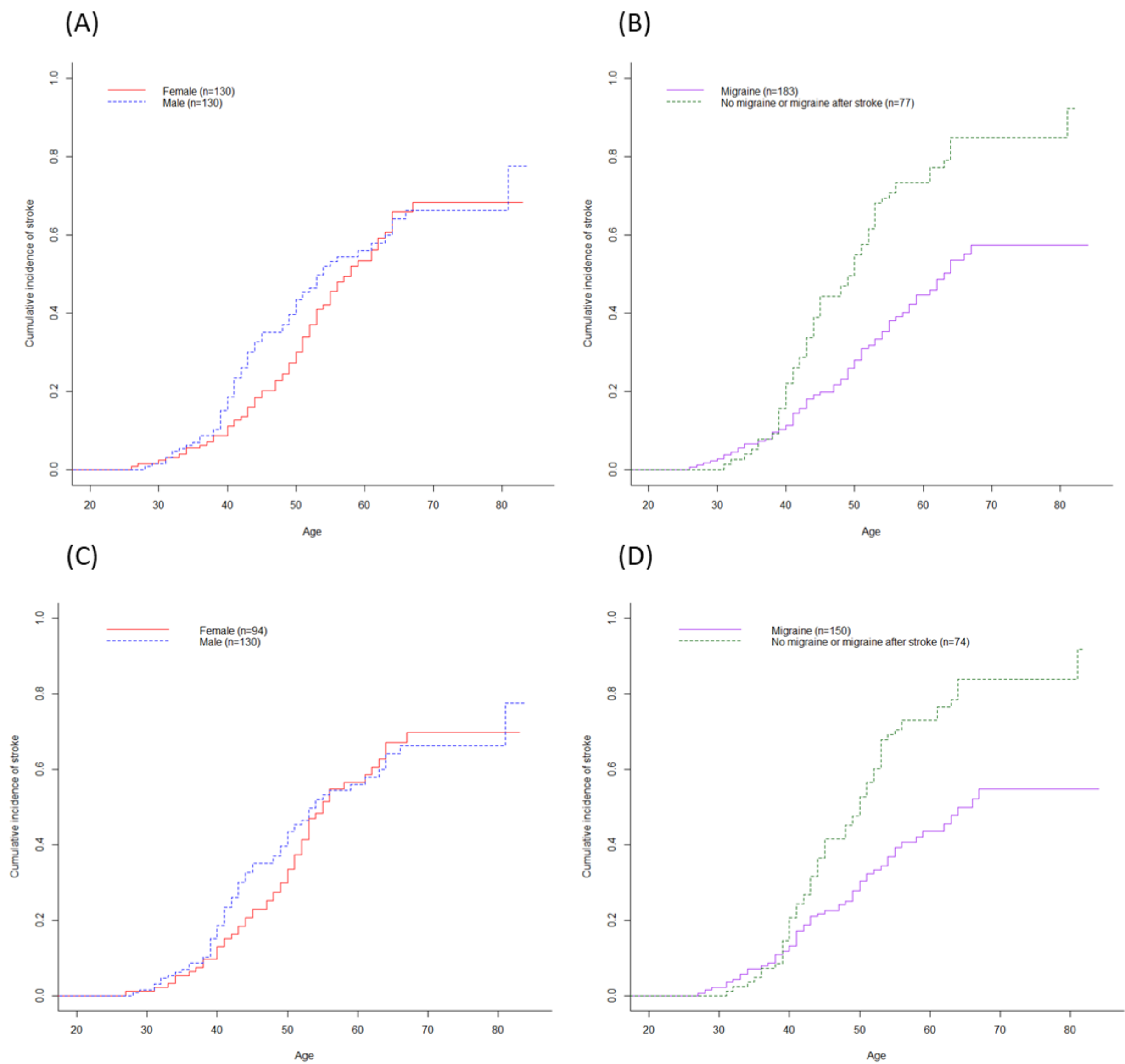


Figure 3-4: *Cumulative incidence of stroke in (A) and (C) males and females, and (B) and (D) migraineurs and non-migraineurs, in an analysis with an equal number of males and females (A and B). (C) and (D) show the same analysis in 130 males and 94 females both with equal proportions of migraineurs. As demonstrated on competing risks analysis in both approaches, migraineurs had a lower hazard ratio than non-migraineurs, while the hazard ratio for female sex was not significant.*

Chapter 4: A candidate gene study in presumed sporadic, early-onset small vessel disease

Outline

4.1 Introduction

4.2 Methods

- 4.2.1 Study population
- 4.2.2 Baseline characteristics of patients
- 4.2.3 Gene and transcript selection
- 4.2.4 Bioinformatics
- 4.2.5 Filtering of variants

4.3 Results and discussion

- 4.3.1 Sequencing quality control
- 4.3.2 Copy Number Variants (CNVs)
- 4.3.3 Single Nucleotide Variants (SNVs) and Indels

4.4 Conclusions

List of tables

Table 4-1	Baseline characteristics of 993 patients included in this study
Table 4-2	SVD-related candidate genes included on the ThromboGenomics sequencing panel
Table 4-3	Filters for single nucleotide variants and indels called on ThromboGenomics.
Table 4-4	Individuals with more than one CNV in the 15 candidate genes screened
Table 4-5	CNVs appearing in more than one individual
Table 4-6	Variants called in each gene in 993 SVD cases where a relaxed filter has been applied
Table 4-7	Single nucleotide variants with a high MAF identified on the relaxed filter
Table 4-8	Individuals with both CNVs and SNVs (passing the stringent filter) in candidate genes
Table 4-9	NOTCH3 cysteine-altering variants identified on both relaxed and stringent filters
Table 4-10	Heterozygous HTRA1 variants identified on stringent filtering
Table 4-11	COL4A1 and COL4A2 variants passing the stringent filter, and previously reported COL4A1 mutations identified in this cohort
Table 4-12	TREX1 variants passing both stringent and relaxed filters
Table 4-13	Variants passing stringent filters in candidate genes associated with related phenotypes
Table 4-14	Compound heterozygous individuals carrying both CNVs and SNVs (passing relaxed filter only) in the ABCC6 gene
Table 4-15	Patients found to have SNVs on more than one SVD candidate gene

List of figures

Figure 4-1	Examples of coverage profiles
Figure 4-2	Heterozygous HTRA1 variants passing the stringent filter, identified in 12 individuals
Figure 4-3	Missense and frameshift COL4A1 and COL4A2 variants passing the stringent filter, and COL4A1 variants previously associated with SVD

4. A candidate gene study in presumed sporadic, early-onset small vessel disease

4.1 Introduction

Cerebral small vessel disease (SVD) is estimated to cause up to a fifth of all ischaemic strokes and is a leading cause of vascular dementia. The main causes of SVD are cerebral amyloid angiopathy and arteriosclerosis, with hypertension being the major risk factor. The latter form of SVD, which is a disease of the small perforating arteries supplying the white and deep grey matter of the brain, and is also responsible for subcortical or deep intracerebral haemorrhages,[4] is considered in this chapter.

The majority of SVD cases are thought to be sporadic, arising in the elderly and hypertensive population. However, SVD is also the subtype of stroke that is most likely to present as a familial disease caused by a mutation in a single gene. The most common monogenic form of SVD is Cerebral Autosomal Dominant Arteriopathy with Subcortical Infarcts and Leukoencephalopathy (CADASIL), which is caused by cysteine-changing mutations in the epidermal growth factor-like repeat (EGFr) regions of the NOTCH3 gene.[39] Non-cysteine-changing variants have been reported as CADASIL-causing, however these associations remain unproven and controversial.[244]

Familial SVD is a rare group of diseases, with CADASIL estimated to occur in approximately 2 to 4 per 100 000 population.[40,41] A previous study of a population of presumed-sporadic, MRI-defined SVD stroke patients below the age of 70 (the UK Lacunar Stroke Study) showed that only five of 994 patients carried a CADASIL-causing cysteine-changing mutation[1], and no patients carried Fabry disease-causing mutations.[151] This cohort of patients is studied in this chapter.

However, the relative heterogeneity of symptoms and severity of disease features in pedigrees with CADASIL and other forms of familial SVD raises suspicion that there may remain many undiagnosed cases in the population. Many affected relatives in pedigrees with diagnoses of monogenic SVD have milder, non-specific features of familial disease such as

[1] In this study only exons 3, 4, 5, 6, 11, 18, 19 and 22 of the NOTCH3 gene were screened for mutations.

migraine, strokes and white matter abnormalities, and in the context of concomitant risk factors these patients may not often lead to a suspicion of familial disease, and would have been classed as 'sporadic' SVD, or remained undiagnosed. In addition, the five NOTCH3 mutation-carrying 'sporadic' stroke patients mentioned above had few clinical features specific to CADASIL, and three did not have any family history of stroke.[151] Hence, we speculate that many patients with SVD-related lacunar strokes which are presumed to be 'sporadic' may in fact have an underlying monogenic cause. This chapter aims to investigate the prevalence of rare variants in known familial SVD genes in a preliminary study of the same cohort of patients from the UK Lacunar Stroke Study, using a High-Throughput Sequencing (HTS) platform known as ThromboGenomics.[197]

4.2 Methods

4.2.1 Study population

Over an 8-year period from 2005, 1247 patients with suspected lacunar stroke were recruited from 72 specialist stroke centres across the UK. Patients underwent full investigation for the aetiology of their stroke, and were categorised according to the Trial of Org 10172 in Acute Stroke Treatment (TOAST) classification. One thousand and thirty of these patients with European ancestry and an MRI-confirmed lacunar strokes of presumed SVD origin before the age of 70 fulfilled inclusion criteria of the UK Young Lacunar Stroke DNA study, as previously described.[151] White matter hyperintensities (WMH) were graded on MRI using the semi-quantitative Fazekas scale.[211]

As part of the Young Lacunar Stroke DNA study, 994 DNA samples of sufficient quantity and quality were previously screened for disease-causing mutations in the NOTCH3 and GLA genes. Exons 3, 4, 5, 6, 11, 18, 19 and 22 of the NOTCH3 gene screened using denaturing high-performance liquid chromatography (DHPLC), and in addition, exons 3 and 4 were screened using Sanger sequencing. Up to 90% of mutations in the UK population were previously shown to be in these exons.[245] This screen showed that 5 individuals carried CADASIL-causing mutations. The GLA gene was screened using high resolution melt-curve analysis, covering all exons, all intron/exon junctions and one specific reaction aimed at a common deep intronic

mutation. No Fabry-causing mutations were identified.[151] In this study, DNA of sufficient quality and quantity was available from 993 patients.

4.2.2 Baseline characteristics of patients

Baseline characteristics and cardiovascular risk factor profiles of these 993 patients are summarised in Table 4-1.

Table 4-1: Baseline characteristics of 993 patients included in this study

Age at first stroke in years, mean (S.D.)	60.1 (8.6)
Number of males	705 (71.0%)
BMI, mean (S.D.)	28.7 (6.2)
Number of patients with:	
Hypertension	713 (71.8%)
Diabetes	164 (16.5%)
Hypercholesterolaemia	673 (67.8%)
Smoker (current or previous)	697 (70.2%)
Alcohol excess (≥20 units/week)	290 (29.2%)
Migraine	50 (17.9%)
Myocardial infarction/ coronary artery bypass graft or angioplasty	35 (3.5%)
Peripheral vascular disease	28 (2.8%)
Previous or recurrent strokes	37 (3.7%)
White matter hyperintensities (Fazekas grade ≥2)	326 (32.8%)
Microbleeds present*	63 (18.6%)

S.D.: standard deviation.

*Gradient echo imaging to allow determination of the presence of microbleeds was available in 339 (34.1%) of patients

4.2.3 Gene and transcript selection

DNA samples were quantified, diluted and sequenced using the ThromboGenomics platform at the Cambridge Translational Genomics (CATGO) laboratory, part of the Cambridge Biomedical Research Centre. This platform and analysis pipeline have been described in detail by Simeoni et al.[197] This pipeline identifies variants in 96 genes related to bleeding and platelet disorders. In addition, 15 SVD-related genes selected based on peer-reviewed literature, as listed in Table 4-2, were added to this platform. These 15 genes were associated with familial cerebral small vessel disease (SVD) or its related phenotypes, such as familial hemiplegic migraine (FHM), cerebral amyloid angiopathy (CAA) or connective tissue diseases known to cause stroke. For each gene, we identified the main transcript on which to report variants based on peer-reviewed literature, and submitted these to the Locus Reference Genomic database (LRG) (hhttp://www.lrg-sequence.org/).[246]

The platform was designed to target (capture) all exonic regions, 5' and 3' untranslated regions (UTRs, often regulatory regions) and 50bp into each intron. In addition, 1000 base pairs (bp) upstream of the gene start sites and all variants listed in the Human Gene Mutation Database (HGMD) for those genes were targeted. HGMD is a curated database of genetic variants, or 'mutations', known to be associated to a disorder. They are classified as disease-causing mutations (DM), possibly disease-causing mutations (DM?), disease-associated polymorphisms (DP), disease-associated polymorphisms with functional evidence (DFP) or as variants not associated with disease.

Table 4-2: *SVD-related candidate genes included on the ThromboGenomics sequencing panel.*

Category	Disease	Inheritance	Gene symbol (HGNC)	RefSeq Transcript
Familial SVD	CADASIL	Dominant	NOTCH3	NM_000435
	CARASIL	Dominant and recessive	HTRA1	NM_002775
	COL4A1-related SVD	Dominant	COL4A1	NM_001845
	COL4A2-related SVD	Dominant	COL4A2	NM_001846
	FOXC1-related SVD/ Axenfeld-Rieger Anomaly	Dominant	FOXC1	NM_001453
	Retinal Vasculopathy with Cerebral Leukodystrophy and Systemic manifestations (RVCL-S)	Dominant	TREX1	NM_033629
	Fabry Disease	Dominant X-linked	GLA	NM_000169
Familial CAA	Hereditary Cerebral Haemorrhage with Amyloidosis (HCHWA-Dutch, Flemish, Italian, Piedmont, Iowa, Arctic)	Dominant	APP	NM_000484
	Familial British Dementia	Dominant	ITM2B	NM_021999
	HCHWA-Icelandic	Dominant	CST3	NM_000099
FHM	FHM Type 1	Dominant	CACNA1A	NM_001127221
	FHM Type 2	Dominant	ATP1A2	NM_000702
	FHM Type 3	Dominant	SCN1A	NM_001165963
Connective tissue diseases associated with stroke	Ehlers-Danlos Type IV	Dominant	COL3A1	NM_000090
	Pseudoxanthoma elasticum	Recessive	ABCC6	NM_001171

4.2.4 Bioinformatics

Reads were aligned to the human reference genome build GRCh37 using Burrows-Wheeler Aligner (BWA) 0.7.10.[247] On-target coverage was assessed using BEDtools 2.22.0.[248] Single nucleotide variants (SNVs) and short insertions or deletions (indels) were called using GATK 3.3 HaplotypeCaller,[249] and large copy number variants (CNVs) were called using a custom pipeline based on ExomeDepth 1.1.10.[250] To call CNVs in a sample, ExomeDepth finds the read depth in regions of interest and compares the sample to an optimized reference set of 10 other samples measured in the same batch. ExomeDepth's methods are customized to define a reference set and calculate a number of reads per targeted region. To be able to detect smaller CNVs, the large targeted regions were divided into chunks of no greater than 500bp. Deletions were assigned a homozygous status if the observed/expected read ratio was less than 0.1, and heterozygous if the ratio was between 0.1 and 1 (exclusive). Duplications of one extra copy were called if the ratio was <1.7, and more than one additional copy if the ratio was >1.7. A Bayes factor was calculated for each CNV, reflecting the degree of the confidence of the call.

Samples from 993 SVD patients were processed in 3 different batches. In each batch, the CNVs were called for the autosomal chromosomes using 10 non-SVD reference samples of any sex, and for CNVs in the X chromosome gene (GLA), these were assessed using reference samples of the same sex.

4.2.5 Filtering of variants

4.2.5.1 Single Nucleotide Variants (SNVs) and Indels

All variants within 15bp up- and down-stream of exons or any region for which a bait was designed were interpreted. Two levels of filtering were applied to SNVs and indels – a relaxed and a stringent filter. Using the relaxed filter, all SNVs and indels called were filtered based on their predicted impact on the genes (Ensembl 75 dataset), their presence in HGMD Pro (2016.3) or locus-specific databases (LSDBs), and their minor allele frequencies (MAFs) in the genome Aggregation Database (gnomAD), a database of 123,136 whole exome sequences and 15,496 whole genome sequences curated by the Broad Institute (http://gnomad.broadinstitute.org/) using SnpEff 4.0.[251] Variants with a MAF of <0.1 or >0.9

within the dataset, and which are present in HGMD or other locus-specific databases were retained. A summary of criteria is provided in Table 4-3**.**

A more stringent filter was also applied on the data. All SNVs and indels called were filtered based on their minor allele frequency in the genome Aggregation Database (gnomAD). Variants with a minor allele frequency of less than 1 in 10^{-4} were included in the analysis. Combined Annotation Dependent Depletion (CADD) scores, computed using an *in silico* predictive tool which integrates multiple annotations to provide an estimate of deleteriousness (http://cadd.gs.washington.edu), were also computed. Variants with a CADD score of 15 and above, and which were presumed to cause disease by haploinsufficiency were allowed to pass this stringent filter. A summary of this filter is provided in Table 4-3.

Table 4-3: *Filters for single nucleotide variants and indels called on ThromboGenomics.*

RELAXED FILTER

Fulfilling all of the following criteria:

- AF <0.1 or AF >0.9 in ThromboGenomics
- Is at the same position as previously described in HGMD or variant identified in another locus-specific database
- AF <0.025 or >0.975 or not present in gnomAD

or

Fulfilling all of the following criteria:

- Has a predicted moderate or high impact on splice-region on translation in the selected gene transcript according to SnpEff
- AF <0.001 or >0.99 in gnomAD, or not present in gnomAD
- Has <4 alternate alleles (to guard against errors in repetitive regions)

STRINGENT FILTER

Fulfilling all of the following criteria:

- MAF < 0.0001 in gnomAD
- CADD score ≥ 15
- Presumed to cause disease by haploinsufficiency (i.e. known autosomal dominant inheritance only)

Variants were compared with those reported in literature, as well as on ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/), an archive of genomic variation and its relationship to disease, and the DatabasE of genomiC varIation and Phenotype in Humans using Ensembl Resources (DECIPHER)[2], a web-based database of genomic variants.[252]

Copy Number Variants

All CNVs within the candidate gene regions with a Bayes factor of 20 and above were included in analysis.

4.3 Results and discussion

4.3.1 Sequencing quality control

To identify low quality regions to be excluded from the analysis, the sequencing coverage profile of the ThromboGenomics platform was assessed in each batch of samples. Individual coverage profiles for each gene showed that targeted regions of transcripts were well-covered (>100X), beyond a level sufficient for sensitive variant calling (20X), with an exception in the ABCC6 gene where the 5th percentile showed low coverage in the 3' untranslated region (UTR). This is likely due to the 3'UTR of ABCC6 not being included in the target region during the development of capture probes, possibly because there are a high number of repetitive regions in the region. Examples of coverage plots are shown in Figure 4-1.

[2] This study makes use of data generated by the DECIPHER community. A full list of centres who contributed to the generation of the data is available from http://decipher.sanger.ac.uk and via email from decipher@sanger.ac.uk. Funding for the project was provided by the Wellcome Trust.

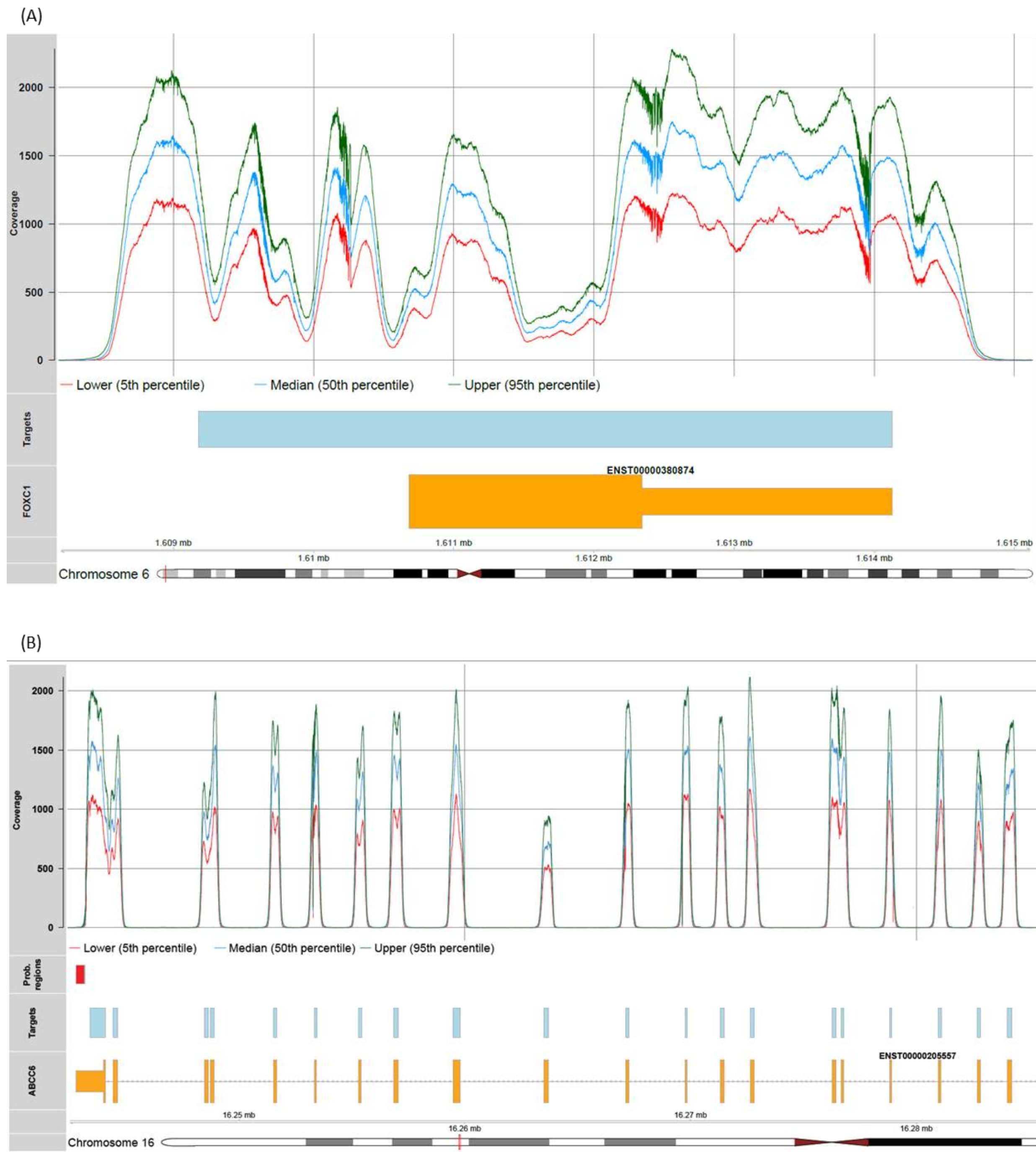


Figure 4-1: *Example of a coverage profile of (A) the FOXC1 gene, showing good coverage (>500X) across all targeted regions of the transcript, and (B) a section of the ABCC6 gene, showing poor coverage (<20X) at the 3' UTR region which has not been included as a target in the design of capture sequences. The coverage is shown for the 5th percentile (red line), 50th percentile (blue line) and 95th percentile (green line). The regions targeted by the capture array are represented in blue ('targets') and the Ensembl transcript model in yellow. The problematic region in ABCC6 is highlighted in red. ABCC6 is on the reverse stand, and is thus read from right (5'UTR) to left (3'UTR).*

4.3.2 Copy Number Variants (CNVs)

A total of 50 CNVs (34 deletions and 13 duplications) were called in 102 individuals. Excluding CNVs with a Bayes' factor of less than 20, 17 CNVs (11 deletions and 6 duplications) in 44 individuals (4.4%) remained. All CNVs were in a heterozygous state, however 6 individuals were found to have more than one CNV as shown in Table 4-4. Two patients were compound heterozygous for deletions in SCN1A, affecting both exons 1 to 8 or 1 to 9 including the promoter, and exons 17 to 25 of the gene. Individuals with both CNVs and SNVs are discussed in the section under SNVs.

A duplication spanning 45 genes including COL4A1 and COL4A2 in the 13q33-q34 locus has previously been described in a female singleton with early-onset SVD.[87] This individual presented at the age of 44 with diffuse white matter hyperintensities, multiple microbleeds and multiple subcortical lacunar infarcts, as well as vertebrobasilar dolichoectasia, and there was no clear family history of SVD. In this cohort we identified a duplication of exons 2 to 52 of COL4A1 and exons 3 to 48 of COL4A2 in a male presenting at the age of 29 with a lacunar stroke, having had only a history of cigarette smoking in terms of his cardiovascular risk factors. He had only mild white matter disease, and no microbleeds were identified on MRI.

__Table 4-4:__ Individuals with more than one CNV in the 15 candidate genes screened.

ID	Sex	Age	Clinical features	Imaging features	Family history	CNV coordinates (GRCh37)	CNV type	Gene	Exons affected	Remarks
UK02-026	M	56	Untreated hypertension Former smoker	Isolated lacunar infarct Fazekas score 0 No gradient echo performed	None	chr16:16305992-16306134	Deletion	ABCC6	6 of 31	
						chr16:16313384-16313589	Deletion	ABCC6	4 of 31	
UK03-005	M	56	Untreated hypertension Untreated hyperlipidaemia Smoker BMI 33.8	Isolated lacunar infarct Fazekas score 0 No gradient echo performed	None	chr16:16302564-16306134	Deletion	ABCC6	6 and 7 of 31	
						chr2:166852486-166856328	Deletion	SCN1A	22 – 24 of 26	
UK03-006	M	54	Untreated hypertension Untreated diabetes Former smoker BMI 35	WMH Fazekas score 2 No gradient echo performed	Father – stroke at 51	chr2:166850621-166872280	Deletion	SCN1A	1 – 8 of 26 Promoter	
						chr2:166904111-166931663	Deletion	SCN1A	17 – 25 of 26	
UK03-002	F	57	Treated hypertension Untreated hyperlipidaemia	WMH Fazekas score 3 No gradient echo performed	None	chr2:166903246-166931663	Deletion	SCN1A	1 - 9 of 26 Promoter	
						chr2:166850621-166872280	Deletion	SCN1A	17 - 25 of 26	

ID	Sex	Age	Clinical features	Imaging features	Family history	CNV coordinates (GRCh37)	CNV type	Gene	Exons affected	Remarks
UK29-004	M	28	Untreated hyperlipidaemia Smoker Depression	Isolated lacunar infarct Fazekas score 0 No microbleeds on gradient echo	None	chr13:110801276-110959240	Duplication	COL4A1	2 – 52 of 52	
						chr13:110960366-111165394	Duplication	COL4A2	3 – 48 of 48	
UK32-008	M	47	Untreated hyperlipidaemia Former smoker	Isolated lacunar infarct Fazekas score 0 No gradient echo performed	Mother – stroke at 40	chr2:166848511-166931663	Duplication	SCN1A	1 – 26 of 26 Promoter	Also has an SNV in FOXC1
						chr2:189838116-189877481	Duplication	COL3A1	1 – 51 of 51 Promoter	

One individual was found to have a duplication of the entire length of the ABCC6 gene (*chr16:16243414-16276815)* including its promoter. Homozygous or compound heterozygous mutations in ABCC6 are associated with pseudoxanthoma elasticum, an autosomal recessive connective tissue disease. Duplication of this gene has not previously been associated with disease, while a deletion of the whole gene has been found in a compound heterozygous patient with pseudoxanthoma elasticum.[253] Duplications are challenging to interpret: depending on the insertion of the duplication, this variant may or may not disrupt the expression of the gene, and expression of the duplication may not have clinical impact. Functional studies such as RNA sequencing are necessary to determine the impact of the duplication on gene function.

Four individuals were found to have a deletion of exon 23 to 29 of the ABCC6 gene. This deletion has previously been found to be present in a high frequency of Caucasian patients with pseudoxanthoma elasticum, where these individuals also carried another missense or nonsense mutation on the other ABCC6 allele. In these four patients two were also found to have missense variants in the ABCC6 gene, as discussed in the section on SNVs (Table 4-14). The deletion is thought to be mediated by Alu repeats (AluSX) occurring at a high frequency in this region of the gene.[254] Overall, the ABCC6 gene is highly susceptible to CNVs, and this is thought to be due to its high proportion of long and short interspersed nuclear elements which provide sites for recombination, such as the Alu repeats.[253] This repetitiveness also complicates the calling of CNVs. For example, the two CNVs called in individual UK02-026 affecting exons 4 and 6 may in fact be the same CNV that has been poorly called (Table 4-4).

Two other regions of CNVs were also found in multiple patients. The first is a region spanning exon 22 and 23 in SCN1A, deletions of which were identified in three individuals albeit of varying lengths. The second is a region spanning exon 1 to 8 in SCN1A which was found in two individuals. Autosomal dominant mutations in SCN1A are associated with Familial Hemiplegic Migraine Type 3.

CNVs identified in more than one patient are summarised in Table 4-5.

Table 4-5*: CNVs appearing in more than one individual.*

CNV	CNV Type	Gene affected	Exons affected	Number of individuals	Remarks
chr2:166903246-166931663	Deletion	SCN1A	1 – 9 of 26 Promoter	1	Overlapping deletion
chr2:166904111-166931663	Deletion	SCN1A	1 – 8 of 26 Promoter	1	
chr2:166850621-166872280	Deletion	SCN1A	17 – 25 of 26	2	Overlapping deletion
chr2:166852486-166856328	Deletion	SCN1A	22 – 24 of 26	1	
chr2:166854516-166856328	Deletion	SCN1A	22 – 23 of 26	1	
chr16:16248454-16259803	Deletion	ABCC6	23 to 29 of 31	4	
chr16:16302564-16306134	Deletion	ABCC6	6 – 7 of 31	3	
chr16:16305992-16306134	Deletion	ABCC6	6 of 31	12	
chr16:16313384-16313589	Deletion	ABCC6	4 of 31	11	

Table 4-4 footnotes: This table includes CNVs which appear in only one individual, but overlap with that found in another individual.

4.3.3 Single Nucleotide Variants (SNVs) and Indels

SNVs and indels were subject to two levels of filtering. The relaxed filter resulted in 17358 SNVs filtered in 909 individuals across the 15 candidate genes. A summary of the number of each type of variant identified on the relaxed filter is provided in Table 4-6 . The stringent filter resulted in 87 SNVs called in 93 individuals.

Sixteen variants were identified with a high MAF on application of the relaxed filter (>0.9 in ThromboGenomics and >0.975 or absent in gnomAD, or >0.99 or absent in gnomAD). These imply that a minority of individuals did not carry the identified minor variant that has been called as the major variant in the reference genome. These variants, the majority of which are intronic in ABCC6, are described in Table 4-7.

Table 4-6: *Variants called in each gene in 993 SVD cases where a relaxed filter has been applied.*

Gene	Stop-gained	Missense	Intron	Splice donor/ acceptor	Upstream	3'UTR	Synonymous	Frameshift	Inframe deletion	Inframe insertion	Splice region	Total
ABCC6	1	31	18	0	0	0	8	2	0	0	4	60
APP	0	9	2	0	3	0	1	0	0	0	1	15
ATP1A2	0	8	4	0	0	0	0	0	0	0	2	12
CACNA1A	0	34	14	0	0	0	6	0	4	3	5	62
COL3A1	0	16	2	0	0	0	3	0	0	0	3	21
COL4A1	0	19	4	0	1	0	0	0	0	0	5	23
COL4A2	0	21	3	0	1	0	0	1	0	0	3	24
CST3	1	0	0	0	0	0	0	1	0	0	0	2
FOXC1	0	3	0	0	1	1	1	1	4	3	0	13
GLA	0	3	0	0	1	0	2	0	0	0	0	6
HTRA1	1	10	1	0	0	0	1	0	0	0	2	13
ITM2B	0	3	0	0	0	0	0	0	0	0	0	3
NOTCH3	0	50	6	1	0	0	2	0	0	1	9	59
SCN1A	0	15	5	0	2	0	5	0	0	0	2	27
TREX1	0	3	0	0	0	0	0	1	1	0	0	5

Table 4-7: *Single nucleotide variants with a high MAF identified on the relaxed filter.*

dbSNP ID	Impact	Gene	Nucleotide change	Amino acid change	No. of SVD patients without this variant (%)
rs11866320	Intron	ABCC6	c.2787+62T>C		43 (4.3)
rs1516446	Missense	COL3A1	c.4059T>G	p.His1353Gln	43 (4.3)
rs2051504	Intron	APP	c.1090+194G>A		42 (4.2)
rs2672586	Intron Splice site	HTRA1	c.1274+8G>a		41 (4.1)
rs2820581	Intron	ATP1A2	c.381+16C>T		42 (4.2)
rs2854246	Intron	ATP1A2	c.13-22T>C		42 (4.2)
rs536174	Missense	COL4A1	c.1663A>C	p.Thr555Pro	42 (4.2)
rs566839	Upstream	SCN1A			42 (4.2)
rs6416668 **rs111973031**	Missense	ABCC6	c.2542A>G	p.Met848Val	42 (4.2)
rs7193932	Intron	ABCC6	c.1432-48G>A		42 (4.2)
rs7194043	Intron	ABCC6	c.2995+142C>T		42 (4.2)
rs7199864	Intron	ABCC6	c.1432-215A>G		43 (4.3)
rs7201980	Intron	ABCC6	c.2788-127A>G		43 (4.3)
rs7500834	Synonymous	ABCC6	c.2400A>G	p.Gly800Gly	43 (4.3)
rs8043704	Intron	ABCC6	c.2788-388A>C		240 (24.2)
rs9932889	Intron	ABCC6	c.1431+325G>A		45 (4.5)

Individuals with both CNVs and SNVs

Compound heterozygous individuals with both CNVs and SNVs in the ABCC6 gene have previously been identified in patients with pseudoxanthoma elasticum. Five patients were found to have both CNVs and SNVs passing the relaxed filter in ABCC6, and these are discussed in a later section of this chapter under 'ABCC6'.

No individuals were found to have both a CNV and an SNV passing the stringent filter in the same gene. However, with the stringent filter, four individuals were found to carry SNVs in a separate gene, in addition to the CNVs described earlier, as summarised in Table 4-8.

One individual was found to have duplications of the entire lengths of the SCN1A and COL3A1 genes, as well as a frameshift variant at the 381st amino acid residue of FOXC1. Heterozygous frameshift variants at positions 34 and 274 have previously been associated with SVD.[91] All 44 individuals with CNVs passing the Bayes factor filter were also found to carry at least one SNV passing the relaxed SNV filter.

One other individual with a deletion in exon 1 of the ABCC6 gene was also found to carry a novel COL4A2 missense variant (p.Trp1512Arg) in the NC1 domain of the protein. This region of the gene is thought to be responsible for tropocollagen stabilisation.[255] The nearest heterozygous variant previously associated with SVD in this region is p.Ala1690Thr,[159] also in the NC1 domain.

Table 4-8: Individuals with both CNVs and SNVs (passing the stringent filter) in candidate genes

ID	Sex	Age	Risk factors and clinical features	Imaging features	Family history	CNV gene	CNV type	CNV (GRCh37)	SNV gene	SNV impact	SNV	Remarks
UK07-026	M	58	Untreated hyperlipidaemia, Current smoker BMI 32.5	Isolated lacunar infarct Fazekas 0 No microbleeds	Mother – stroke at 65	ABCC6 Exon 6	Deletion	chr16:16305992-16306134	FOXC1	Inframe insertion	c.204_205insGCCGCA p.Thr68_Pro69insThrProGln	Novel variants
UK11-039	M	56	Treated hypertension Untreated hyperlipidaemia Smoker	Isolated lacunar infarct Fazekas 1 No microbleeds	Mother – stroke (age unknown)	ABCC6 Exon 6	Deletion	chr16:16305992-16306134	CACNA1A	Missense	c.6839G>T p.Arg2280Leu	Novel variants
UK32-008	M	47	Untreated hyperlipidaemia Former smoker	Isolated lacunar infarct Fazekas 0 No gradient echo performed	Mother – stroke at 40	SCN1A Exon 1-26	Duplication	chr2:166848511-166931663	FOXC1	Frameshift	c.1141_1142insG p.Ala381_Gly382fs	Novel variants. Frameshift variants associated with SVD.
						COL3A1 Exon 1 - 51	Duplication	chr2:189838116-189877481				
UK85-011	F	53	Untreated hyperlipidaemia BMI 37.2	Isolated lacunar infarct Fazekas 1 No gradient echo performed	None	ABCC6 Exon 1	Deletion	chr16:16317234-16317371	COL4A2	Missense	c.4534T>C p.Trp1512Arg	Novel variants. COL4A2 missense variant in NC1 domain

Table 4-8 footnotes: Ages specified are the age at onset of the presenting stroke, and risk factors specified are at the time of the presenting stroke. Family history was collected for first degree relatives (parents, offspring and siblings) only.

NOTCH3

Cysteine-altering variants within the EGFr region (exons 2 to 24) of the NOTCH3 gene are known to be associated with CADASIL. Five individuals were previously found to have five different CADASIL-causing mutations on screening of exons 3, 4, 5, 6, 11, 18, 19 and 22,[42] and these individuals and variants were again identified in this study. However, there were six additional individuals also identified to carry cysteine-altering variants – of which two (p.Arg169Cys and p.Cys76Tyr) were in exons that had initially been screened on both DHPLC and Sanger sequencing.

One of the variants, a cysteine-altering variant in exon 22 (p.Cys1222Gly) which was also detected in the previous study, did not pass the stringent filter as it was present at a minor allele frequency of 1.1×10^{-4} in gnomAD, but was identified on relaxed filtering. This variant has previously been reported as a pathogenic mutation in CADASIL,[256] however it was also identified in a significant proportion of non-SVD controls on whole genome sequencing (Chapter 5).

In total, 11 individuals were found to be heterozygous for CADASIL-causing mutations in the NOTCH3 gene, giving an overall diagnostic yield of 1.1%. Six of these individuals had a first-degree relative with a history of stroke, and seven had white matter hyperintensities on MRI (Fazekas score of 2 and above). The majority of these patients had multiple cardiovascular risk factors at the time of their presenting stroke. There were no other insertions or deletions resulting in a gain or loss of a cysteine variant found in this cohort. The two individuals who did not have confluent white matter hyperintensities carried variants in the latter part of the EGFr region (p.Cys1119Tyr and p.Cys1222Gly in exons 20 and 22), lending further support to the hypothesis that mutations in this part of the protein may result in a milder clinical phenotype.[59] There were no other variants (non-cysteine-altering) which passed the stringent filter. The clinical features of, and mutations identified in these 11 individuals are summarised in Table 4-9.

Table 4-9: *NOTCH3 cysteine-altering variants identified on both relaxed and stringent filters*

ID	Sex	Age	Risk factors	Imaging features	Family history	Variant	Exon	dbSNP ID	Filter passed	MAF in gnomAD	Remarks	References
UK81-002	M	61	Treated hypertension Treated hyperlipidaemia Treated diabetes Former smoker BMI 34	WMH Fazekas score 3 Bilateral anterior temporal pole involvement No gradient echo performed	None	c.227G>A p.Cys76Tyr	3	-	Relaxed Stringent	-	Not identified on DHPLC/ Sanger	[59] p.Cys76Arg[245] and p.Cys76Trp[257,258] reported. Variant also identified on WGS of the same patient (Chapter 5)
UK16-059	M	46	Untreated hyperlipidaemia Smoker BMI 30 Migraine with aura	WMH Fazekas score 3 No gradient echo performed	Father – stroke at 36	c.505C>T p.Arg169Cys	4	rs28933696	Relaxed Stringent	-	Identified on DHPLC/ Sanger	[57,259]
UK18-002	M	47	Untreated hyperlipidaemia Alcohol excess BMI 30	WMH Fazekas score 3 Minor left anterior temporal pole involvement No gradient echo performed	Father – stroke at 42 Sib – stroke at <57						Not identified on DHPLC/ Sanger	

ID	Sex	Age	Risk factors	Imaging features	Family history	Variant	Exon	dbSNP ID	Filter passed	MAF in gnomAD	Remarks	References
UK18-010	M	58	Treated hypertension Treated hyperlipidaemia Smoker Second stroke at 62	WMH Fazekas score 3 No gradient echo performed	Mother – stroke at 70	c.619C>T p.Arg207Cys	4	-	Relaxed Stringent	8.1×10^{-6}	Identified on DHPLC/ Sanger	[260]
UK05-049	M	35	Smoker Alcohol excess	WMH Fazekas score 2 No microbleeds on gradient echo	None	c.967T>A p.Cys323Ser	6	-	Relaxed Stringent	-	Identified on DHPLC /Sanger	Internal CADASIL database
UK39-010	F	67	Untreated hypertension Treated hyperlipidaemia Alcohol excess BMI 45 Depression Migraine with aura TIA	WMH Fazekas score 3 Bilateral anterior temporal pole involvement	Mother – stroke at 60	c.1162T>C p.Cys388Arg	7	-	Relaxed Stringent	-	Not previously screened	[261]
UK05-058	F	65	Untreated hypertension Untreated hyperlipidaemia Smoker	WMH Fazekas score 3 No microbleeds on gradient echo	None	c.1759C>T p.Arg587Cys	11	-	Relaxed Stringent	3.6×10^{-5}	Identified on DHPLC /Sanger	[259,262]

ID	Sex	Age	Risk factors	Imaging features	Family history	Variant	Exon	dbSNP ID	Filter passed	MAF in gnomAD	Remarks	References
UK15-006	M	53	BMI 35	Isolated lacunar infarct Fazekas score 1 No microbleeds on gradient echo	Father – stroke at 77	c.3356G>A p.Cys1119Tyr	20	-	Relaxed Stringent	8.1×10^{-6}	Not previously screened	Internal CADASIL database
UK18-021	M	50	Untreated hyperlipidaemia Current smoker	Multiple lacunar infarcts Fazekas score 1 No gradient echo performed	Mother – stroke at 59	c.3664T>G p.Cys1222Gly	22	rs199638166	Relaxed	1.1×10^{-4}	Not identified on DHPLC	[256]
UK25-009	F	67	Treated hypertension Untreated hyperlipidaemia Smoker BMI 29	WMH Fazekas score 3 Microbleeds on gradient echo	None						Not identified on DHPLC	
UK28-024	M	55	Untreated hypertension Treated hyperlipidaemia Treated diabetes Smoker TIAs	WMH Fazekas score 2 No microbleeds on gradient echo	None						Identified on DHPLC	

Ages provided are that at the time of first lacunar stroke. Family history was collected for first degree relatives (parents, offspring and siblings) only.

On relaxed filtering, and excluding cysteine-altering variants in exons 2 to 24, 144 individuals (14.5%) carried 51 unique variants passing filters in the NOTCH3 gene respectively. All individuals were heterozygous for these variants. Two patients each carried two NOTCH3 variants (p.Pro496Leu/p.His1235Leu), however one of the variants (p.Pro496Leu) has a minor allele frequency of 1.1×10^{-2} in gnomAD. Two individuals shared a disruptive inframe insertion (p.Arg12_Arg13insArgArgArg), however this does not introduce or remove a cysteine residue.

While only cysteine-changing mutations are thought to be disease-causing, it has been suggested that other non-cysteine altering variants may be associated with disease.[263] However, these associations are often not replicated, are in patients who have not had the entire EGFr region of the gene sequenced, and have not had the diagnosis of CADASIL confirmed on post-mortem analysis or skin biopsy. In this study, one of these variants, c.509A>G (p.His170Arg) was identified in 9 patients. This variant has been also been described as a polymorphism[57] and is present in a significant number of individuals in control databases (Chapter 5), with a frequency of 1.7×10^{-3} in gnomAD, indicating that this variant is not pathogenic for CADASIL.

HTRA1

Missense and nonsense HTRA1 mutations have been reported in Cerebral Autosomal Recessive Arteriopathy with Subcortical Infarcts and Leukoencephalopathy (CARASIL).[61] More recently, heterozygous missense and nonsense mutations in HTRA1 have also been identified in patients with an autosomal dominant form of SVD.[67,174] The HTRA1 gene encodes for the high temperature requirement serine protease A1, and is thought to cause a vasculopathy via the loss of inactivation of the TGFβ pathway.[136]

Eight missense and 1 nonsense variants in the HTRA1 gene were identified in 12 heterozygous individuals (1.2%) on stringent filtering. There were no compound heterozygotes or homozygotes. Two variants had previously been associated with disease: p.Arg302Ter has been reported in both autosomal recessive and autosomal dominant disease,[61,264] while p.Asp450His was reported in autosomal dominant disease.[67] p.Arg302Ter has been demonstrated to result in a mutant protein with 21 to 50% of normal protease activity.[61] These variants were each identified in one patient.

Five novel variants were found in the trypsin domain of the protein, with one variant (p.Arg227Trp) appearing in four individuals. This variant is within the trypsin domain of HTRA1, close to the trypsin active site in position 220. A summary of the clinical features and heterozygous HTRA1 variants identified is provided in Table 4-10, and a plot of the position of each variant on a linearized HTRA1 protein molecule is provided in Figure 4-2.

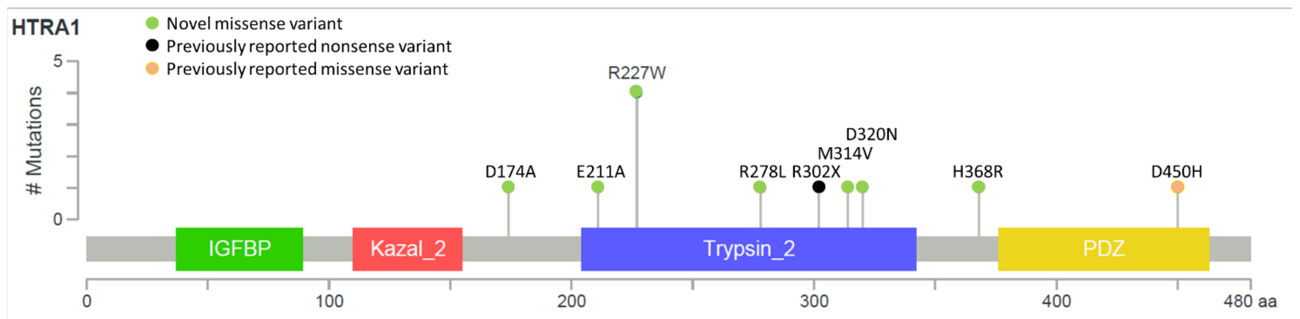


Figure 4-2: *Heterozygous HTRA1 variants passing the stringent filter, identified in 12 individuals.*

Table 4-10: Heterozygous HTRA1 variants identified on stringent filtering

<table>
<tr><th>ID</th><th>Sex</th><th>Age</th><th>Risk factors</th><th>Imaging features</th><th>Family history</th><th>Variant</th><th>Protein domain</th><th>dbSNP ID</th><th>MAF in gnomAD</th><th>Remarks</th><th>Ref</th></tr>
<tr><td>UK05-073</td><td>F</td><td>68</td><td>Treated hyperlipidaemia
Smoker</td><td>Isolated lacunar infarct
Fazekas score 1
No microbleeds on gradient echo</td><td>None</td><td>c.521A>C
p.Asp174Ala</td><td>Between Kazal and Trypsin</td><td>rs763221423</td><td>4.1×10^{-6}</td><td>Novel variant</td><td></td></tr>
<tr><td>UK05-083</td><td>M</td><td>56</td><td>Untreated hyperlipidaemia
Smoker
Alcohol excess
Depression</td><td>Multiple lacunar infarcts
Fazekas score 1
No microbleeds on gradient echo</td><td>None</td><td>c.632A>C
p.Glu211Ala</td><td>Between Kazal and Trypsin</td><td>-</td><td>-</td><td>Novel variant</td><td></td></tr>
<tr><td>UK17-018</td><td>M</td><td>59</td><td>Treated hypertension
Treated hyperlipidaemia
Treated diabetes</td><td>WMH
Fazekas score 3
No gradient echo performed</td><td>None</td><td rowspan="2">c.679C>T
p.Arg227Trp</td><td rowspan="2">Trypsin</td><td rowspan="2">rs151165026</td><td rowspan="2">7.9×10^{-5}</td><td rowspan="2">Novel variant</td><td rowspan="2"></td></tr>
<tr><td>UK10-007</td><td>M</td><td>62</td><td>Untreated hypertension
Untreated hyperlipidaemia
Former smoker
BMI 30
Migraine without aura</td><td>Isolated lacunar infarct
Fazekas score 0
No microbleeds on gradient echo</td><td>Father – stroke at 80</td></tr>
</table>

ID	Sex	Age	Risk factors	Imaging features	Family history	Variant	Protein domain	dbSNP ID	MAF in gnomAD	Remarks	Ref
UK39 -011	M	60	Treated hypertension Untreated hyperlipidaemia Alcohol excess BMI 40	WMH Fazekas score 3 No gradient echo performed	None						
UK82 -009	M	61	Treated hypertension Treated hyperlipidaemia Former smoker Alcohol excess	Isolated lacunar infarct Fazekas score 0 No microbleeds on gradient echo	None						
UK08 -001	M	50	Treated hypertension Smoker	WMH Fazekas score 3 No gradient echo performed	Father – stroke (age unknown)	c.834C>A p.Phe278Leu	Trypsin	-	-	Novel variant	
UK01 -029	M	43	Treated hypertension Treated hyperlipidaemia Treated diabetes Smoker BMI 30 Depression	Multiple lacunar infarcts Fazekas score 0 No microbleeds on gradient echo	Father – stroke at 64	c.904C>T p.Arg302Ter	Trypsin	rs113993970	1.1×10^{-5}	Reported in autosomal recessive and dominant disease	[61,264]

ID	Sex	Age	Risk factors	Imaging features	Family history	Variant	Protein domain	dbSNP ID	MAF in gnomAD	Remarks	Ref
UK82-010	M	63	Untreated hypertension Untreated hyperlipidaemia Migraine with aura	Isolated lacunar infarct Fazekas score 1 Microbleeds on gradient echo	None	c.940A>G p.Met314Val	Trypsin	rs758342671	8.1×10^{-6}	Novel variant	
UK36-002	M	41	None	Isolated lacunar infarct Fazekas score 0 No gradient echo performed	None	c.958G>A p.Asp320Asn	Trypsin	rs376449340	1.2×10^{-5}	Novel variant	
UK30-001	F	62	Untreated hyperlipidaemia BMI 36	WMH Fazekas score 2 No gradient echo performed	Mother – stroke (age unknown)	c.1103A>G p.His368Arg	Between Trypsin and PDZ	-	-	Novel variant	
UK05-059	F	46	Untreated hypertension Smoker BMI 29	Isolated lacunar infarct Fazekas score 0 No microbleeds on gradient echo	None	c.1348G>C p.Asp450His	PDZ	-	7.2×10^{-6}	Previously reported in autosomal dominant disease	[67]

Ages provided are that at the time of first lacunar stroke. Family history was collected for first degree relatives (parents, offspring and siblings) only.

On relaxed filtering only three additional variants were identified. These were p.Pro4Ser in the signal peptide domain, and p.Gln151Lys and p.Asp336Asp in the trypsin domain. p.Gln151Lys is a variant that was recently reported as a heterozygous mutation in a male singleton from a pedigree with an autosomal dominant pattern of inheritance of disease. This individual had progressive ataxic gait, dysarthria, dysphagia and cognitive decline from the age of 59, and had two affected relatives (mother and maternal uncle). His brain MRI showed white matter changes, microbleeds and atrophy.[265] In this study the variant was identified in a female with a stroke occurring at the age of 68, no family history of stroke and no white matter hyperintensities or microbleeds on MRI. This variant, occurring in the Kazal-like domain of the HTRA1 protein (and thus not near the serine protease active site) is predicted to have a moderate impact on VEP, and is rare, with a frequency of 0.00016 in gnomAD.

COL4A1 and COL4A2

COL4A1 and COL4A2 encode the α1 and α2 chains of the collagen IV protein respectively. Missense mutations in COL4A1 and COL4A2, typically but not always affecting the glycine residue in the repetitive Gly-X-Y region of the collagen chains, were first found to be associated with paediatric porencephaly and hemiplegia. These were subsequently found to be associated with cerebral small vessel disease, and ischaemic and haemorrhagic subcortical lacunar strokes.[87] Patients may also have muscular or renal manifestations of the disease.[85] Mutations affecting the C-terminal or NC1 domain of the gene, thought to be responsible for triple helix assembly,[266] have also been associated with SVD.[165]

Ten individuals (1.0%) carried 9 missense COL4A1 variants passing the stringent filter. All variants were present in the heterozygous state, and there were no compound heterozygotes. Only one of the variants affected a glycine residue in the triple helix repeat regions of the protein (p.Gly332Arg), and this variant has previously been reported to likely be benign on ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/variation/311067/). One variant (p.Pro1337Leu) was shared by two individuals. This variant is found in a low complexity region, and has also previously been reviewed on ClinVar as a likely benign variant in the context of COL4A1-related SVD (https://www.ncbi.nlm.nih.gov/clinvar/variation/311030/). Only one other variant was found in the triple-helix repeat region, and three were found in the NC1 or

C4 domain associated with the assembly of the tropocollagen molecule.[266] The remainder of the variants were found in low complexity regions.

One previously reported pathogenic variant passing the stringent filter, p.Pro352Leu, was present in one patient. This variant was previously described in a single case of presumed-sporadic intracerebral haemorrhage (ICH), and was shown on functional analyses to impair the secretion of COL4A1 into the extracellular space.[156] Two other previously reported variants (p.Arg1238Cys,[267] p.Gln1316Glu[268]), which were found in one individual each, only passed the relaxed filter. Both variants are found in a large proportion of unrelated controls on whole genome sequencing and other control database, as discussed in detail in Chapter 6. All three variants were found in low complexity regions of the protein. All variants passing the stringent filter, as well as all previously reported mutations are summarised in Table 4-11. A plot of these variants is provided in Figure 4-3.

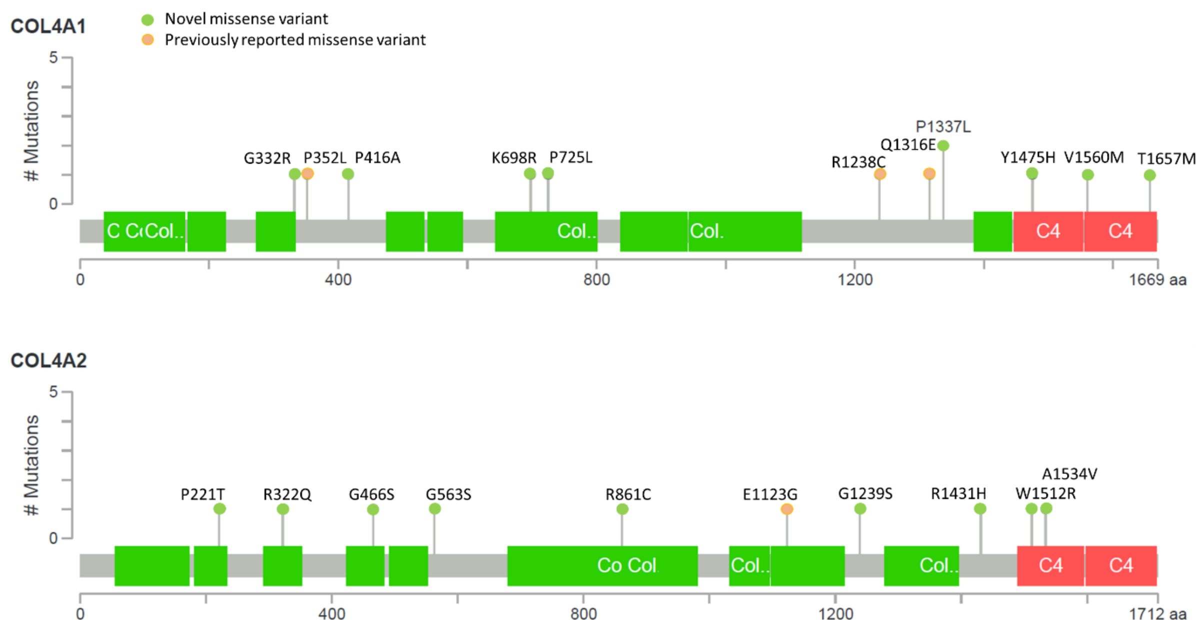


***Figure 4-3**: Missense and frameshift COL4A1 and COL4A2 variants passing the stringent filter, and COL4A1 variants previously associated with SVD.*

Nine missense COL4A2 variants in 9 individuals (0.9%) passed the stringent filter. There were no compound heterozygous or homozygous individuals. One of these variants (p.Gly466Ser) alters a glycine residue in the Gly-X-Y triple helix repeat region, and two variants were in the NC1 or C4 domain.

None of these variants passing the stringent filter had previously been reported as disease-causing, however one variant associated with disease, which passed only the relaxed filter, was identified in 34 individuals. This variant (p.Glu1123Gly) had previously been reported in a paediatric case of porencephaly[269], an adult with probable cerebral amyloid angiopathy (CAA)-related intracerebral haemorrhage (ICH), and an adult with an ICH in the basal ganglia.[159] This variant was also demonstrated to impair secretion of COL4A1 and COL4A2, and was not identified in 142 age-matched controls.[159]. This variant affects the X residue in the Gly-X-Y triple helix repeat region. In these reports there was no evidence for segregation of the variant with disease as the patients were singletons,[159] and this variant also has a high frequency in control databases (MAF 9.5 x 10^{-3} in gnomAD) beyond that expected for a highly penetrant monogenic disease as detailed in Chapter 6. It is thus possible that this variant is a polymorphism which has erroneously been described as a disease-causing mutation. Figure 4-3 illustrates the position of all variants passing the stringent filter, as well as p.Glu1123Gly. The clinical features of individuals carrying variants which passed the stringent filter are summarised in Table 4-11.

Additional variants identified in 16 individuals on relaxed filtering were scattered across the gene, with six missense variants in triple helix repeat regions, four in low complexity regions and two in the NC1 domain. One frameshift variant (p.Gly1423_Pro1424fs) was identified in a low complexity region near the NC1 domain. This variant would theoretically disrupt the NC1 domain, and thus may have implications on tropocollagen assembly.

On stringent filtering, none of the patients were found to have variants in both COL4A1 and COL4A2. One patient was found to carry a missense variant (p.Pro427Leu) in COL4A2 and a splice region variant in COL4A1 (c.652_653insCAC), with both variants passing only the relaxed filter.

Table 4-11*: COL4A1 and COL4A2 variants passing the stringent filter, and previously reported COL4A1 mutations identified in this cohort.*

Gene	ID	Sex	Age	Risk factors	Imaging features	Family history	Variant	Protein domain	MAF in gnomAD	Remarks	References
COL4A1	UK14-001	F	53	Treated hypertension Treated hyperlipidaemia Smoker Alcohol excess Depression	WMH Fazekas score 2 No gradient echo performed	Father – stroke at 60	c.994G>C p.Gly332Arg	Triple helix repeat region	4.3×10^{-5}	ClinVar: likely benign variant	ClinVar
	UK32-020	F	70	Treated Hypertension Migraine with aura	Multiple lacunar infarcts Fazekas score 1 No gradient echo performed	Mother – stroke at 70 Father – stroke at 81 Sibling – stroke at 64	c.1055C>T p.Pro352Leu	Low complexity region	6.5×10^{-5}	Previously associated with ICH, impairs COL4A1 secretion	[156]
	UK85-005	F	45	BMI 37	Isolated lacunar infarct Fazekas score 0 No gradient echo performed	None	c.1246C>G p.Pro416Ala	Low complexity region	4.9×10^{-5}	Novel variant	
	UK32-035	F	63	Untreated hypertension Untreated hyperlipidaemia Former smoker BMI 32 Depression Migraine with aura	WMH Fazekas score 2 No gradient echo performed	None	c.2093A>G p.Lys698Arg	Triple helix repeat region	5.9×10^{-5}	Novel variant	

Gene	ID	Sex	Age	Risk factors	Imaging features	Family history	Variant	Protein domain	MAF in gnomAD	Remarks	References
	UK05-032	F	55	Untreated hypertension Untreated hyperlipidaemia Smoker BMI 32	Isolated lacunar infarct Fazekas score 0 No gradient echo performed	Not known	c.2174C>T p.Pro725Leu	Triple helix repeat region	-	Novel variant	
	UK05-016	M	51	Untreated hypertension Untreated hyperlipidaemia Smoker Alcohol excess	Multiple lacunar infarcts Fazekas score 0 No gradient echo performed	None	c.3712C>T p.Arg1238Cys	Low complexity region	1.9×10^{-4}	Did not pass stringent filter. Previously associated with SVD.	[267]
	UK05-075	F	42	Smoker BMI 29	Isolated lacunar infarct Fazekas score 0 No gradient echo performed	None	c.3946C>G p.Gln1316Glu	Low complexity region	1.3×10^{-4}	Also has high frequency in WGS controls (Chapter 6)	[268]

Gene	ID	Sex	Age	Risk factors	Imaging features	Family history	Variant	Protein domain	MAF in gnomAD	Remarks	References
	UK108-003	M	42	Alcohol excess BMI 33	Isolated lacunar infarct Fazekas score 0 No gradient echo performed	None	c.4010C>T p.Pro1337Leu	Low complexity region	7.9×10^{-5}	ClinVar: likely benign variant	ClinVar
	UK28-001	F	68	Untreated hyperlipidaemia	Isolated lacunar infarct Fazekas score 0 No microbleeds on gradient echo	None					
	UK05-086	M	38	Untreated hypertension Smoker BMI 30	Isolated lacunar infarct Fazekas score 0 No gradient echo performed	Unknown on both maternal and paternal sides	c.4423T>C p.Tyr1475His	NC1 domain	1.6×10^{-5}	Novel variant	
	UK11-064	M	65	Former smoker	WMH Fazekas score 3 No gradient echo performed	Father – stroke age 60	c.4678G>A p.Val1560Met	NC1 domain	1.2×10^{-5}	Novel variant	

Gene	ID	Sex	Age	Risk factors	Imaging features	Family history	Variant	Protein domain	MAF in gnomAD	Remarks	References
	UK80-007	M	69	Former smoker Alcohol excess	WMH Fazekas score 2 No microbleeds on gradient echo	None	c.4970C>T p.Thr1657Met	NC1 domain	2.4×10^{-5}	Novel variant	
COL4A2	UK111-006	M	59	Untreated hypertension Untreated hyperlipidaemia Former smoker Alcohol excess BMI 37	WMH Fazekas score 2 No gradient echo performed	Father – stroke at 65	c.661C>A p.Pro221Thr	Triple helix repeat region	1.5×10^{-5}	Novel variant	
	UK01-028	M	58	Untreated hyperlipidaemia	WMH Fazekas score 3 No microbleeds on gradient echo	None	c.965G>A p.Arg322Gln	Triple helix repeat region	3.6×10^{-5}	Novel variant	
	UK05-055	M	65	Untreated hypertension Untreated hyperlipidaemia Smoker	Isolated lacunar infarct Fazekas score 0 No gradient echo performed	Unknown on both maternal and paternal sides	c.1396G>A p.Gly466Ser	Triple helix repeat region	5.4×10^{-5}	Novel variant	

Gene	ID	Sex	Age	Risk factors	Imaging features	Family history	Variant	Protein domain	MAF in gnomAD	Remarks	References
	UK31-001	M	56	Treated hypertension Untreated hyperlipidaemia BMI 29	Isolated lacunar infarct Fazekas score 1 No gradient echo performed	Unknown on both maternal and paternal sides	c.1687G>A p.Gly563Ser	Low complexity	-	Novel variant	
	UK32-025	M	53	Treated hyperlipidaemia Smoker BMI 40	Isolated lacunar infarct Fazekas score 0 No gradient echo performed	Father – stroke at 64	c.2581C>T p.Arg861Cys	Triple helix repeat region	2.7×10^{-5}	Novel variant	
	UK32-009	M	65	Alcohol excess	WMH Fazekas score 2 No gradient echo performed	None Unknown on paternal side	c.3715G>A p.Gly1239Ser	Low complexity	-	Novel variant	
	UK80-019	M	61	Untreated hypertension Untreated hyperlipidaemia Smoker BMI 30	Isolated lacunar infarct Fazekas score 1 No gradient echo performed	Sibling – stroke at 54	c.4292G>A p.Arg1431His	Low complexity	2.9×10^{-5}	Novel variant	

Gene	ID	Sex	Age	Risk factors	Imaging features	Family history	Variant	Protein domain	MAF in gnomAD	Remarks	References
	UK85-011	M	62	Untreated hypertension Former smoker	Multiple lacunar infarcts Fazekas score 0 No gradient echo performed	None	c.4534T>C p.Trp1512Arg	NC1 domain	6.2×10^{-5}	Novel variant	
	UK05-013	M	58	Treated hypertension Treated hyperlipidaemia Treated diabetes BMI 34 Depression	Multiple lacunar infarcts Fazekas score 0 No gradient echo performed	Father – stroke at 57. Mother – stroke at 82	c.4601C>T p.Ala1534Val	NC1 domain	1.2×10^{-5}	Novel variant	

Ages provided are that at the time of first lacunar stroke. Family history was collected for first degree relatives (parents, offspring and siblings) only.

FOXC1

The FOXC1 gene encodes for the Forkhead box C1 transcription factor. Autosomal dominant FOXC1 missense, nonsense and frameshift mutations, as well as deletion or duplication of the gene locus (6p25) have been associated with ocular abnormalities described as the Axenfeld-Rieger Syndrome.[91,270] Some of these mutations have also been found to be associated with white matter abnormalities.[91]

None of the variants reported in literature on FOXC1-associated SVD were identified in this cohort, both on relaxed or stringent filtering. We identified two variants in two individuals (0.2%) on stringent filtering: one inframe insertion (p.Thr68_Pro69insThrProGln) in a heterozygous patient who also carries a deletion in the ABCC6 gene, and one frameshift variant (p.Ala381_Gly382fs) in a heterozygous patient also carrying duplication of the COL3A1 and SCN1A genes. The inframe insertion was sited in a low complexity region near the forkhead box domain, and the frameshift variant was also in a low complexity region, some distance away from the nearest frameshift variant reported in disease at position 274. The FOXC1 gene has only 1 exon, thus this frameshift variant would conventionally be considered a disruptive, high impact variant. These cases are summarised in Table 4-8.

Eleven other variants found in 17 individuals passed the relaxed filter. One individual carried two variants – an inframe deletion (p.Arg27_Ala30del), and a missense variant (p.Pro297Ser). There were no additional frameshift variants identified. The majority of these variants were inframe deletions or insertions, however none of the variants affected the forkhead box region.

TREX1

Originally described as multiple distinct syndromes, Retinal Vasculopathy with Cerebral Leukodystrophy and Systemic Manifestations (RVCL-S) was recently found to arise due to frameshift mutations near the C-terminus of the TREX1 gene.[78] The TREX1 gene encodes the 3’ repair exonuclease, and missense mutations in this gene are also associated with Aicardi-Goutiéres Syndrome,[271] a neurological disease with paediatric onset, and familial chilblain lupus.[272]

In this study, 3 missense variants (p.Gly197Ala, p.Arg229Gly, p.Lys230Asn), one inframe deletion and one frameshift variant found in 5 individuals (0.5%) passed the stringent filter, and there were no additional variants identified on relaxed filtering. The only frameshift variant identified (p.Ala194fs) lies 41 residues upstream of the nearest previously reported frameshift variant in RVCL-S (p.Val235fs), away from the C-terminus. It was also identified in the same patient on a separate study using Sanger sequencing, and was reported as p.Ala139Valfs*21 on a different isoform.[273] The variant was found to almost completely abolish function in an assay of TREX1 nuclease activity in this study.[273] The p.Ala194fs variant was also previously reported in a paediatric case of Aicardi-Goutiéres syndrome, with cerebral calcification, white matter atrophy, global developmental delay, progressive microcephaly, congenital glaucoma as well as skin and cardiac abnormalities on the DECIPHER database (Patient 303873).[252] This child was also found to have a missense TREX1 variant (p.Arg169His), where the frameshift was paternally inherited, and the missense variant was maternally inherited.

All variants were present in a heterozygous state. A summary of these novel variants is provided in Table 4-12.

__Table 4-12__: TREX1 variants passing both stringent and relaxed filters

ID	Sex	Age	Risk factors	Imaging features	Family history	Variant	dbSNP ID	MAF in gnomAD	Remarks	Ref
UK26-015	F	48	Untreated hypertension Untreated hyperlipidaemia Untreated diabetes BMI 31 Depression	WMH Fazekas score 3 Microbleeds on gradient echo	None	c.581delC p.Ala194fs	rs763229085	2.8×10^{-5}	Reported in paediatric case of Aicardi Goutiéres syndrome (Patient 303873)	252
UK92-007	F	42	Untreated hypertension Smoker BMI 30 Depression	Isolated lacunar infarct Fazekas score 0 No gradient echo performed	Sibling – stroke at 29 Sibling 2 – stroke at 47	c.590G>C p.Gly197Ala	-	-	Novel	
UK16-010	F	56	Treated hypertension Treated hyperlipidaemia Treated diabetes BMI 40 Depression	WMH Fazekas score 2 No gradient echo performed	None	c.685A>G p.Arg229Gly	rs759481016	1.1×10^{-5}	Novel	
UK05-113	M	64	Untreated hypertension	Isolated lacunar infarct Fazekas score 1 No microbleeds on gradient echo	None	c.690G>T p.Lys230Asn	-	-	Novel p.Val235fs previously reported	78

ID	Sex	Age	Risk factors	Imaging features	Family history	Variant	dbSNP ID	MAF in gnomAD	Remarks	Ref
UK06-008	F	60	Treated hypertension Treated hyperlipidaemia BMI 32	Isolated lacunar infarct Fazekas score 0 No gradient echo performed	None	c.1024_1041del CTGCTGGCCCCA CTGGGT p.Leu342_Gly347del	rs79318303	6.1×10^{-5}	Novel p.Leu287fs previously reported	[78]

Ages provided are that at the time of first lacunar stroke. Family history was collected for first degree relatives (parents, offspring and siblings) only.

GLA

Fabry disease is an X-linked dominant metabolic storage disease caused by mutations in the GLA gene, which causes abnormal depositions of globotriaosylceramide in the cells of multiple organ systems, including the small blood vessels of the brain.[274] Six variants identified in 21 individuals passed the relaxed filter, however none of these were found on application of the stringent filter. Ten of these individuals were hemizygous males and the remainder were heterozygous females. One of these variants (p.Asp313Tyr), identified in two females and four males, was previously described in Fabry disease, and has been demonstrated in in-vitro assays to result in decreased alpha-galactosidase A enzyme activity.[275] However, this variant was previously also identified in a child whose clinical features did not fit with the diagnosis of Fabry disease,[276] and multiple submitters to ClinVar have listed this variant as 'likely benign' or of 'uncertain significance' according to various guidelines or criteria (https://www.ncbi.nlm.nih.gov/clinvar/variation/10738/).

Candidate genes associated with related phenotypes

Variants were also identified on stringent filtering across the candidate genes associated with familial hemiplegic migraine (FHM) (CACNA1A, SCN1A, ATP1A2), cerebral amyloid angiopathy (APP, ITM2B, CST3) and Ehlers Danlos Syndrome Type IV (COL3A1). All individuals were heterozygous for these variants, and there were no compound heterozygous individuals. The majority of variants were found outside of domains of the proteins previously associated with disease. A summary of findings is provided in Table 4-13.

Table 4-13: Variants passing stringent filters in candidate genes associated with related phenotypes

Related phenotype	Gene	Variant	dbSNP ID	No. of individuals	MAF in gnomAD	Remarks	Ref
Ehlers Danlos (IV)	COL3A1	c.74A>G p.Gln25Arg	-	1	-	In signal peptide	
		c.1856C>T p.Pro619Leu	rs373838193	1	4.3×10^{-5}	In low complexity region VUS on ClinVar for Aortic aneurysm/ dissection and Ehlers Danlos Type IV	https://www.ncbi.nlm.nih.gov/clinvar/variation/199718/
		c.1864C>T p.Pro622Ser	rs772638774	2	3.3×10^{-5}	In low complexity region	
		c.226A>G p.Asn76Asp	rs142045411	1	6.2×10^{-5}	In von Willebrand factor Type C domain	
		c.1165A>T p.Asn389Tyr	rs200394946	1	9.7×10^{-5}	In collagen triple helix repeat domain. VUS on ClinVar for Aortic Aneurysm/dissection and Ehlers Danlos Type IV	https://www.ncbi.nlm.nih.gov/clinvar/variation/199718/
Cerebral amyloid angiopathy	APP	c.373G>A p.Asp125Asn	-	1	-	In amyloid A4 N-terminal heparin binding domain	
		c.682G>A p.Val228Ile	rs755841034	2	4.1×10^{-6}	In low complexity region, near p.Val225Ala, a VUS associated with familial Alzheimer's disease	

Related phenotype	Gene	Variant	dbSNP ID	No. of individuals	MAF in gnomAD	Remarks	Ref
		c.727G>A p.Asp243Asn	rs750279232	1	3.3×10^{-6}	In coiled coil domain	
		c.736G>A p.Glu246Lys	rs147485129	1	1.8×10^{-5}	In coiled coil domain	
		c.1859C>T p.Pro620Leu	rs759926843	1	1.6×10^{-5}	In low complexity region	
		c.1909+5G>A Intron and splice region	rs200271509	1	1.5×10^{-5}	In coiled coil domain	
		c.2148C>T p.Ile716Ile	rs145564988	1	6.9×10^{-5}	In C-terminus domain. Ile716Thr reported on ClinVar (no clinical interpretation available). Ile716Val associated with familial Alzheimer's disease.	https://www.ncbi.nlm.nih.gov/clinvar/variation/98240/) [277]
	CST3	c.277G>T p.Glu93Ter	-	1	4.1×10^{-6}	p.Leu94Gln reported in familial Icelandic dementia	[15]
		c.277_296delGAGCTGGGCCGAACCACGTG p.Glu93fs	rs765138253	1	6.1×10^{-5}		
	ITM2B	c.92C>G p.Pro31Arg	rs150336652	1	8.9×10^{-5}	Low complexity region	

Related phenotype	Gene	Variant	dbSNP ID	No. of individuals	MAF in gnomAD	Remarks	Ref
Familial Hemiplegic Migraine	ATP1A2	c.1262G>A p.Arg421Gln	rs139499540	1	5.5×10^{-5}	Between E1-E2 and Cation ATPase domains Reported as VUS for FHM on ClinVar	https://www.ncbi.nlm.nih.gov/clinvar/variation/406194/
		c.2146G>T p.Val716Leu	-	1	-	Near metal ion binding sites at residues 714 and 718. p.Asp718Asn reported as pathogenic in FHM	[278]
	CACNA1A	c.130G>A p.Ala44Thr	rs201398669	1	2.5×10^{-5}	Low complexity region	
		c.633T>C p.Ser211Ser	rs202216404	1	6.2×10^{-5}	Ion transport domain p.Ser211Asn reported as VUS for unspecified phenotype.	https://www.ncbi.nlm.nih.gov/clinvar/variation/426949/
		c.1083-108_1083-101delCACACACA	-	1	-	Intron variant	
		c.2524G>C p.Glu842Gln	-	1	-	Between ion transport domains. p.Ala841Ser reported as VUS for unspecified phenotype.	https://www.ncbi.nlm.nih.gov/clinvar/variation/385371/
		c.2812G>C p.Gly938Arg	rs771423362	1	4.1×10^{-5}	Between ion transport domains. Reported as VUS for unspecified phenotype	https://www.ncbi.nlm.nih.gov/clinvar/variation/422044/

Related phenotype	Gene	Variant	dbSNP ID	No. of individuals	MAF in gnomAD	Remarks	Ref
		c.2944G>A p.Gly982Ser	-	1	-	Low complexity region	
		c.2960G>A p.Arg987Gln	-	1	-	Low complexity region. p.Arg987Pro reported as VUS for unspecified phenotype.	https://www.ncbi.nlm.nih.gov/clinvar/variation/387025/
		c.3226G>A p.Ala1076Thr	rs554091859	1	8.6×10^{-5}	Between ion transport domains. Reported as VUS (unspecified phenotype)	https://www.ncbi.nlm.nih.gov/clinvar/variation/429962/
		c.3265G>A p.Gly1089Ser	rs201311000	1	3.3×10^{-5}	Between ion transport domains. Gly1089Cys reported as VUS for unspecified phenotype	https://www.ncbi.nlm.nih.gov/clinvar/variation/391877/
		c.3436G>A p.Val1146Ile	rs376365775	1	1.6×10^{-5}	Between ion transport domains. Reported as VUS for unspecified phenotype	https://www.ncbi.nlm.nih.gov/clinvar/variation/195470/
		c.4646A>G p.Gln1549Arg	rs759263620	1	4.1×10^{-5}	Between ion transport domains	
		c.5494G> p.Val1832Met	rs376815942	1	1.2×10^{-5}	GPHH (voltage dependent L type calcium channel) domain	
		c.5795A>G p.Asn1932Ser	-	1	4.1×10^{-6}	Voltage gated calcium channel IQ domain	
		c.5936A>C p.Glu1979Ala	rs774657158	1	8.3×10^{-6}	Outside of Ca channel IQ domain	

Related phenotype	Gene	Variant	dbSNP ID	No. of individuals	MAF in gnomAD	Remarks	Ref
		c.6400C>T p.Arg2134Cys	rs121908235	1	6.2×10^{-5}	Outside of Ca channel IQ domain VUS for unspecified phenotype and associated with episodic ataxia type 2 (impact uncertain)	https://www.ncbi.nlm.nih.gov/clinvar/variation/68440
		c.6644A>C p.His2215Pro	-	1	3.8×10^{-5}	Low complexity region	
		c.6715C>T p.Arg2239Trp	rs759576380	1	5.2×10^{-5}	Outside of Ca channel IQ domain	
		c.6758G>A p.Arg2253Gln	rs752950486	1	1.6×10^{-5}	Outside of Ca channel IQ domain	
		c.6839G>T p.Arg2280Leu	-	1	-	Patient also has a deletion in ABCC6 Outside of Ca channel IQ domain	
		c.7027G>A p.Gly2343Ser	-	1	6.5×10^{-5}	Outside of Ca channel IQ domain	
		c.7166G>A p.Arg2389Gln	-	1	-	Outside of Ca channel IQ domain	
	SCN1A	c.136G>A p.Glu46Lys	rs769582667	1	3.2×10^{-5}	Low complexity region	
		c.1457C>G p.Ala486Gly	rs777120925	1	5.8×10^{-5}	Low complexity region	

Related phenotype	Gene	Variant	dbSNP ID	No. of individuals	MAF in gnomAD	Remarks	Ref
		c.1738C>G p.Arg580Gly	-	1	-	Cytoplasmic domain of voltage-gated Na ion channel. p.Arg580Gln reported as VUS for early infantile epileptic encephalopathy. p.Arg580Ter reported as pathogenic for severe myoclonic epilepsy in infancy	https://www.ncbi.nlm.nih.gov/clinvar/variation/167646/ [279]
		c.1843G>A p.Gly615Arg	-	1	4.1×10^{-6}	Cytoplasmic domain of voltage-gated Na ion channel.	
		c.3394G>A p.Glu1132Lys	-	1	-	Sodium ion transport-associated domain	
		c.4690C>T p.Arg1564Cys	rs121918807	1	7.7×10^{-5}	Ion transport domain	
		c.5750G>A p.Arg1917His	-	1	-	Outside of ion transport domain	

VUS: Variant of uncertain significance

ABCC6

On stringent filtering, none of the 993 individuals carried more than one ABCC6 variant, and none were homozygous for any ABCC6 variants. As a result, all ABCC6 variants were filtered out on the basis that haploinsufficiency would not be a mechanism for disease.

On relaxed filtering, 33 individuals were found to be compound heterozygotes for missense ABCC6 variants. None of these individuals carried more than one variant present at a minor allele frequency in gnomAD of less than 1×10^{-4}. Of the 60 variants detected, there were 27 missense variants present at a frequency of less than 1×10^{-4}, 8 missense, 1 frameshift and 1 intron and splice region variants were previously reported as being associated with autosomal recessive disease.

Some patients with pseudoxanthoma elasticum were previously found to carry a CNV and an SNV in the ABCC6 gene. In this cohort, 5 individuals had both CNVs and missense variants in the ABCC6 gene. Two of these individuals had CNVs and SNVs that were both previously reported as being associated with disease. These two individuals would thus have a genetic profile compatible with previously documented cases of pseudoxanthoma elasticum, although neither of these individuals were known to have clinical features of pseudoxanthoma elasticum at the time of recruitment. These findings are summarised in Table 4-14.

***Table 4-14**: Compound heterozygous individuals carrying both CNVs and SNVs (passing relaxed filter only) in the ABCC6 gene*

ID	CNV identified	Exons affected by CNV	Missense variant	Remarks	Ref
UK02-008	chr16:16305992-16306134 Deletion	6	c.3190C>T p.Arg1064Trp	Missense variant previously reported as pathogenic	[280]
UK03-008	chr16:16248454-16259803 Deletion	23-29	c.2974G>A p.Gly992Arg	Missense variant previously reported as pathogenic on ClinVar. Exon 23-29 deletion known to be associated with disease	https://www.ncbi.nlm.nih.gov/clinvar/variation/433524/ [281]
UK102-010	chr16:16313384-16313589 Deletion	4	c.3671C>T p.Thr1224Ile	Both variants novel	
UK29-002	chr16:16248454-16259803 Deletion	23-29	c.1171A>G p.Arg391Gly	Missense variant previously reported as pathogenic. Exon 23-29 deletion associated with disease.	[282] [281]
UK29-007	chr16:16302564-16306134 Deletion	6-7	c.2600C>A p.Pro867His	Both variants novel	

Individuals with SNVs in multiple genes

Two individuals were found to each have two variants passing the stringent filter in different candidate genes. One individual had a cysteine-altering variant in exon 7 of the NOTCH3 gene, which would be expected to account for all of the patient's clinical features. These variants are summarised in Table 4-15.

Table 4-15: *Patients found to have SNVs on more than one SVD candidate gene.*

ID	Gene	Variant 1	MAF in gnomAD	Gene	Variant 2	MAF in gnomAD
UK05-032	COL4A1	c.2174C>T p.Pro725Leu	-	CACNA1A	c.3436G>A p.Val1146Ile	1.6×10^{-5}
UK39-010	CACNA1A	c.5494G>A p.Val1832Met	1.2×10^{-5}	NOTCH3	c.1162T>C p.Cys388Arg	-

4.4 Conclusions

In this study of 993 patients with presumed-sporadic, early-onset SVD stroke, we have demonstrated the use of a multi-gene HTS platform to screen for rare variants associated with monogenic SVD and related phenotypes. Using this platform, we have achieved robust coverage of all coding regions of the 15 candidate genes.

This cohort of patients was previously screened for NOTCH3 mutations on a limited number of exons. In this study we have found that two mutations were initially missed on both DHPLC and Sanger sequencing, and that a total of 11 (1.1%) patients carried CADASIL-causing mutations. This provides further evidence that HTS methods, with their high coverage of bases, are more sensitive than conventional sequencing techniques.

We demonstrate that previously reported (or theoretical, in the case of CADASIL) disease-causing mutations across 15 genes associated with SVD and related phenotypes were identified in this cohort at two different levels of filtering. This includes CADASIL-causing mutations, autosomal dominant SVD-associated mutations in HTRA1 and a frameshift mutation in TREX1 previously associated with cerebral white matter disease. This suggests that single mutations may in fact account for a few cases of presumed-sporadic stroke with a

low clinical index of suspicion for monogenic disease. The caveats here are that few known mutations have been reported for each gene, and that the ThromboGenomics platform only calls variants in the coding, upstream and downstream variants. There may thus be many more monogenic disease-causing variants in the non-coding regions or other genes that are not known, but are also causative of strokes in this population.

The identification of known variants on only the relaxed filter also calls into question the pathogenicity of these variants, particularly variants which have also been found in high frequencies in control databases. We found that one cysteine-altering mutation in NOTCH3, a missense HTRA1 variant and a COL4A2 variant previously associated with autosomal dominant disease are present at higher frequencies than other known mutations and were thus excluded on more stringent filtering strategies. In contrast, a large number of variants were found to pass the stringent filter in CACNA1A. This suggests that in larger genes conventional filtering strategies may be limited in their utility to narrow down the list of possible causative variants, without reducing the risk of excluding truly pathogenic ones. In this example, additional pedigrees sharing the same phenotype and genotype, as well as large-scale functional screening may lend further evidence for pathogenicity in these variants. These issues call into question both the arbitrary filter thresholds that we have set, as well as whether these variants are truly highly penetrant disease-causing mutations.

There are a few possible explanations for the relatively high frequency of known variants in this cohort of patients. It is possible that these variants cause a milder form of disease only associated with stroke and not the other syndromic features (which would likely have led to exclusion from the study due to suspicion of familial disease). For example, although NOTCH3 mutation carriers did tend to have more severe white matter hyperintensities, this was not the case for mutation carriers in other genes, where many individuals had only isolated lacunar infarcts, a lack of a convincing clinical history (e.g. that of migraine), or a lack of a family history which would have suggested a likely familial disease. We have also identified patients carrying both CNVs and SNVs in the ABCC6 gene, all of which have previously been associated with pseudoxanthoma elasticum. None of these patients were clinically diagnosed with pseudoxanthoma elasticum or documented to carry clinical features of this disease.

While it may be that some of these variants are indeed disease-causing but result in milder phenotype in the context of co-existing cardiovascular risk factors, for some of these variants

in genes such as COL4A2 these may have been erroneously associated with the disease in the first place. This may have been due to the limited availability of controls prior to the development of large-scale control databases. One example here is the variants identified in COL4A2, which have been demonstrated on functional assays to result in impaired secretion of both COL4A1 and COL4A2 chains as well as the accumulation of protein in the endoplasmic reticulum leading to stress in the endoplasmic reticulum.[159] Two of these variants were found in large numbers of controls. Functional in-vitro assays are limited in their ability to recapitulate disease features and thus may not adequately serve as a surrogate for the true impact of variants in vivo. The heterogeneity of disease features between and within families with suspected familial SVD raises the question of whether 'pathogenic' variants are indeed present in high background frequencies in the general population, and whether these may be responsible for 'sporadic' cases of stroke. This is also discussed in Chapter 6.

There is also the possibility that carrying multiple variants (rare or otherwise) together may increase the risk for early-onset SVD. Beyond examining known pathogenic variants, future work should also investigate the hypothesis that multiple variants in different genes may contribute to disease risk in this cohort of patients, by comparing the burden of rare variants in cases and controls. However, this study is limited by the lack of a suitable control dataset apart from databases such as gnomAD. An ideal set of controls would be age- and ethnicity-matched, MRI-phenotyped to confirm the absence of stroke or cerebral disease, and sequenced on the same platform.

We have shown that the ThromboGenomics HTS platform can be useful for the simultaneous screening of multiple genes, suggesting its potential use as a diagnostic tool for patients with suspected familial forms of SVD. With milder phenotypes of monogenic disease now coming to the fore, we speculate that this platform may also be useful for those with presumed sporadic disease and a lower index of clinical suspicion for familial disease.

In this study we have found that a limited screen of the coding regions of a small number of candidate genes already raises more questions than answers, by calling multiple variants of uncertain significance. The clinical interpretation of these variants remains a significant challenge, and understanding the background carrier rate of these purported 'disease-causing mutations' will significantly boost the clinical utility of rapid sequencing tools. This question is further explored in a larger set of controls using whole genome sequencing in **Chapter 6**.

Chapter 5: Rare variants in suspected familial SVD: a candidate gene approach using whole genome sequencing

Outline

5.1 Introduction

5.2 Methods

- 5.2.1 NIHR BioResource Rare Diseases Pilot Study (BRIDGE)
- 5.2.2 Selection of candidate genes
- 5.2.3 Whole genome sequencing (WGS)
- 5.2.4 Variant filtering
- 5.2.5 Modelling of NOTCH3 variants

5.3 Results and discussion

- 5.3.1 Baseline characteristics
- 5.3.2 Copy Number Variants (CNVs)
- 5.3.3 Single Nucleotide Variants (SNVs)

5.4 Conclusions

5.5 Appendix

List of tables

Table 5-1 List of candidate genes investigated

Table 5-2 Clinical features of 118 index cases with suspected familial cerebral SVD

Table 5-3 Coding NOTCH3 variants passing filters, identified in SVD index cases

Table 5-4 Missense and nonsense HTRA1 variants passing filters, identified in SVD index cases

Table 5-5 Clinical features of probands with heterozygous missense or nonsense HTRA1 variants

Table 5-6 Coding COL4A1 and COL4A2 variants passing filters, identified in SVD index cases

Table 5-7 Clinical features of probands with COL4A1 and COL4A2 variants

Table 5-8 Coding variants passing filters in all other genes, which have not been excluded based on segregation analysis.

Table 5-9 Summary of variants identified in all 118 pedigrees.

Table 5-10 Non-coding SNVs passing filters identified in 118 probands.

List of figures

Figure 5-1	Summary of selection criteria and recruitment of participants to the BRIDGE-SVD study
Figure 5-2	Pedigree F102
Figure 5-3	MR images of the proband in Pedigree F102
Figure 5-4	Proportion of each type of variants called in candidate genes across all SVD participants which passed filters.
Figure 5-5	Pedigree F37
Figure 5-6	T2-weighted MR imaging of the proband of pedigree F37
Figure 5-7	3D modelled EGFr domain 1 of the NOTCH3 protein with an overlay of the variants identified in the proband of Pedigree F37
Figure 5-8	Heterozygous missense and nonsense variants identified in HTRA1
Figure 5-9	Pedigree F05
Figure 5-10	MR images of pedigree F05
Figure 5-11	Pedigrees F122 and F34
Figure 5-12	MR imaging of the proband in pedigrees F122 and F34
Figure 5-13	Heterozygous missense and frameshift variants identified in COL4A1
Figure 5-14	MR imaging of the proband from Pedigree F104
Figure 5-15	T2-weighted FLAIR MR imaging of the proband from Pedigree F32
Figure 5-16	Pedigree F08
Figure 5-17	MR imaging of the proband from Pedigree F08
Figure 5-18	Pedigree F18
Figure 5-19	MR imaging of the proband and her brother from Pedigree F18
Figure 5-20	Pedigree F103
Figure 5-21	MRI T2-FLAIR images of the proband from Pedigree F103

5. Rare variants in suspected familial SVD: a candidate gene approach using whole genome sequencing

5.1 Introduction

Cerebral small vessel disease (SVD) is the most common subtype of stroke to present as a familial disease. Cerebral Autosomal Dominant Arteriopathy with Subcortical Infarcts and Leukoencephalopathy (CADASIL) is the most common form of monogenic SVD, however in recent years a number of new genes have been found to also be associated with familial SVD, either through traditional linkage analysis in pedigrees, or in combination with whole exome sequencing. Some examples are COL4A1 and COL4A2-related SVD, or Retinal Vasculopathy with Cerebral Leukodystrophy and Systemic manifestations (RVCL-S). The clinical features and pathogenesis of these diseases have been discussed in Chapter 1.

Although each of these monogenic diseases are typified by distinctive neurological or non-neurological features which may guide the diagnostic process, patients often do not develop these features of the syndrome. They may instead present with non-specific features of SVD at an earlier age than expected, in the context of cardiovascular risk factors, alongside a vague family history of strokes or dementia. Hence, despite recent advances, there remain many families with suspected monogenic SVD where no known genetic cause has been identified. These patients have been described as having a NOTCH3-negative CADASIL-like syndrome.[207,283] However, testing for mutations in multiple genes one after another using traditional Sanger sequencing for mutations in known disease-associated genes can be time-consuming and costly, and often do not lead to a genetic diagnosis. In clinical practice, most clinicians only test for mutations in the NOTCH3 gene.

With the rapid decline in cost of high-throughput sequencing (HTS) over the past two decades, HTS can play a significant role in the diagnosis of rare genetic diseases, and it is likely that this technology will eventually become routinely used in the clinic. A comparison of different HTS techniques such as WGS and whole exome sequencing has briefly been discussed in Chapter 2. Whole genome sequencing (WGS) allows the simultaneous analysis of multiple candidate genes in patients with a suspected genetic disease, circumventing the diagnostic ‘odyssey’ of investigating one gene after another. In addition to the protein-coding regions, non-coding

variants such as splice-site variants are also captured with WGS. These regions form the bulk of the genome, and may also lead to disease, as seen in other collagen-related disorders.[284] Non-coding regions are not typically captured on other forms of HTS such as whole exome sequencing or targeted capture sequencing, as summarised in Chapter 4. This chapter explores the use of whole genome sequencing (WGS) in the investigation of patients with suspected familial SVD with unconfirmed genetic aetiology.

In this chapter we aim to investigate two hypotheses about this population of patients. The first is that rare, previously described monogenic causes of SVD are present in this population, and rare variants in genes identified in associated phenotypes such as familial hemiplegic migraine, cerebral amyloid angiopathy and leukodystrophy are also important in this population.

The second hypothesis is that rare variants in genes identified in genome-wide association studies (GWAS) in SVD may also play a role in this population of familial SVD. A recent meta-analysis of GWAS in sporadic stroke by the MEGASTROKE collaboration by the International Stroke Genetics Consortium has identified multiple loci associated with stroke reaching genome-wide sequence, with one locus found to be associated with the SVD subtype. GWAS analyses of SVD-associated phenotypes (white matter hyperintensities, small vessel ischaemic strokes, small vessel intracerebral haemorrhage and migraine) have also identified 12 loci at levels of significance strongly suggestive for association (unpublished data). We aim to investigate if these loci may also harbour rare monogenic disease-associated variants.

This chapter examines the prevalence of both known and novel rare variants in monogenic SVD genes in families with suspected familial SVD where no genetic diagnosis has been made. It will also investigate the presence of rare variants in other candidate genes such as those from associated phenotypes (cerebral amyloid angiopathy, familial hemiplegic migraine, connective tissue disease associated with stroke and leukoencephalopathy), and those from gene loci identified in GWAS of sporadic SVD.

5.2 Methods

5.2.1 NIHR BioResource Rare Diseases Pilot Study (BRIDGE)

Over a 24-month period from October 2014 to October 2016, patients with suspected familial SVD were identified and recruited across 12 sites in the UK, as part of the National Institute of Health Research (NIHR) BioResource Rare Disease study (BRIDGE), a multicentre whole-genome sequencing (WGS) study of approximately 10,000 patients. Affected and unaffected relatives of these patients were also recruited to the study. The term 'singleton' is used here to refer to index cases where no relatives have been recruited.

Patients with MRI-confirmed early-onset SVD were selected based on clinical suspicion of familial disease. Further details about the recruitment process and selection criteria are discussed in Chapter 4. SVD was one of fourteen rare disease categories in the study, and the list of sites, studies and overall recruitment figures are summarised in **Chapter 2**.

Funding for the project was provided by the National Institute for Health Research (NIHR, grant number RG65966). The study was approved by Research Ethics Committees in the UK and appropriate national ethics authorities in non-UK enrolment centres.

5.2.2 Selection of candidate genes

Seventy-three genes across the genome were selected for investigation. These were selected based on peer-reviewed literature, and include those previously described in monogenic SVD and cerebral amyloid angiopathy and related phenotypes such as familial hemiplegic migraine (FHM), leukodystrophy, connective tissue diseases known to cause stroke and other amyloidosis-causing genes.

Gene loci identified with strongly suggestive evidence ($p < 5 \times 10^{-6}$) of being associated with sporadic subcortical ischaemic or haemorrhagic strokes of presumed small vessel origin, white matter hyperintensities on MRI, or migraine in GWAS studies were selected as candidate regions. Genes which the identified SNPs were expression quantitative trait loci (eQTLs) of, and the two nearest genes to the lead SNP were included in the gene list. Nearby genes (within 1Mb) that were part of the Forkhead family of transcription factors were also included. The complete list of genes is summarised in Table 5-1.

__Table 5-1:__ List of candidate genes investigated

Category	HGNC ID	
Monogenic non-amyloid SVD (n=10)	• NOTCH3 • HTRA1 • COL4A1 • COL4A2 • FOXC1	• TREX1 • SNORD118 • CTSA • GLA • CECR1
Monogenic cerebral amyloid angiopathy and other amyloidosis (n=6)	• APP • ITM2B • CST3	• GSN • PRNP • TTR
Related phenotype: familial hemiplegic migraine (n=6)	• ATP1A2 • ATP1A3 • CACNA1A	• KCNK18 • PRRT2 • SCN1A
Related phenotype: leukodystrophy (n=19)	• ADAR • CSF1R • DARS2 • EARS2 • EIF2B1-5 • FAM126A • GFAP	• HEPACAM • LMNB1 • MLC1 • TUBB4A • RNASEH2A-C • SAMHD1
Related phenotype: connective tissue diseases (n=2)	• COL3A1 • ABCC6	
Sporadic SVD/WMH/migraine gene loci from GWA studies (n=30)	• BGLAP • C1QL1 • CARF • DEGS2 • EFEMP1 • EVL • FOXC2 • FOXF1 • FOXF2 • FOXJ1 • FOXL1 • FOXQ1 • HAAO • ICA1L • JPH3	• KIF18B • LOX • MIR216A • MIR216B • MIR217 • NBEAL1 • OXER1 • PMF1 • PRDM16 • SH3PXD2A • TRIM47 • TRIM65 • WDR12 • ZCCHC14 • ZNF474

5.2.3 Whole genome sequencing (WGS)

Whole genome sequencing (WGS) was performed by Illumina on HiSeq X Ten generating 150bp paired-end reads per lane with minimum coverage of 15X for at least 95% of the genome. Reads were aligned to the hg19 (GRCh37) build of the human genome using the Isaac Aligner, and variants were called using the Isaac VariantCaller.[285]

5.2.4 Variant filtering

5.2.4.1 Single nucleotide variants (SNVs)

Variants were classified according to their annotation using the Ensembl 81 Variant Effect Predictor (VEP), the Exome Variant Server ESP6500SI-V2 (evs.gs.washington.edu/EVS/) and the Human Gene Mutation Database (HGMD) 2016.2. These were then filtered according to the Combined Annotation Dependent Depletion (CADD) score, computed using an in silico predictive tool which integrates multiple annotations to provide an estimate of deleteriousness (http://cadd.gs.washington.edu). All variants with a CADD score below 15 were excluded from analysis. The remaining variants were then filtered according to their minor allele frequency (MAF) in control databases, with variants present at a frequency of greater than 3 in 10 000 being excluded. These include the Exome Aggregation Consortium (ExAC) v0.3 (exac.broadinstitute.org/downloads), Phase 1 of the UK10K, and the UK10K TWINS release databases. Variants with a MAF greater than 3 in 10 000 in unrelated SVD patients in BRIDGE were also excluded. Allele frequencies in the genome Aggregation Database (gnomAD) (http://gnomad.broadinstitute.org/) were retrospectively reviewed for all variants passing filters.

Variants that were present in relatives but not in sequenced index cases, as well as variants that did not segregate with the pattern of inheritance of disease in the pedigree were excluded from further analysis. For example, if an affected individual carried a rare variant but this was not found in another affected relative, this variant was not considered as a possible causative variant. Where a known pathogenic mutation is identified in one gene (e.g. known cysteine-altering variant in NOTCH3), rare variants identified in other genes are not considered further. Where relatives' disease statuses are unknown and segregation is not

possible to demonstrate, the variants carried by the affected proband are retained for consideration.

The diagnostic yield was calculated as a percentage of all index SVD cases sequenced, while the allele frequency was calculated as a percentage of all affected individuals, including those from the same pedigree.

5.2.4.2 Copy Number Variants (CNVs)

Copy number variants (CNVs) refer to structural variants, insertions, deletions, duplications, inversions and translocations generally greater than 1kb in size. In this study, CNVs were called using the Manta Structural Variant Caller (http://github.com/Illumina/manta), which calls structural variants and insertions or deletions (indels) from mapped paired-end sequencing reads,[286] and the Canvas Copy Number Variant Caller (http://github.com/Illumina.canvas), which identifies regions of the genome that are not present, present once, or more than twice in a sample assumed to be mostly formed of paired chromosomes.[287] Only CNVs within protein-coding regions with an allele frequency of less than 0.001 were considered.

5.2.5 Modelling of NOTCH3 variants

Epidermal Growth Factor-like repeat (EGFr) sequences of the NOTCH3 protein were obtained from Uniprot (protein ID: Q9UM47), and aligned using the multiple sequence alignment tool of Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/). 3D modelling of the EGFr1 domain was performed using the Swiss-Model tool. EGFr domains were visualised using the Chimera software.[288] Mapping of NOTCH3 variants was performed by Gido Gravesteijn, at the Leiden University Medical Centre, The Netherlands.

5.3 Results and discussion

5.3.1 Baseline characteristics

One hundred and forty-three index cases with suspected familial SVD, and 137 of their relatives were recruited according to the criteria described in Chapter 2. Whole genome sequencing data was available from 246 participants (121 index cases, 125 relatives). Of these, three index cases were no longer eligible after recruitment, having received diagnoses of CADASIL confirmed on clinical genetic testing (n=2) or neurosarcoidosis (n=1). Four other participants were also excluded from further analyses as the probands from these pedigrees had not successfully been sequenced at the time of writing. Of the remaining 239 individuals included in this analysis, 118 were index cases and 121 were relatives. There was a total of 143 individuals with MRI-confirmed diagnoses of SVD.

The clinical features of all 118 index cases at baseline are described in Table 5-2. Samples from these participants were sequenced alongside 9110 controls (patients and relatives from other rare disease studies in BRIDGE), and 7348 of these were unrelated individuals. These were included in the analysis as a control population. (Figure 5-1).

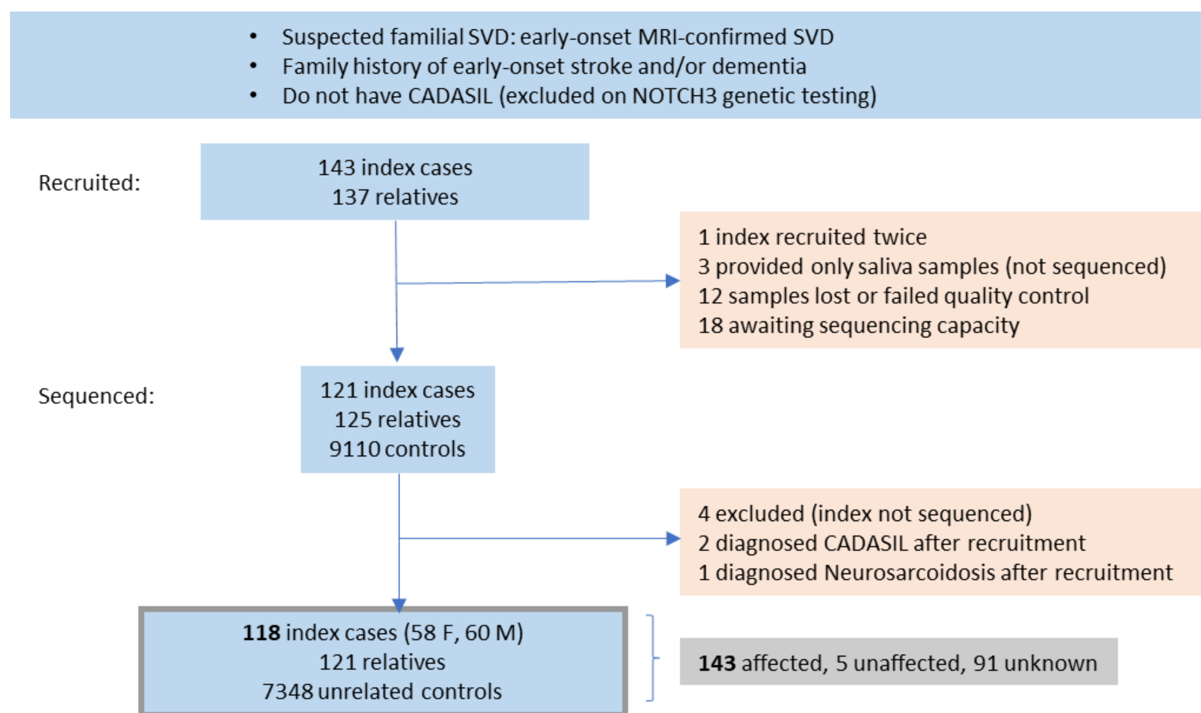


***Figure 5-1**: Summary of selection criteria and recruitment of participants to the BRIDGE-SVD study*

Table 5-2: Clinical features of 118 index cases with suspected familial cerebral SVD

	Male	Female	Total
All patients	60	58	118
Age at presentation (mean ± SD)	52.5 ± 8.7	52.5 ± 11.0	52.5 ± 9.9
Presenting feature (major neurological feature)			
• **Stroke**	33	7	40
• **TIA**	0	3	3
• **Cognitive impairment**	4	5	9
• **Dementia**	0	0	0
• **Encephalopathy**	1	5	6
• **Seizure**	0	2	2
Migraine (n, %)	21 (35.0)	40 (69.0)	61 (51.7)
Age at onset (mean ± SD)	30.9 ± 14.4	30.9 ± 13.5	30.9 ± 13.7
Migraine with aura	12 (20.0)	33 (56.9)	45 (38.1)
Stroke (n, %)	49 (81.7)	18 (31.0)	67 (56.8)
Ischaemic stroke	44 (73.3)	17 (29.3)	61 (51.7)
Haemorrhagic stroke	8 (13.3)	2 (3.4)	10 (8.5)
Age at onset of stroke (mean ± SD)	53.1 ± 6.2	53.0 ± 10.7	53.0 ± 7.7
Encephalopathy (n, %)	3 (5.0)	8 (13.8)	11 (9.3)
Age at onset (mean ± SD)	59.0 ± 2.9	50.3 ± 17.5	52.9 ± 15.3
Cognitive impairment	16 (26.7)	24 (41.4)	40 (33.9)
Age at onset	51.4 ± 8.6	52.5 ± 10.5	52.1 ± 9.8
Dementia	0	5 (8.6)	5 (4.2)
Age at onset	-	52.3 ± 7.6	52.3 ± 7.6
Seizures	4 (6.7)	6 (10.3)	10 (8.5)
Age at onset	48.0 ± 18.7	33.0 ± 12.3	39.0 ± 16.9

5.3.2 Copy Number Variants (CNVs)

Across the genome, 354 CNVs were called in 184 individuals from 96 different pedigrees. All CNVs that were not present in the index case of a pedigree, or did not segregate with the pattern of inheritance of disease within pedigrees were excluded, leaving 155 remaining CNVs in 98 individuals from 79 pedigrees.

5.3.2.1 CNVs in candidate genes

Two SVD singletons were found to have large, non-overlapping deletions spanning 12 to 13 genes on chromosome 13q13 and 13q31. A large chromosome 13q33-q34 duplication spanning 45 genes including COL4A1 and COL4A2 was previously described in a singleton with early-onset SVD. This individual did not have a clear family history of SVD,[87] and no other CNVs on chromosome 13 have previously been associated with SVD. In this study, the CNVs did not overlap with the COL4A1 or COL4A2 gene region.

Chromosome 6p25 duplications affecting the FOXC1 gene have also been identified in previous reports of FOXC1-associated SVD. In this study, one singleton was found to have a deletion in 6p24.3 affecting the GCNT2 gene, and two affected individuals in one family were found to have a deletion in the same chromosome (6p24.1), affecting the TBC1D7 gene, as described in **Pedigree F102**. No CNVs were found to affect the FOXC1 gene.

The PITX2 gene, which acts in the FOXC1 pathway leading to both familial SVD and the Axenfeld-Rieger Syndrome, is found on chromosome 4. No CNVs were identified in the PITX2 gene region in our study, with the nearest deletion found in a singleton residing more than 5.8 million bases downstream in the RP11-55L3.1 gene.

No other CNVs were identified in the 73 candidate genes.

Pedigree F102

A white male of Italian ancestry (Figure 5-2**, III-1**) presented with a haemorrhagic stroke in the pons at the age of 52, and was found on MR imaging to have white matter hyperintensities, lacunar infarcts and microbleeds in both the cortical and subcortical regions. (Figure 5-3**A-C**) He had no previous history of migraine or psychiatric problems.

His mother (**II-4**) had a history of multiple subcortical ischaemic and haemorrhagic strokes from the age of 74. Her past medical history was only significant for treated anxiety, and her MRI showed multiple lacunes, white matter hyperintensities and multiple microbleeds in the cortex. (Figure 5-3**D-F**) Her father had died of a cerebral haemorrhage at the age of 48, and her sister had a diagnosis of dementia, although the type of dementia had not been specified.

Both the proband and his mother were found to have a heterozygous deletion of chromosome 6: 13310619-13337175, affecting the TBC1D7 gene. TBC1D7 (TBC1 Domain Family Member 7) is a component of the tuberous sclerosis complex (TSC), which has roles in limiting cell growth. Deletions resulting in frameshift and premature termination mutations have been associated with autosomal recessive macrocephaly/ megalencephaly syndrome. This disease is characterised by the abnormal enlargement of the cranium and brain at birth or in childhood, and intellectual disability. Brain MRI in these patients have shown enlarged corpus callosum and cerebral calcifications, or were normal. [289,290] Autosomal dominant mutations in TBC1D7 have not been described in disease, and it is not known histopathologically if this disease is on the SVD spectrum.

Sequencing and MR imaging to confirm the presence of the maternal aunt (**II-5**) diagnosed with dementia, as well as other unaffected relatives may serve to provide further evidence for segregation of this variant with disease in the pedigree.

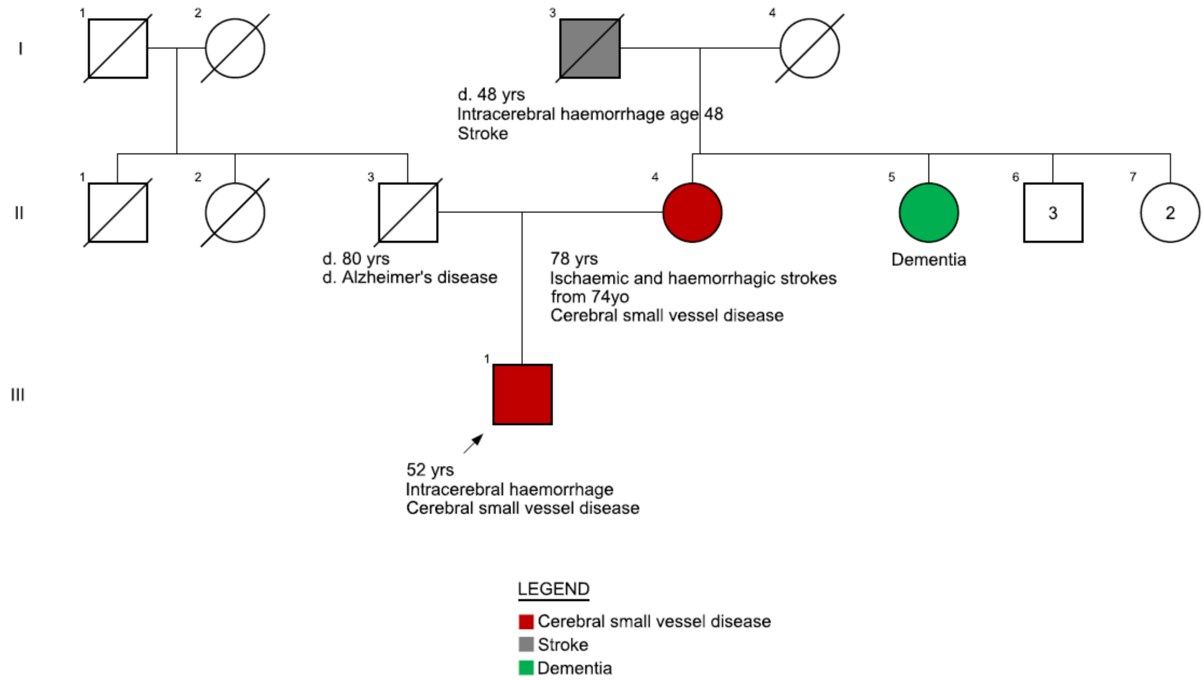


Figure 5-2*: Pedigree F102. Both affected individuals in pedigree F102 (III-1 and II-4) have a 6p24.1 deletion affecting the TBC1D7 gene.*

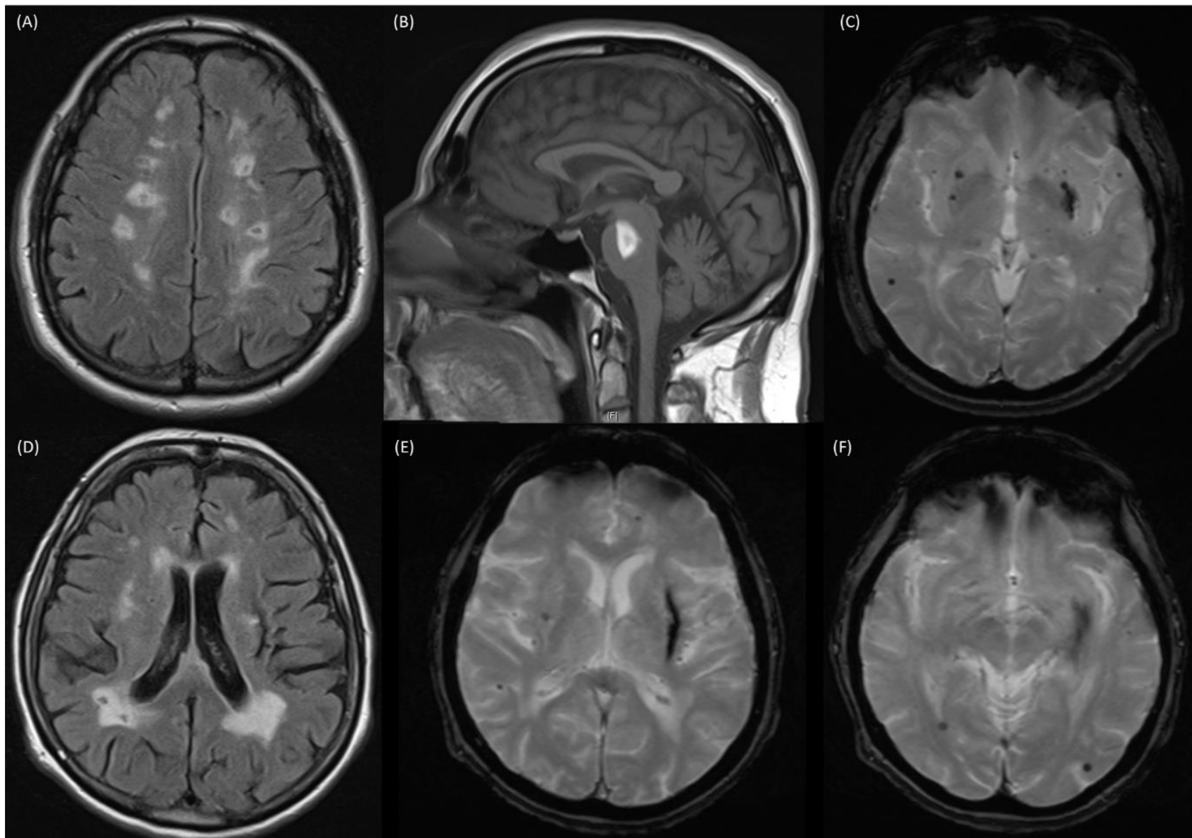


Figure 5-3: *MR images of III-1 of Pedigree F102 (A) T2-weighted FLAIR image showing white matter hyperintensities and lacunes, (B) T1 image showing acute pontine haemorrhage, and (C) Gradient echo imaging showing cortical and subcortical microbleeds and an old left-sided haemorrhage. His mother, II-4 had similar features on (D) T2-weighted FLAIR imaging, showing white matter hyperintensities and lacunes, and (E and F) cortical and subcortical microbleeds and an old left sided subcortical haemorrhage.*

5.3.2.2 Other CNVs

There were no other identical CNVs that were present in more than one singleton or pedigree. Two singletons were found to have overlapping CNVs in chromosome 16: one with a 300 000bp deletion, and the other a 6000bp deletion within this region. The overlapping gene, C16orf45, has previously been associated with Lisch's endothelial corneal dystrophy. While corneal opacities and other ocular abnormalities have been described in familial SVD, these two index cases were not known to have any ocular abnormalities.

A summary of all CNVs identified is provided in the appendix.

5.3.3 Single Nucleotide Variants (SNVs)

In 118 SVD index cases and 121 affected or unaffected relatives, a total of 337 rare variants in the 73 candidate genes passed filters. An average of 4.6 variants per gene passed filters, with majority of these being non-coding variants. The proportions of each type of variant passing filters from each candidate gene is summarised in Figure 5-4.

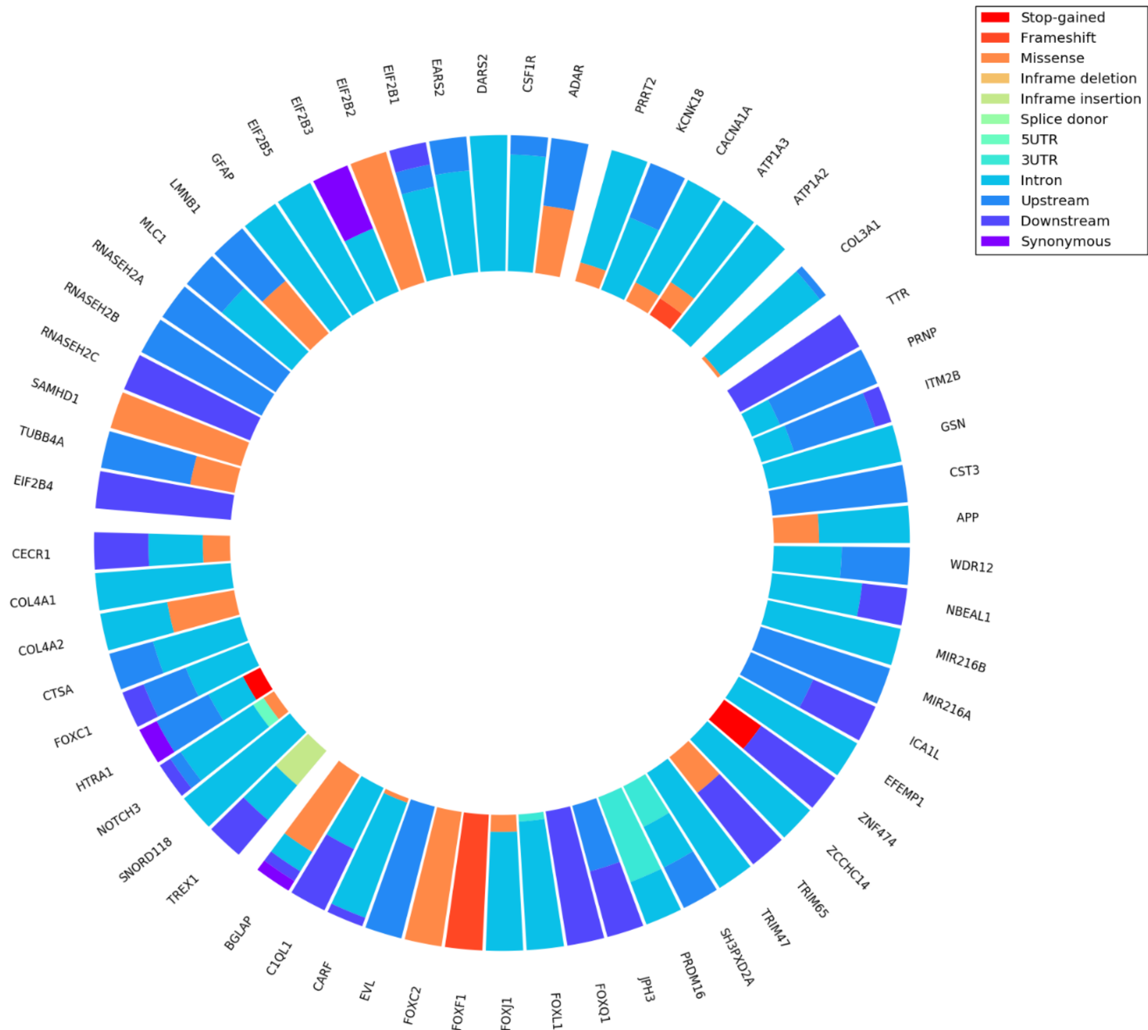


Figure 5-4: *Proportion of each type of variants called in candidate genes across all SVD participants which passed filters. This includes variants excluded from analysis e.g. where the variant was not present in the index case, or where variants did not segregate with the pattern of inheritance of disease. Genes have been organised according to the category from which they were selected. Those with no variants passing filters have not been included in this figure. Legend notes: 3'UTR: 3-prime untranslated region, 5'UTR: 5-prime untranslated region.*

CADASIL-causing and other rare NOTCH3 variants are identified in a presumed CADASIL-negative cohort of familial SVD patients

CADASIL is caused by cysteine-altering variants in the epidermal growth factor repeat-like (EGFr) region of the NOTCH3 gene. These mutations do not lead to a loss- or gain-of-function, but result in the abnormal deposition of the NOTCH3 ectodomain in the walls of the small blood vessels.[119]

In this study, 3 cysteine-altering variants were identified in 3 index cases: one previously reported variant p.Arg169Cys (exon 4),[124,259] a novel variant p.Cys76Tyr (exon 3), and a variant p.Cys212Arg (exon 4) which has not been reported in literature but has been identified in another patient with a clinical diagnosis of CADASIL, seen at the UK CADASIL National Referral Clinic.

As with all probands included in this analysis, all three patients previously had a genetic diagnosis of CADASIL excluded prior to WGS. Two were screened on Sanger sequencing (exons 3 and 4) and denaturing high performance liquid chromatography (DHPLC) (exons 3, 4, 5, 6, 11, 18, 19 and 22), while the remaining patient was screened for mutations in exons 2 to 24 on Sanger sequencing alone. None of the patients were found to have any mutations. In this cohort of 118 index cases, this gave a diagnostic yield of 2.5%, and a combined allele frequency of 2.1% (of all affected cases). The patient with a p.Cys76Tyr mutation was subsequently re-tested in a clinically-accredited laboratory using Sanger sequencing, which validated the presence of this mutation.

Two pedigrees were found to carry non-Cysteine changing missense variants. Two of these variants were identified in a singleton, as discussed below (**Pedigree F37**). The remaining variant, p.Thr328Ile, was found in all three members of a pedigree (one affected, two of unknown disease status). All variants identified were in a heterozygous state. A summary of coding NOTCH3 variants identified in index cases is provided in Table 5-3. Non-coding variants identified in NOTCH3 are summarised in the appendix.

__Table 5-3:__ Coding NOTCH3 variants passing filters, identified in SVD index cases. Variants that were not found in the index case of the pedigree, or did not segregate with the disease in the pedigree, are excluded from this list.

Variant	Consequence	CADD	Diagnostic yield	AF	Unrelated controls	ExAC	gnomAD	UK10K	Remarks
c.634T>C p.Cys212Arg	Missense	25.1	0.0042	0.0035	-	-	-	-	• Not reported in literature • Present in 1 patient seen at UK National CADASIL Referral Clinic. • p.Cys212Cys present at an AF of 8.1×10^{-6} in gnomAD
c.227G>A p.Cys76Tyr	Missense	27.5	0.0042	0.0035	-	-	-	-	• Novel • p.Cys76Arg and p.Cys76Trp previously reported[245,257,258]
c.505C>T p.Arg169Cys	Missense	27.5	0.0042	0.0035	-	-	-	-	Previously reported[57,259]
c.133G>C p.Asp45His	Missense	27.6	0.0042	0.0035	-	1.7×10^{-5}	8.3×10^{-6}	-	Identified in cis in one index case
c.154G>A p.Gly52Arg	Missense	29.5	0.0042	0.0035	-	1.7×10^{-5}	1.2×10^{-5}	-	
c.983C>T p.Thr328Ile	Missense	27.4	0.0042	0.0035	-	-	1.1×10^{-5}	0.00028	Identified in 1 pedigree – 1 affected, 2 unknown disease status

AF: allele frequency in all affected SVD cases.

Diagnostic yield: frequency in index cases only

Pedigree F37: Two non-cysteine altering missense mutations in cis on NOTCH3 gene

A 57-year-old white British female (Figure 5-5**, III-1**) presented with cognitive impairment and recurrent subcortical SVD-related lacunar infarcts at the age of 57. She had a history of migraine with aura. On MR imaging of the brain at age 57, she was found to have bilateral white matter hyperintensities affecting the external capsules, but not the anterior temporal poles. (Figure 5-6). She did not have any granular osmiophilic material (GOM) on electron microscopy of her skin biopsy, although her small blood vessels had an abnormally thick basement membrane.

Her father (**II-3**) had had 6 strokes from the age of 62. Her brother (**III-2**) had a similar history of migraines, having his first stroke at age 53. DNA collected from her identical twin sisters (**III-3 and III-4**) who were asymptomatic and had normal MR imaging of the brain, and daughter (**IV-1**), who had a history of migraine, depression and anxiety, and a normal MRI at the age of 40, was sequenced.

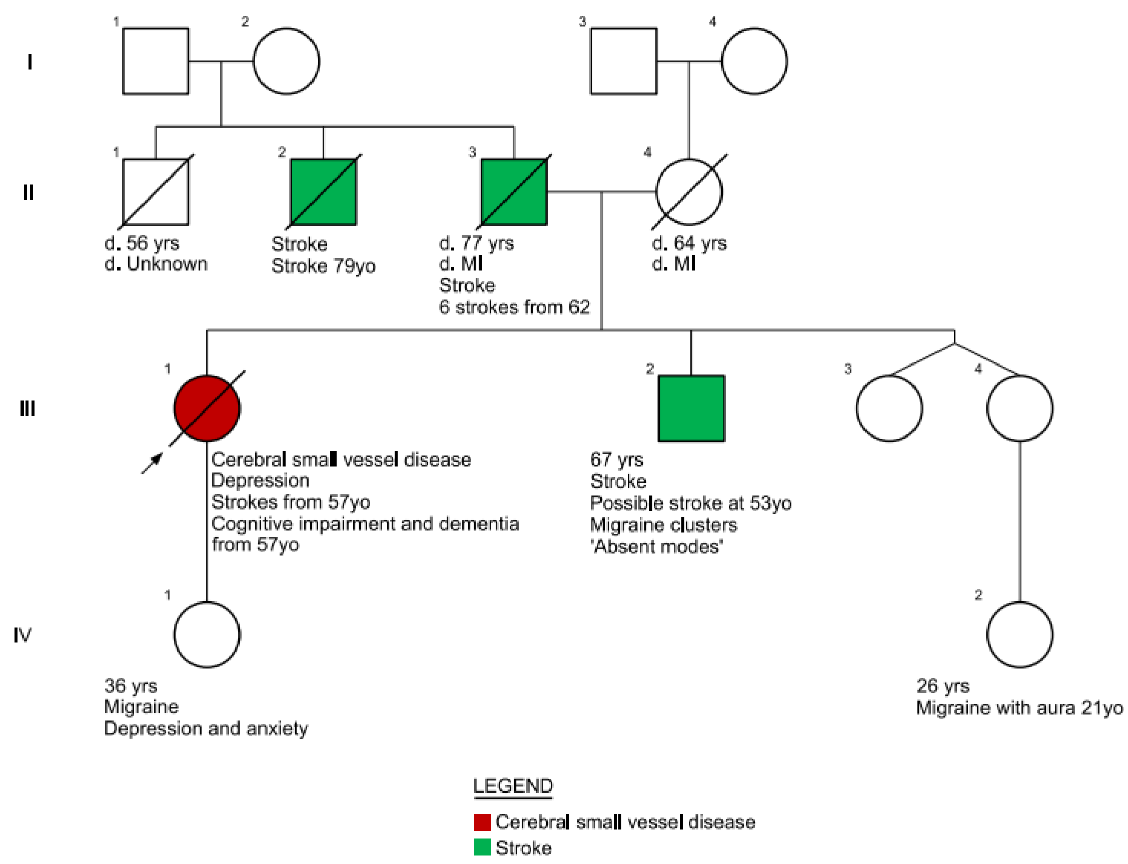


Figure 5-5: *Pedigree F37*

The proband was found to carry two rare non-cysteine-altering NOTCH3 variants in cis (on the same haplotype), p.Asp45His and p.Gly52Arg. None of her relatives who were sequenced (III-3, IV-1) were found to carry either of these variants, thus these variants segregate with the presumed pattern of inheritance of the disease based on MRI phenotyping.

The two variants lie six residues apart, and both lie within the first EGFr domain of the NOTCH3 gene. These residues were both close to the two cysteine residues which are predicted to form one disulphide bond.

On three-dimensional mapping of the variants on the NOTCH3 wild-type model, the backbone was predicted to deviate slightly from the wild-type model, resulting in a change in the orientation of the second cysteine residue of the EGFr domain. (Figure 5-7**A** and Figure 5-7**B**) However, this model does not demonstrate consequences for disulphide bridge formation, and it is thus not possible to conclude that this patient has CADASIL resulting from altered cysteine pairing.

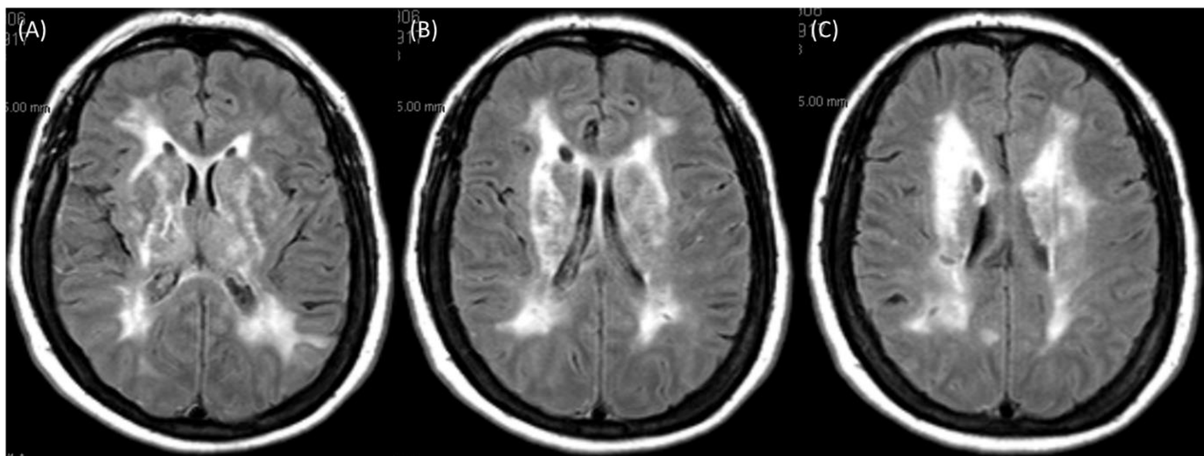


Figure 5-6: *T2-weighted MR imaging of the proband of pedigree F37, showing confluent white matter hyperintensities (A-C) affecting the external capsules (B).*

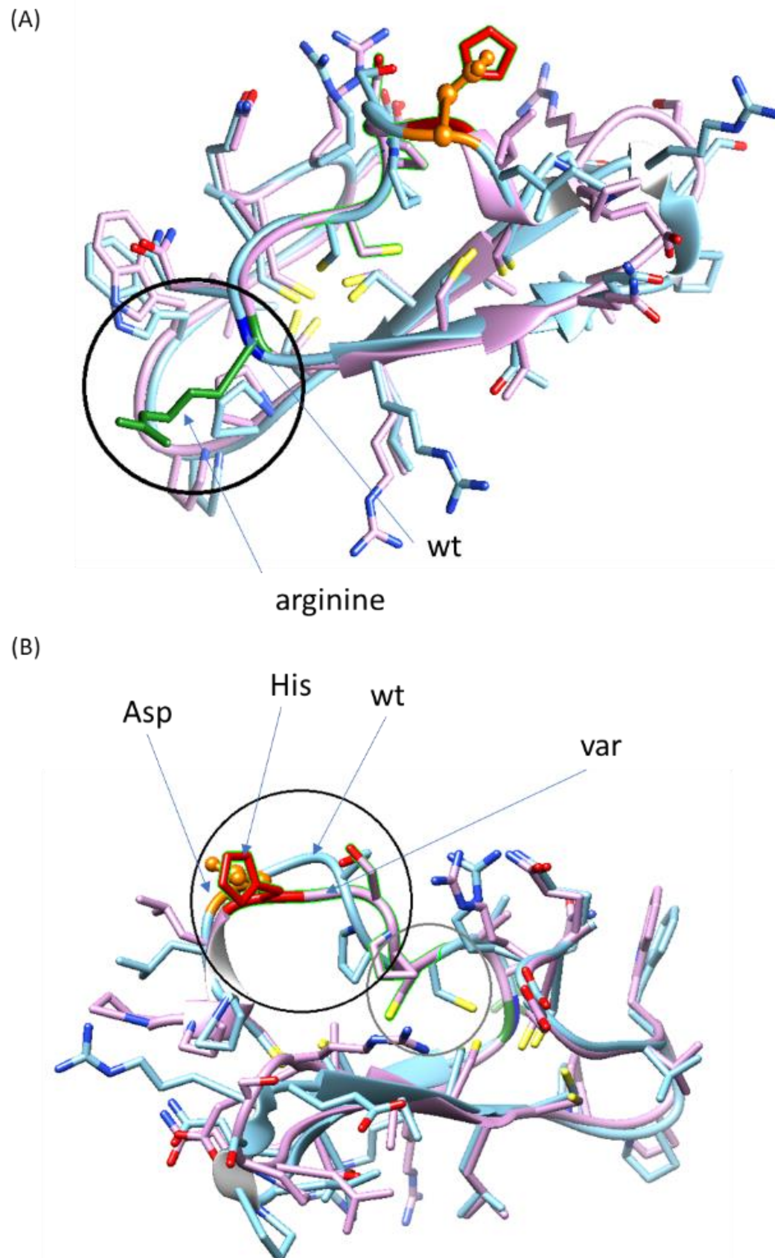


***Figure 5-7**: 3D modelled EGFr domain 1 of the NOTCH3 protein with an overlay of the variants identified in the proband of Pedigree F37. (A) The p.Gly52Arg variant is circled, with the wildtype glycine residue marked in dark blue, and the variant arginine is marked in green. This variant is not predicted to cause a major change in the EGFr domain backbone. (B) The Asp45His variant is circled, with the wildtype domain in light blue and variant domain marked in purple. The second cysteine residue of this domain is predicted to have a slightly altered position compared to the wildtype, although it is not known whether this will affect disulphide bond formation. Modelling was performed by Gido Gravesteijn.*

5.3.3.1 *HTRA1 mutations are the next most common cause of undiagnosed familial SVD after CADASIL*

Mutations in the HTRA1 gene are associated with an autosomal recessive form of familial SVD known as CARASIL,[61] and in recent years heterozygous carriers of HTRA1 variants have also been found to have a milder form of symptomatic familial SVD without the extra-neurological features of CARASIL.[67,264]

Nine HTRA1 variants were identified in 10 index cases, as summarised in Table 5-4. Of these, five variants in six index cases were coding (missense or nonsense variants), and three of these (p.Pro285Leu, p.Val175Met, p.Arg370Ter) have previously been reported in either or both CARASIL and autosomal dominant HTRA1-associated familial SVD. (Figure 5-8) Two novel missense variants were identified, and one of these cases is discussed below (**Pedigree F05**). The overall diagnostic yield was 5.1% of index cases, with a combined allele frequency of 6.3% in all affected cases. The clinical features of all five index cases with missense and nonsense variants are summarised in **Table 5-5** .

Three intronic variants were identified in three probands (two singletons), and one relative whose disease status is unknown. These variants are summarised in the appendix at the end of the chapter. Intronic mutations have not previously been reported in CARASIL or autosomal dominant familial SVD.

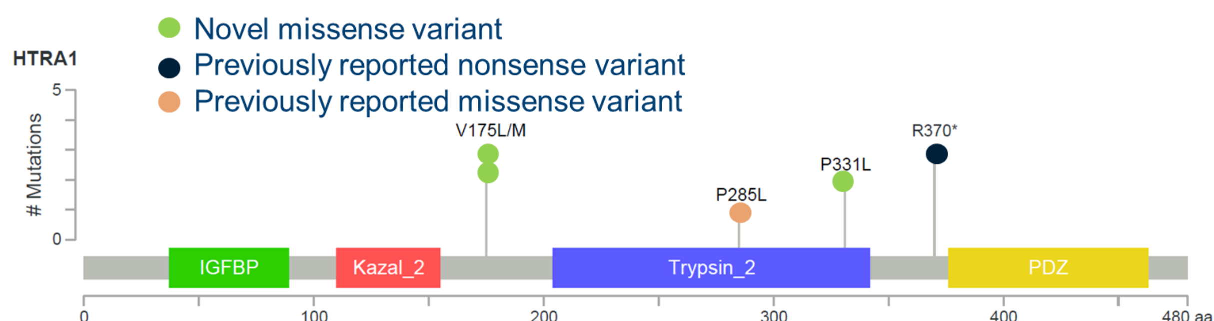


Figure 5-8: *Heterozygous missense and nonsense variants identified in HTRA1. p.Pro285Leu and p.Pro331Leu are found within the trypsin domain of HTRA1. Variants at or near these positions have been found to result in inactive protein on in-vitro functional analyses, while p.Arg370Ter has been demonstrated to result in nonsense-mediated mRNA decay in patient fibroblast cells.*

Pedigree F05: a novel HTRA1 variant found in three affected individuals in a pedigree

A white British female (Figure 5-9**, III-1**) presented with an encephalopathic episode at the age of 46, and was diagnosed with cognitive impairment within the same year. She had a history of migraine with aura during her pregnancy at 38 years old.

A year later, her 51-year-old sister (**III-2**) was also admitted during an acute confusional episode subsequently diagnosed as encephalopathy. This sister had a history of recurrent TIAs, migraine with and without aura and anorexia nervosa. She subsequently suffered a fatal stroke at age 52, although it is unclear if this was a haemorrhagic stroke or an ischaemic stroke with haemorrhagic transformation.

Their mother (**II-2**) was relatively fit and well until the age of 70, when she developed recurrent TIAs manifesting as transient dysphasic episodes. These episodes did not result in any positive findings on diffusion-weighted MRI, and she had no positive changes on electroencephalogram (EEG) testing. Their paternal grandfather had also suffered a stroke at the age of 65.

On MR imaging, all three of the pedigree had confluent white matter changes with subcortical lacunar infarcts. (Figure 5-10) The proband also had white matter abnormalities involving the anterior temporal pole. Screening of the NOTCH3 gene in both the proband and her sister, and the GLA gene in the sister, revealed no mutations. On examination of their skin biopsies, no granular osmiophilic material (GOM) were detected, although the sibling was noted to have re-duplication of the basement membrane.

DNA from whole blood was collected from the proband and her sister, Whole genome sequencing revealed a heterozygous p.Pro331Leu variant in both sisters. Sanger sequencing performed on DNA from saliva from their mother confirmed that she also carried this novel variant.

p.Pro331Leu is a variant in the trypsin-like serine protease inhibitor domain of the HtrA1 enzyme. This domain performs the key enzymatic function of the protein (Figure 5-8). Previous functional analyses have demonstrated that a mutation near this position (p.Ser328Ala) results in a protein that is able to form the HtrA1 trimer, but is catalytically inactive.[174]

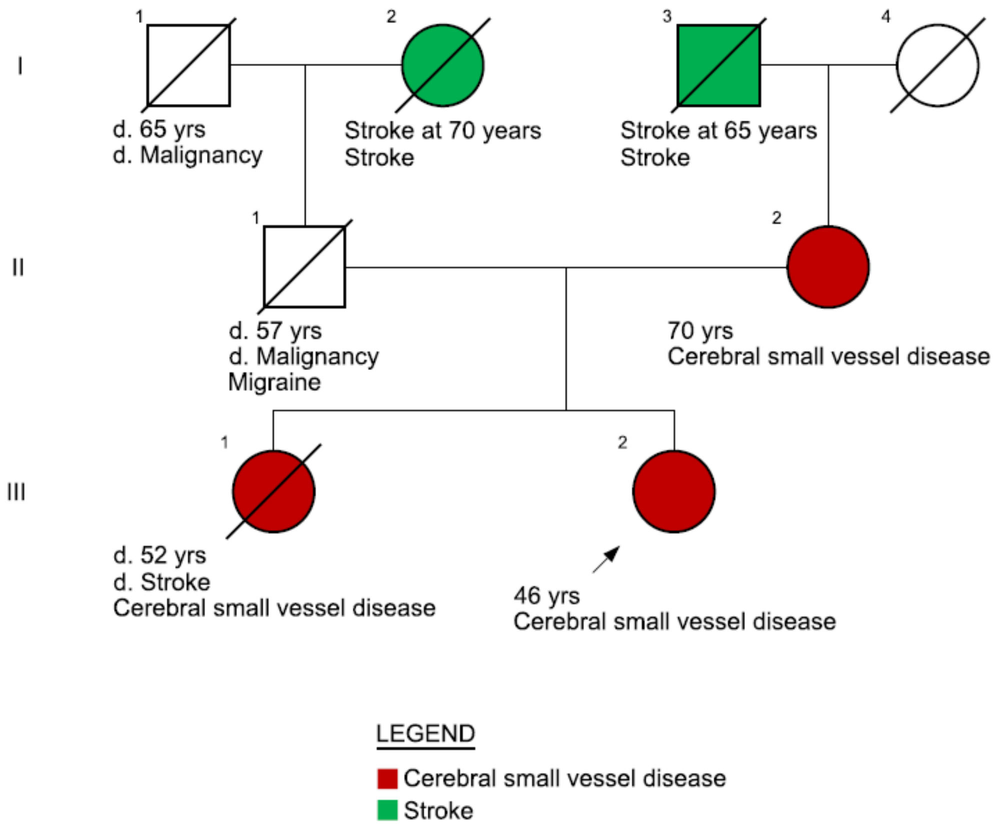


Figure 5-9: *Pedigree F05*

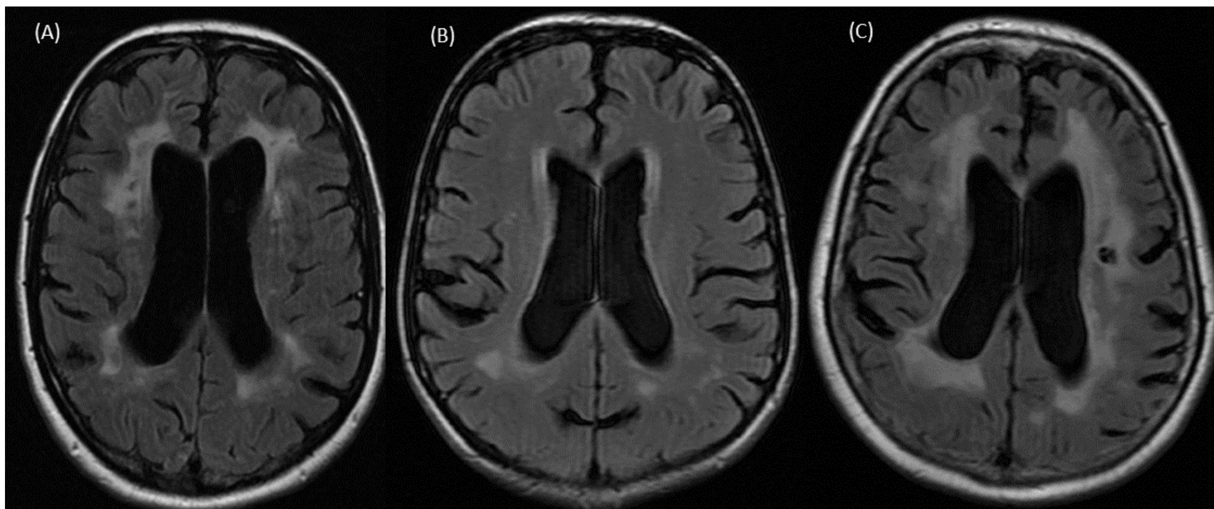


Figure 5-10: *MR images of pedigree F05. MRI T2-FLAIR images of (A) The proband (III-1), (B) her sister (III-2) and (C) their mother (II-2)*

Pedigrees F122 and F34: a CARASIL -causing HTRA1 variant found in two pedigrees

In this cohort we identified three affected individuals from two pedigrees each carrying a single copy of the p.Arg370Ter variant in HTRA1. This variant is absent from 7348 unrelated controls, and is present in less than 3 in 10 000 individuals in control data sets (1.6×10^{-5} in gnomAD) (Table 5-4).

The first variant carrier is the proband of Pedigree F122 (Figure 5-11**A**), a 68-year-old white British male (**II-3**) who presented complaining of memory problems. He struggled with DIY tasks that he previously carried out with ease, and had lost confidence in driving. On cognitive testing he showed deficits in attention, verbal fluency and recall, in keeping with vascular cognitive impairment.

In the preceding year he had multiple seizure episodes, during which he would become unrousable for about half an hour, and would then be confused. He would often be incontinent of urine during these episodes, but would not be observed to have any stiffening or twitching.

On MR imaging of the brain, he was found to have extensive white matter changes and an old thalamic infarct (Figure 5-12). His sister (**II-2**), who was also found to carry the variant, was diagnosed with vascular dementia at the age of 71. She was also found to have white matter changes on her MRI suggestive of vascular disease. Their mother (**I-2**) had had a history of migraine with aura from the age of 25, and had a stroke at 70 years. Their brother (**II-1**) had an episode of collapse in his 20s, although little else was known about this episode.

The third carrier, from Pedigree F34 (Figure 5-11**B**), was a 48-year-old white British male (**III-1**) who presented with a subcortical ischaemic lacunar stroke. He had a previous history of complicated migraines with aura, and aura without headache from the age of 25, during which he would develop unilateral facial numbness. His MRI showed patchy white matter hyperintensities, bilateral lacunar infarcts and numerous microbleeds in both the cortical and subcortical regions (Figure 5-12). His mother (**II-2**) suffered her first stroke at age 32, dying of a subsequent stroke at the age of 57.

The p.Arg370Ter variant in HTRA1 was previously described in two consanguineous families with CARASIL.[61,291] Heterozygous variant carriers in both the Japanese and Turkish families

were asymptomatic at the time of sequencing, although no brain imaging had been performed. In vivo analyses of the Japanese CARASIL patient's fibroblasts showed that this mutation results in a loss of function of the resulting HTRA1 protein, by causing nonsense-mediated mRNA decay.[61] Leukocytes from a heterozygous carrier of p.Arg370Ter in a consanguineous Japanese family showed only the presence of wild-type HTRA1 mRNA,[61] thus providing further evidence of nonsense-mediated decay of the variant mRNA strand resulting from this variant.

The HtrA1 enzyme is a homotrimeric protein. Disease-causing mutations are thought to result in either the loss of high temperature requirement serine protease A1 (HtrA1) enzyme function, or the failure to form a functional HtrA1 trimer, impairing the function of HTRA1 on the TGFβ pathway.[292] It was previously suggested that heterozygous missense mutations could exert a dominant negative effect by affecting the function of the wild-type protein, and that nonsense mutations resulting in haploinsufficiency would not cause SVD.[67] However, a heterozygous loss-of-function mutation p.Arg302Ter which was previously described in CARASIL was subsequently identified in a Japanese patient with SVD and no extra-neurological features of CARASIL.[264] This variant also results in nonsense-mediated mRNA decay.

The presence of p.Arg370Ter in three affected individuals in two unrelated pedigrees, and its low frequency in control data sets thus also lend support for the pathogenicity of this variant in the heterozygous state. Despite the absence of a mutant HTRA1 protein in another heterozygous carrier of this variant, the variant may still lead to haploinsufficiency and thus a milder form of the disease.

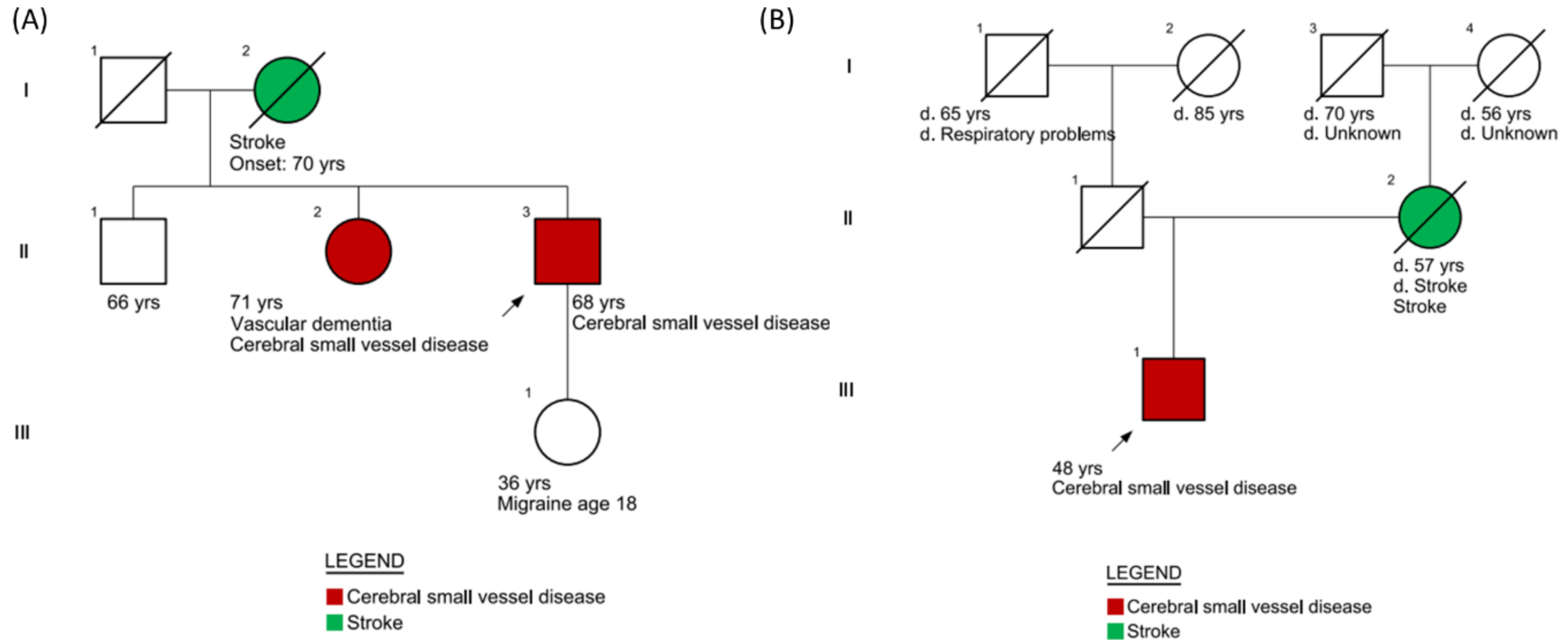


Figure 5-11*: Pedigrees F122 (A) and F34 (B). Both probands, and II-2 of pedigree F122 carry a single copy of the p.Arg370Ter variant in the HTRA1 gene.*

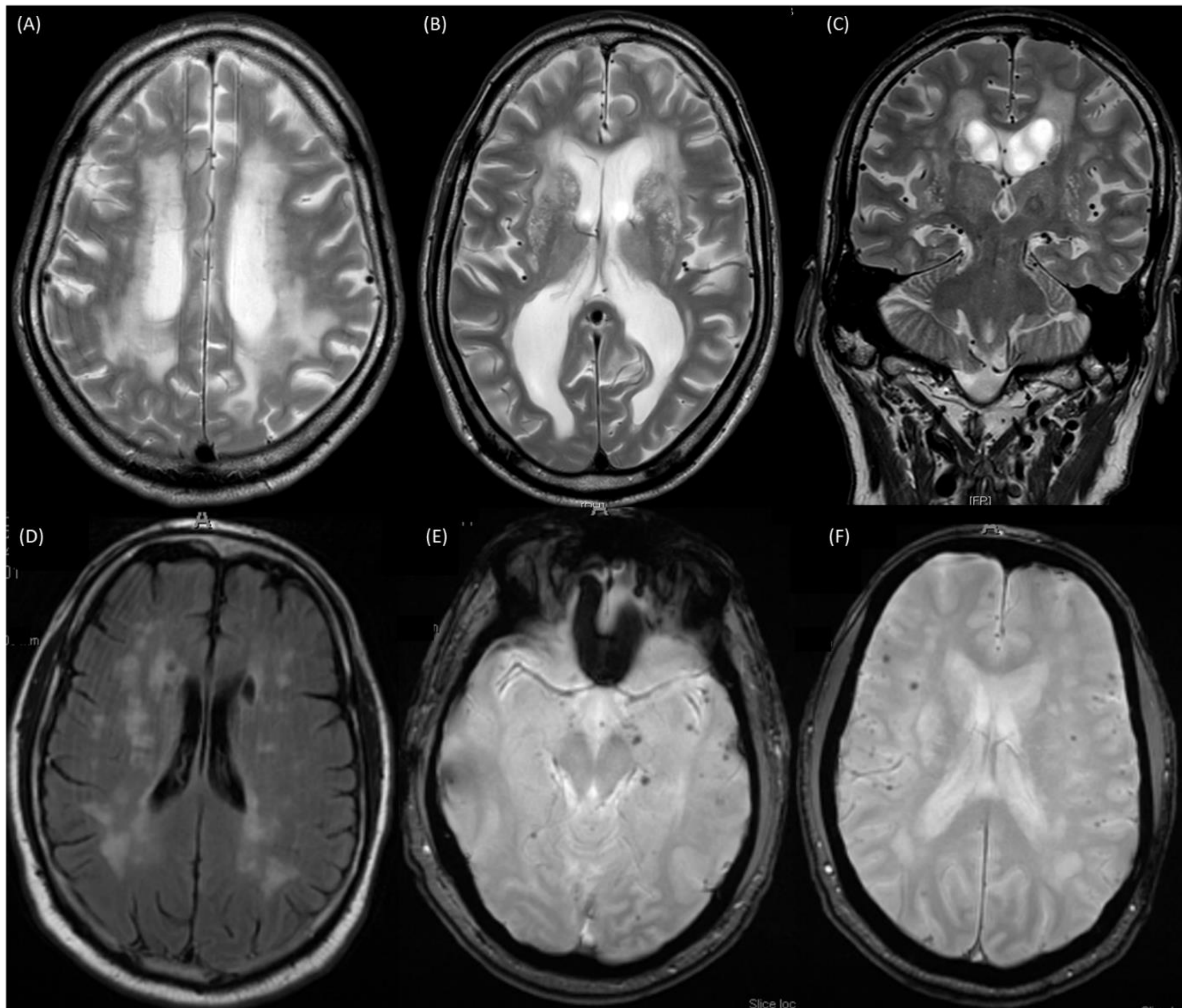


Figure 5-12: *MR imaging of the proband in pedigree F122 showing (A) and (B) confluent white matter hyperintensities, and (C) a left thalamic infarct on T2 imaging. This patient did not have any clinical strokes. The proband in pedigree F34 had patchy white matter hyperintensities on T2-FLAIR imaging (D), and multiple microbleeds in subcortical and cortical regions on susceptibility weighted imaging (E and F).*

Table 5-4: *Missense and nonsense HTRA1 variants passing filters, identified in SVD index cases. Variants that were not found in the index case of the pedigree, or did not segregate with the disease in the pedigree, are excluded from this list.*

Variant	Consequence	CADD	Diagnostic yield	AF	Unrelated controls	ExAC	GnomAD	UK10K	Previously reported?	Functional impact?
c.992C>T p.Pro331Leu	Missense.	35	0.0042	0.0035	-	-	4.1×10^{-6}	-	Novel	Mutation near this position (p.Ser328Ala) results in inactive protein[174]
c.1108C>T p.Arg370Ter	Nonsense.	44	0.0085	0.0070	-	8.2×10^{-6}	1.6×10^{-5}	-	Reported in CARASIL[61,291]	Loss of function involving nonsense-mediated mRNA decay demonstrated in patient fibroblasts[61]
c.523G>C p.Val175Leu	Missense.	32	0.0042	0.0035	-	-	-	-	Novel. p.Val175Met previously reported (see below)	Mutation near this position (p.Ala173Pro) reported in both CARASIL and autosomal dominant disease, results in catalytically inactive enzyme secreted at lower levels (autodegradation)[67]
c.523G>A p.Val175Met	Missense.	33	0.0042	0.0070	-	8.2×10^{-6}	4.1×10^{-6}	-	Previously reported in autosomal dominant SVD[265]	

Variant	Consequence	CADD	Diagnostic yield	AF	Unrelated controls	ExAC	GnomAD	UK10K	Previously reported?	Functional impact?
c.854C>T p.Pro285Leu	Missense.	35	0.0042	0.0035	-	-	4.1×10^{-6}	-	Previously reported in CARASIL, mutation affecting same residue reported in autosomal dominant SVD.[174,293]	Protein is proteolytically inactive. Variant is within activation loop domain needed to induce its active conformation[68]

AF: allele frequency in all affected patients. Unrelated controls: unrelated, non-SVD patients in BRIDGE.

Diagnostic yield: frequency in index cases only

Table 5-5: Clinical features of probands with heterozygous missense or nonsense HTRA1 variants

Pedigree	Variant	Clinical features	Sex	Age at onset*	MRI brain findings	Significant family history
F39	p.Pro285Leu	• Migraine aura without headache • Depression • Executive dysfunction	Female	45	Confluent WMH	• Father – recurrent strokes, cognitive impairment from age 58. • Paternal aunt, uncle died of strokes in early 60s.
F05	p.Pro331Leu	• Migraine with aura • Encephalopathy • Cognitive impairment	Female	46	Confluent WMH involving anterior temporal poles	MRI confirmed SVD in sister (migraine, TIAs, encephalopathy, psychiatry, stroke) and mother (TIAs). All three carry this variant.
F95	p.Val175Met	• Complicated migraines • Encephalopathy • Muscle aches and stiffness (diagnosed 'polymyalgia rheumatica')	Female	59	Confluent WMH involving anterior temporal poles, microhaemorrhages	Sister – one episode of dysphasia, recurrent vertigo from age 59, diagnosed with benign paroxysmal positional vertigo, incidental finding of WMH on MRI. Also carries variant.
F28	p.Val175Leu	• TIA • Small vessel ischaemic stroke • Depression, emotional lability • Parkinsonism: limb rigidity, bradykinesia, mask-like facies	Male	58	Confluent WMH involving anterior temporal poles, microhaemorrhages.	Mother – dementia from 80s.

Pedigree	Variant	Clinical features	Sex	Age at onset*	MRI brain findings	Significant family history
F34	p.Arg370Ter	• Small vessel ischemic stroke • Migraine with and without aura	Male	48	Confluent WMH, multiple lacunar infarcts, microbleeds predominantly in cortical but also in subcortical regions.	Mother – stroke at 32 years, died of stroke at 57.
F122		• Cognitive impairment • Recurrent encephalopathy • Mild hearing impairment • Migraine	Male	68	Confluent WMH and brain atrophy	Sister – MRI confirmed SVD, diagnosed vascular dementia at age 61. Both proband and sister carry variant

*Age at onset of first major clinical feature, such as stroke, encephalopathy, seizures, cognitive impairment or dementia.

5.3.3.2 Novel coding and non-coding mutations in COL4A1 and COL4A2 may be associated with haemorrhagic SVD

Mutations in, and duplications of the COL4A1 and COL4A2 variants are associated with a form of familial SVD characterised by white matter abnormalities and both ischaemic and haemorrhagic subcortical strokes.[87] Mutations were first identified in infants with hemiparesis and cognitive delay, who were found to have porencephaly.[86] Other rare clinical features include schizencephaly and retinal vascular tortuosity.[269]

Type IV collagen α-chains each have a similar structure: they have a non-collagenous domain (NC1) at their C-terminus, multiple tripeptide Gly-X-Y repeats forming a long central triple-helical domain, and a short N-terminal triple-helical 7S domain.[115] Typical SVD-causing mutations affect the glycine residue of the Gly-X-Y repeats, although variants have also been described in other domains.[165] Disease is thought to result from either haploinsufficiency, or a dominant-negative effect of heterozygous variants.[294]

Six and 10 index cases were found to carry novel heterozygous variants passing filters in the COL4A1 and COL4A2 genes respectively. Although no known disease-causing mutations were identified in this cohort, one patient with a history of subcortical haemorrhagic strokes was found to have a glycine-altering variant within the Gly-X-Y tripeptide repeat region in COL4A1 which also results in a frameshift and premature termination codon (p.Gly1369ArgfsTer33). This variant not only affects the triple-helical domain but also alters the sequence of the remaining peptide, and is thus likely to be deleterious. (Figure 5-13)

One proband carried variants in both COL4A1 and COL4A2. This patient had both a missense variant in COL4A1, and an intronic variant in COL4A2, although the COL4A2 variant was also shared with an asymptomatic sibling. These variants are summarised in Table 5-6, and the clinical features summarised in Table 5-7.

One pedigree, comprising two affected and three unaffected individuals, was studied in a separate study which identified a 3' untranslated region variant (c.*32G>A). This study further demonstrated the segregation of this variant in the extended pedigree of 10 affected and 10 unaffected individuals.[295] Although this variant was also identified in the two affected individuals recruited to this study, the variant did not pass filters as it had a CADD score of 8.3. The variant was not present in any unrelated controls. Only one other singleton was found to

have a 3’ UTR variant (c.*658T>A), however this variant had a CADD score of 12.9 and thus did not pass the filter.

Although mutations in the 3’ untranslated region, and splice sites of the COL4A1 gene have been described,[165,209] intronic variants have yet to be reported in both COL4A1 and COL4A2. In this cohort we found intronic variants passing filters in the COL4A1 and COL4A2 genes in two pedigrees with haemorrhagic strokes. These pedigrees are described in Table 5-7 and **Pedigree F104**.

A pedigree with a missense variant in COL4A1 is also discussed later in the chapter (**Pedigree F103**).

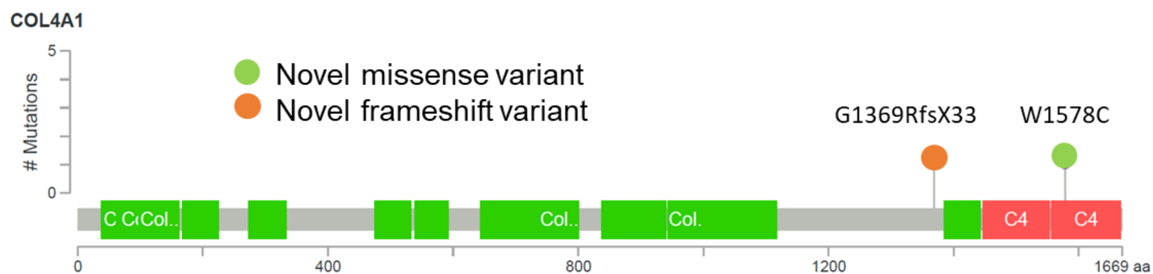


Figure 5-13: *Heterozygous missense and frameshift variants identified in COL4A1. p.Gly1369ArgfsTer33 is found within the Gly-X-Y tripeptide repeat of the α1 procollagen chain. p.Trp1578Cys of COL4A1 is found within the NC1 domain, a region which uses methionine-lysine covalent bonds to stabilise the collagen trimer.*

Pedigree F104

A 50-year-old white British gentleman presented following his second subcortical haemorrhage, having suffered his first episode three years before. At the time of his first bleed he was found to be hypertensive, however his blood pressure had been well-controlled at the time of the second.

Following his first stroke he had developed significant depression and anxiety, and was also diagnosed with cognitive impairment, showing problems with executive functioning. This was in keeping with the features seen on MRI, including white matter hyperintensities and subcortical microbleeds (Figure 5-14)

His paternal grandfather died of a stroke at the age of 76, and there were no first-degree relatives who were known to have had strokes, dementia or other neurological symptoms, although his mother had died prematurely at the age of 64 from a malignancy. On whole genome sequencing, he was found to carry an intronic deletion in the COL4A2 gene (c.100-1916_100-1877delAGGGCTGTACTAGCTCACCTGTGTGG GGGGAGGGCAGTACinsGCGGGCGGGA). This variant was not carried by his asymptomatic father or asymptomatic 58-year-old brother. As both relatives are above the age at which MRI-changes are expected to be seen, imaging in both these individuals would help to rule this variant in or out as a likely cause for his disease.

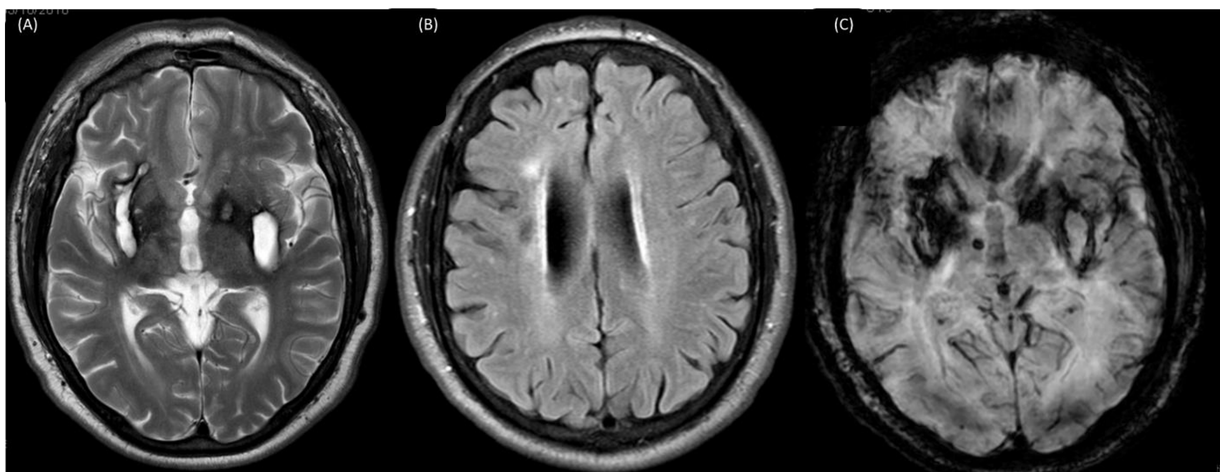


Figure 5-14: *MR imaging of the proband from Pedigree F104, showing (A) bilateral subcortical haemorrhages and (B) white matter hyperintensities on T2-weighted imaging, and (C) a microbleed in the right basal ganglia on susceptibility weighted imaging.*

Table 5-6: *Coding COL4A1 and COL4A2 variants passing filters, identified in SVD index cases. Variants that were not found in the index case of the pedigree, or did not segregate with the disease in the pedigree, are excluded from this list. All variants were absent in unrelated BRIDGE controls.*

Gene	Variant	Consequence	CADD	Diagnostic yield	AF	ExAC	gnomAD	UK10K	Functional impact
COL4A1	p.Gly1369ArgfsTer33	Frameshift	-	0.0042	0.0070	-	-	-	• Affects Gly-X-Y repeat • Likely results in loss of function
	p.Trp1578Cys	Missense	27.7	0.0042	0.0070	-	-	-	• p.Gly1580Arg/Ala previously reported.[269,296] • Within NC1 domain at C-terminus where cys residues highly conserved[266] • Additional cys may affect disulphide bond formation
COL4A2	p.Pro41Ser	Missense	24.9	0.0042	0.0070	2.5×10^{-5}	3.2×10^{-5}	-	• In 7S domain, important in covalent assembly of heterotrimers into procollagen molecule through disulphide bonds and lysine-hydroxylysine crosslinks.[297] • Large proline side chain is next to a cysteine -may affect disulphide bond interactions

NC1: Non-collagenous domain 1

7S: short N-terminal triple helical domain

Table 5-7: *Clinical features of probands with COL4A1 and COL4A2 variants*

Pedigree	Gene, Variant	Clinical features	Sex	Age*	MRI brain findings	Significant family history
F126	**COL4A1** p.Gly1369ArgfsTer33 **COL4A2** c.181-30810delG (Also carries **ADAR, CSF1R** non-coding variants)	**Haemorrhage in basal ganglia**	M	44	Bilateral subcortical haemorrhages, minimal WMH	• Brother – history of 'Bell's palsy'
F103	**COL4A1** p.Trp1578Cys (Also has RNASEH2B frameshift variant)	Migraine with aura Cognitive impairment Depression Dizziness and tinnitus	M	49	Multiple areas of WMH, multiple lacunar infarcts	• Father – stroke from age 65 • Mother – cognitive impairment and depression in early 60s • Daughter – migraines from age 26
F10	**COL4A1** c.144+1918_144+1919insAG	Migraine with aura TIAs	F	70	Confluent WMH affecting external capsules and anterior temporal poles, multiple lacunar infarcts	• Mother – stroke age 52 • Maternal grandmother – stroke in 50s • Maternal aunt – gait abnormalities in 30s
F16	**COL4A1** c.144+6701_144+6705delCTTTA	Migraine without aura Ataxia, slurred speech Cognitive impairment Depression	M	60	Confluent WMH, subcortical microbleeds, multiple lacunar infarcts	• Asymptomatic sibling shares this variant • Father – possible stroke or TIA at age 40

Pedigree	Gene, Variant	Clinical features	Sex	Age*	MRI brain findings	Significant family history
F23	**COL4A1** c.1897+559_1897+561delCCT	TIA Lacunar stroke	M	43	Multiple lacunar infarcts	• Mother – stroke age 59, severe depression • Maternal half-sister – stroke age 49, dementia age 52, severe depression
F29	**COL4A1** c.145-5665_145-5634delCACATATATATATATATACATATATATATATAinsTATT (Also has GSN downstream variant)	**Cortical haemorrhagic strokes**	M	50	Confluent WMH, cortical and subcortical microbleeds, superficial siderosis, multiple lobar haemorrhages. At age 56, this patient fulfils modified Boston criteria for probable CAA.[17]	• Maternal uncle – stroke and dementia (age at onset unknown) • Maternal niece: cerebral palsy, epilepsy, 'microcephaly' • Maternal nephew: 'arachnoid cyst', intellectual delay, epilepsy
F110	**COL4A2** p.Pro41Ser (Also has HTRA1 intron variant)	**Cortical (temporal lobe) haemorrhage**	M	49	Confluent WMH, multiple lacunar infarcts and haemorrhage, microbleeds	• Father – stroke from age 54, dementia
F79	**COL4A2** c.181-19325_181-19312delGTGGGTCCCGGTGTinsCACC (Also has CSF1R intron variant)	Migraine with aura Lacunar stroke Minor hand tremor and motor apraxia	F	61	Confluent WMH	• Consanguineous parents • Mother – TIAs, recurrent ICH from age 47. • Maternal grandmother – stroke from age 40. • Shares intronic variants in COL4A2 and CSF1R with asymptomatic, affected sister
F104	**COL4A2** c.100-1916_100-1877delAGGGCTGTACTAGCTCACCTGTGTGGGGGGAGGGCAGTACinsGCGGGCGGGA	**Subcortical haemorrhagic stroke** Depression Cognitive impairment	M	49	Confluent WMH, bilateral basal ganglia haemorrhages, microbleeds	• Variant not carried by father or two brothers (all asymptomatic) • Paternal grandfather – stroke age 76

Pedigree	Gene, Variant	Clinical features	Sex	Age*	MRI brain findings	Significant family history
F63	**COL4A2** c.100-22218_100-22217dupGC	Lacunar stroke	M	49	Patchy WMH, multiple lacunar infarcts	• Mother – stroke age 55 • Father – stroke age 75
F03	**COL4A2** c.1596+636_1596+637dupGG	Migraine with aura Depression Seizures Cognitive impairment	F	45	Confluent WMH	• Variant not carried by unaffected sibling (alopecia, degenerative disc disease, cervical dystonia, normal brain MRI) • Father – recurrent strokes from age 67 • Paternal uncle – stroke at age 65 • Mother –stroke age 69
F123	**COL4A2** c.180+7416_180+7417delCT	Migraine with aura Gait apraxia 'Myalgic Encephalomyelitis' – generalised aches, pains	F	40	Confluent WMH	• Mother – stroke, vascular dementia diagnosed at age 84 • Two sisters – diagnosed 'myalgic encephalomyelitis'
F35	**COL4A2** c.2425+1717_2425+1725delCATACACAC	Migraine aura without headache Lacunar strokes Cognitive impairment Gait apraxia	M	51	Confluent WMH, lacunar infarcts, microbleeds	• Paternal grandmother – dementia (age at onset not known, father died of non-neurological causes at age 54) • Mother – suicidal depression

Pedigree	Gene, Variant	Clinical features	Sex	Age*	MRI brain findings	Significant family history
F43	**COL4A2** c.99+16231_99+16232delTG (also carries downstream TREX1 variant)	Migraine aura without headache Gait apraxia Depression Cognitive impairment	F	65	Confluent WMH, lacunar infarct	• Brother – stroke from 57, cognitive impairment, depression, 'Meniere's disease' • Sister – migraine with aura • Daughter – Scheurmann's disease

*Age at onset of first major clinical feature, such as stroke, encephalopathy, seizures, cognitive impairment or dementia.

5.3.3.3 CARASAL: a third pedigree with the single known mutation

The CTSA gene encodes the Cathepsin-A enzyme, a subunit of a lysosomal multienzyme complex which has roles in the lysosomal transport, activation and stabilisation of β-galactosidase and neuraminidase-1. This enzyme also inactivates selected neuropeptides and regulates a lysosomal pathway of protein degradation.[34] Recessive (both homozygous and compound heterozygous) mutations in this gene are associated with galactosialidosis, a lysosomal storage disorder with both paediatric and adult onset. Infantile patients present with hydrops fetalis, visceromegaly, psychomotor delay, facial and skeletal dysmorphisms. Juvenile (2-7 years) and adult presentations (age 25-48 years) include myoclonus, ataxia, angiokeratoma and other neurological impairments.[34] One infant with galactosialidosis has been reported to also have cortical-subcortical infarctions.[298]

Heterozygous mutations in the CTSA gene were not identified until recently, in a form of familial SVD described as Cathepsin A-related Arteriopathy with Strokes and Leukoencephalopathy (CARASAL). CARASAL was described in two unrelated Caucasian Dutch families, with affected members in both families carrying the same rare missense variant in the CTSA gene (p.Arg325Cys) inherited in an autosomal dominant pattern.[299] Affected individuals were reported as having headaches, migraines, TIAs, mild cognitive impairment and both ischaemic and haemorrhagic strokes, and were diagnosed with leukoencephalopathy after the age of 40. Post-mortem studies in three affected individuals of a pedigree showed white matter atrophy, small infarcts in the white matter, deep grey matter and cerebellum, extensive leukoencephalopathy and fibrous thickening of the small vessels, as well as a higher number of oligodendrocyte progenitor cells (OPCs) than sporadic SVD controls. No GOM were found on electron microscopy.[299] The molecular pathogenesis of CARASAL has not been established, although the authors speculated that the inhibition of OPC maturation, in addition to vascular impairment, may contribute to the disease of the white matter.[299]

In this study we identified one singleton carrying a single copy of the same variant (p.Arg325Cys) in the CTSA gene, as described in **Pedigree F32**. Another upstream variant was found to be carried by a proband and their asymptomatic offspring, although the disease status of their child is unknown.

Pedigree F32

A white British female presented at the age of 68 with an episode of sudden-onset confusion, during which she was described as being unable to remember or recognise her family. She remained confused for five days after which she improved, but never returned to normal, complaining of memory problems. On cognitive testing she was found to have deficits in attention and orientation, memory, fluency and language, scoring 63/100 on the Addenbrookes Cognitive Examination III. She had another two episodes of confusion over the following year, and these were diagnosed as recurrent encephalopathic episodes. She also complained of fatigue.

She had a history of depression diagnosed at the age of 25 and was treated with citalopram and cognitive behavioural therapy in the community. She suffered a relapse at the age of 67 following a bereavement. On imaging her MRI showed extensive white matter hyperintensities extending into the anterior temporal poles and involving the external capsules. (Figure 5-15)

She had lost contact with her parents, who died in their 60s and 70s of unknown causes, and her brother. She had a sister who has asymptomatic at the age of 75.

She was tested for mutations in exons 2 to 24 of the NOTCH3 gene, common mutations in the POLG gene (for mitochondrial disorders) and the mitochondrial mutation m.3243A>G (causative of Mitochondrial Encephalopathy, Lactic Acidosis and Stroke-like episodes or MELAS), all of which did not yield positive results. A skin biopsy yielded no significant results on electron microscopy. On whole genome sequencing she was found to carry the p.Arg325Cys variant in the CTSA gene. Although little is known about the effects of this variant, the introduction of a cysteine residue may interfere with the formation of disulphide bonds and thus alter the three-dimensional conformation of the protein, or the formation of the lysosomal multienzyme complex.

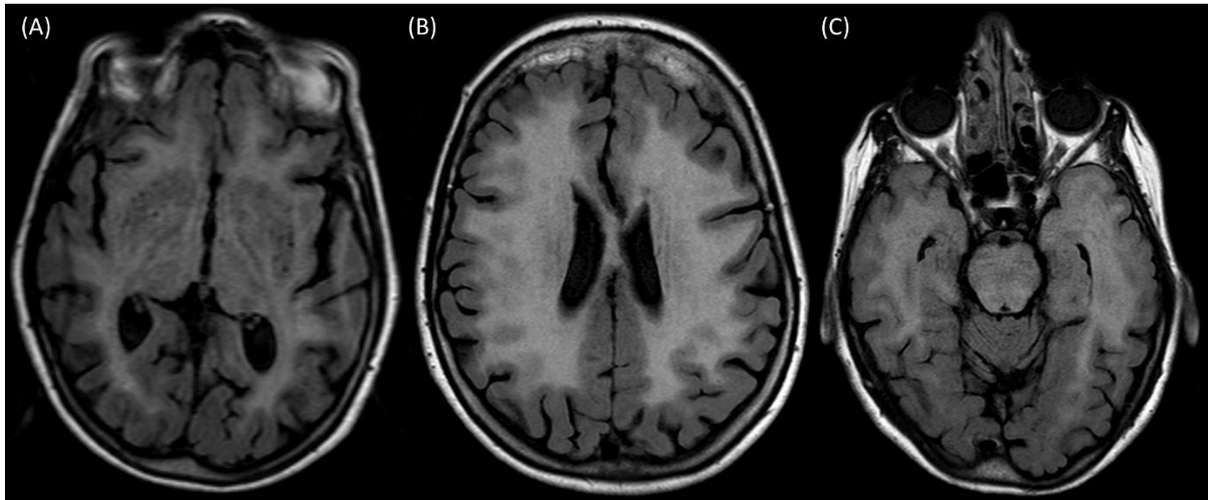


Figure 5-15*: T2-weighted FLAIR MR imaging of the proband from Pedigree F32, showing (A) and (B) confluent white matter hyperintensities (C) involving the anterior temporal poles*

5.3.3.4 Rare variants in familial hemiplegic migraine genes may be associated with SVD

The clinical features of familial hemiplegic migraine (FHM) are similar to some manifestations of cerebral SVD. Patients typically present with complicated migraine episodes and encephalopathy, and have shown reversible changes in cerebral blood flow on MR imaging, but do not have other MRI abnormalities beyond punctate white matter lesions expected in the general migraine population.[300,301]

In this study we have screened all participants for rare variants in six candidate genes associated with familial hemiplegic migraine. Only one coding variant passed our filters, as discussed here in **Pedigree F08**. A summary of all non-coding variants identified in these candidate genes is provided in the appendix.

Pedigree F08

A 66-year-old white British female was referred to the service from the Ear, Nose and Throat (ENT) surgeons. (Figure 5-16**, III-3**) She had been referred to the ENT service with a history of dizziness. On MR imaging she was found to have widespread white matter hyperintensities, most of which were peripheral and which were also affecting the anterior temporal poles. There were also marked perivascular spaces (Figure 5-17). There were also extensive perivascular spaces.

In the same year, she reported concerns with her memory and was found on neuropsychiatric assessment to have an impairment of her verbal and recognition memory, processing speed and complex executive function in keeping with a mild vascular cognitive impairment. She was described by her family as normally being 'an anxious personality', and reported a history of agoraphobia although this was not formally diagnosed.

She had a strong family history of strokes on both sides of the family. Her mother had also developed 'dizzy spells' in her seventies, and died from her first stroke at the age of 75. (**II-4**) Her father (**II-3**) died prematurely in the war, however his sisters (**II-1** and **II-2**) both died from their first strokes in their seventies. Her paternal grandmother (**I-2**) had also died from a stroke at age 67. Her brother (**III-3**) developed TIAs from the age of 79, and was found on MRI to have marked brain volume reduction and features of SVD. Her daughter (**IV-1**) reported experiencing 'word-finding difficulties' at the age of 42 and a past medical history of absence seizures, but had not had an MRI scan.

Both the proband and her brother (**III-3**) were recruited to the study, and both were found to have a novel rare missense variant, p.Gly2002Ser in the CACNA1A gene. The proband's daughter was also recruited, and was not found to carry this variant.

CACNA1A encodes the α_{1A} subunit of CaV2.1, a voltage-sensitive P/Q type calcium channel. CaV2.1 is highly expressed in the central nervous system and is essential to the release of vesicles containing neurotransmitters from the pre-synaptic terminal.[302] Autosomal dominant SNVs and CNVs in the CACNA1A resulting in loss of function of this channel are associated with Familial Hemiplegic Migraine Type I (FHM), Episodic Ataxia Type 2 and Spinocerebellar Ataxia Type 6.[303–305]

Segregation of the variant with disease in this pedigree could potentially be demonstrated with MR phenotyping of the proband's daughter (**IV-1**), as well as sequencing and imaging in the proband's sister (**III-1**).

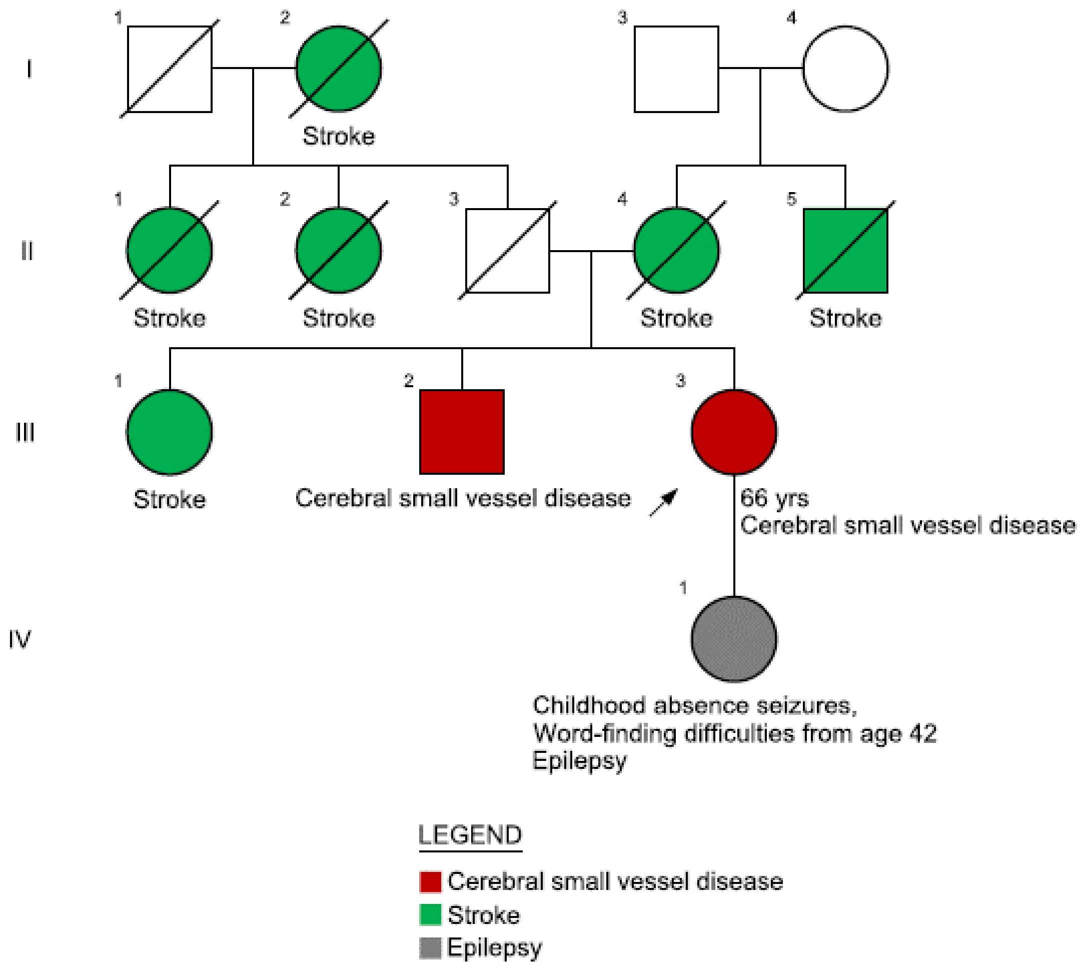


***Figure 5-16**: Pedigree F08*

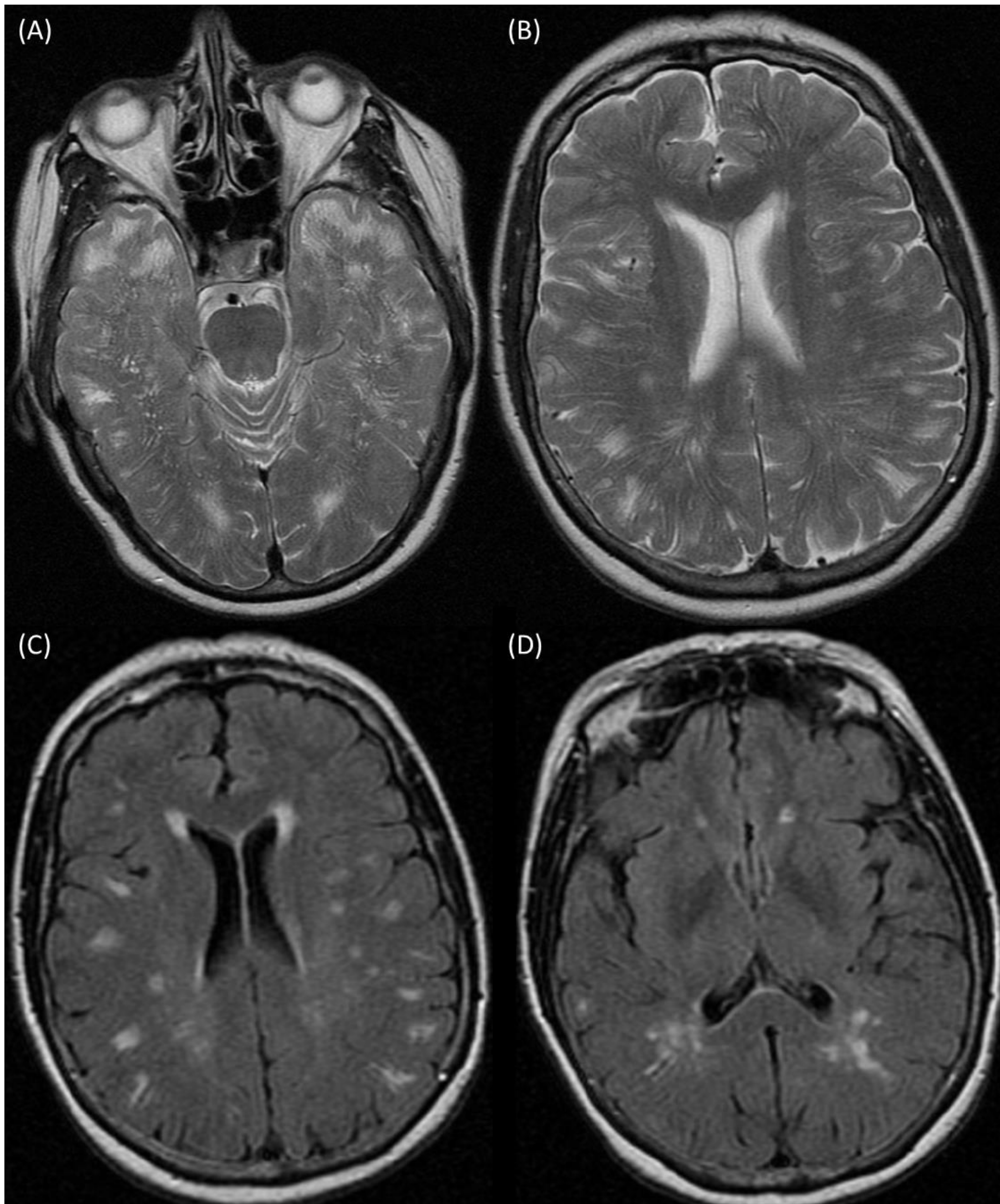


Figure 5-17: *MR imaging of the proband from Pedigree F08 showing (A) white matter hyperintensities in the anterior temporal poles (A), and marked dilated perivascular spaces (B) on T2-weighted imaging. The white matter hyperintensities were mostly peripheral on T2 and T2 FLAIR imaging (B and C), and also affected the periventricular regions (C and D).*

5.3.3.5 Rare variants in genes associated with sporadic SVD may be implicated in monogenic SVD

Common variants in monogenic SVD genes have been demonstrated to have a possible association with features of SVD such as WMH in both stroke and community-based populations.[157,166] We hypothesized that rare variants in gene loci recently identified in GWAS of sporadic SVD may also be associated with monogenic disease.

In this study we have screened 30 genes near loci with suggestive levels of significance in GWAS of sporadic SVD. Coding variants (missense, nonsense, frameshift and in-frame insertions or deletions) were found in seven genes. Only one variant was identified in each of these genes, and none of the variants were found in more than one pedigree. An example of one of these pedigrees is discussed below in **Pedigree F18**, and a summary of these variants is provided in Table 5-8. Non-coding variants identified in these genes are summarised in the appendix at the end of the chapter.

Pedigree F18

A 64-year-old White British female (Figure 5-18**, III-3**) was seen by the ENT surgeons having complained of a longstanding sore throat and swallowing problems. She was diagnosed with gastro-oesophageal reflux disease, however routine MR imaging of the head and neck revealed areas of white matter hyperintensities suggestive of SVD (Figure 5-19A), following which she was referred to the Neurology service. She had no previous history of migraine, stroke and other neurology or psychiatry.

She had a strong family history of strokes on the maternal side. Her mother (**II-2**) had a stroke at the age of 72 and a diagnosis of ‘Parkinson’s disease dementia’, which on retrospect may have been a vascular dementia. Her maternal uncle (**II-3**) also had a diagnosis of ‘Parkinson’s disease’. Her maternal grandfather (**I-3**) died of a stroke at the age of 50 although this was thought to have been secondary to a brain tumour. Her maternal grandmother (**I-4**) died of a stroke at the age of 70.

Her brother (**III-2**) had a history of multiple small vessel ischaemic strokes from the age of 57. At age 58, he was diagnosed with vascular cognitive impairment, having difficulties with attention, concentration, executive function and short-term memory. He also suffered from

tinnitus, dizziness and progressive deafness, and was diagnosed with Meniere's disease. He had a past medical history of depression from the age of 19, and also reported a previous history of 'episodes of confusion' in his late thirties. On MR imaging he was found to have white matter hyperintensities and lacunar infarcts. (Figure 5-19**B**)

Her sister (**III-1**) had a history of migraine aura without headache from the age of 51, but otherwise had no other neurological symptoms and had not had an MRI scan. All three siblings were recruited for whole genome sequencing. Both the proband and her brother were found to carry a rare (absent in gnomAD) p.Val224Leu variant in the FOXC2 gene.

FOXC2 is found within the 16q24 locus, which contains a SNP associated with SVD which reaches genome-wide significance on the largest GWAS in SVD to date.[179] The FOXC2 gene encodes for a member of the Forkhead 'winged-helix' family of transcription factors (Foxc2).[306] The Foxc2 transcription factor is involved in multiple developmental pathways. Both Foxc1 and Foxc2 are critical in cardiovascular development, and are expressed on the endothelium and smooth muscle cells of blood vessels.[306,307] Autosomal dominant mutations in the FOXC2 gene have been described in lymphedema-distichiasis, a congenital form of lymphoedema associated with double rows of eyelashes, cardiac defects, cleft palate, extradural cysts and photophobia.[308]

Like other forkhead transcription factors, it has been suggested that Foxc2 has different roles in development and in the adult. Foxc2 has been shown to have a role in glucose metabolism in the adult, with increased Foxc2 expression counteracting mechanisms associated with type 2 diabetes mellitus and obesity.[309] Altered Foxc2 function leading to impaired glucose metabolism in the brain may thus be a plausible mechanism for the development of SVD in this pedigree.

The variant identified lies 102 residues 3' of the forkhead motif in the Foxc2 protein, and its significance on the Foxc2 protein function and disease is thus unknown. Although the FOXC2 gene has not previously been associated with any SVD-related phenotype, its roles in both vasculature development and glucose metabolism may contribute to the development of SVD.

In this pedigree, MR imaging to confirm the disease status of the proband's sister (III-2) would help to demonstrate or disprove segregation of the variant with disease in the pedigree.

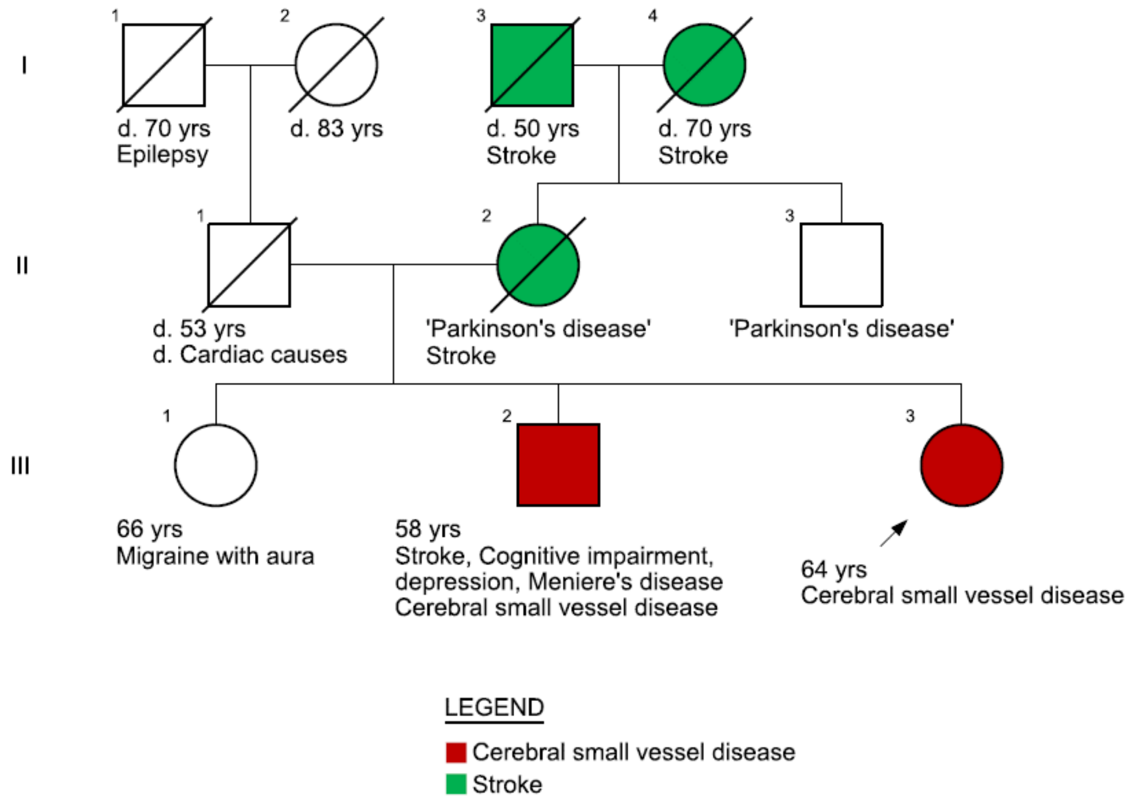


Figure 5-18: *Pedigree F18*

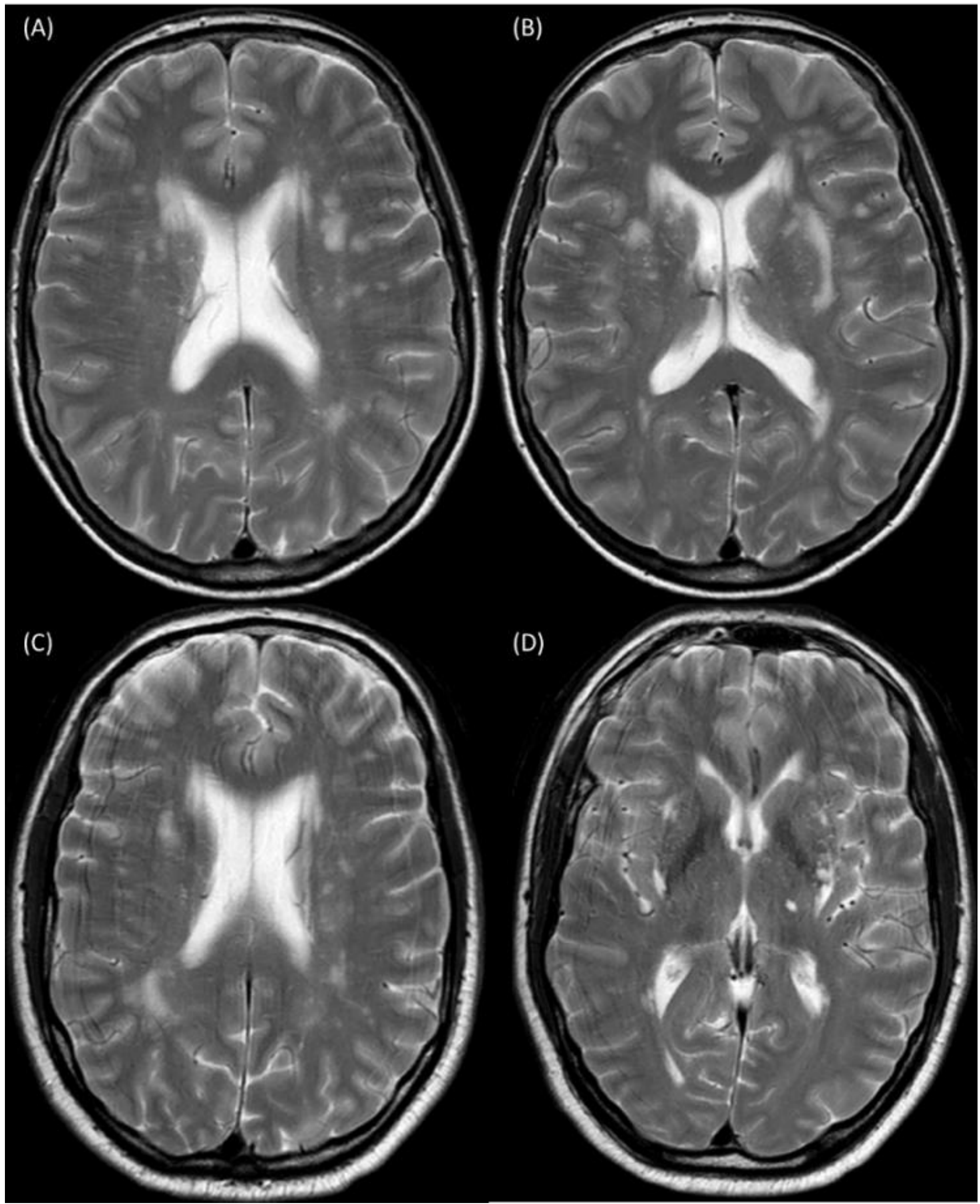


Figure 5-19: *MR imaging of the proband and her brother in pedigree F18. T2-weighted MR imaging of (A and B) the proband, showing white matter hyperintensities, and (B and C) her brother, showing white matter hyperintensities (C) and an old thalamic lacune (D).*

5.3.3.6 Patients with rare variants in multiple candidate genes

Several patients were found to carry variants passing filters in more than one candidate gene, as summarised in the appendix. This is particularly pertinent in larger genes with long non-coding regions (Figure 5-4). For example, there were 34 unique intronic SNVs identified in the CACNA1A gene, which has a non-coding region of 409156bp in length, as opposed to NOTCH3, with 33277bp of non-coding regions and only 2 intronic variants identified. Excluding variants which did not segregate with disease in the pedigrees, there were 18 intronic variants in CACNA1A, of which two were in carriers who had competing genetic diagnoses in the form of missense or nonsense variants in the HTRA1 or NOTCH3 genes (cases described earlier). Although intronic variants may have consequences (such as through the introduction of a new splice site), it is difficult to identify which ones are of significance. An example of a patient carrying rare variants in two different candidate genes is described in **Pedigree F103**.

Pedigree F103

A white British male (Figure 5-20**, III-1**) presented at age of 49 having been diagnosed with cognitive impairment, showing problems in attention, concentration and working memory, as well as in semantic fluency. He also developed depression in that year, and by the following year had new onset dizziness and nausea, was unsteady on his feet, along with tinnitus. He had a past medical history of migraine with visual aura beginning in his teenage years.

On MR imaging of the brain at age 49 he had multiple areas of white matter hyperintensities not involving the anterior temporal poles, as well as lacunar infarcts (Figure 5-21). He did not have evidence of microbleeds or superficial siderosis on susceptibility weighted imaging.

He had a positive family history for stroke, although this did not show a specific pattern of inheritance. His father (**II-1**) suffered his first stroke at age 65 and died of a stroke at age 67, while his mother (**II-2**) had a history of cognitive impairment and depression in her early 60s, dying of unspecified causes at the age of 68. His son (**IV-**2) had a history of migraines, but had not yet had an MRI scan of his brain.

On whole genome sequencing, the proband was found to carry a missense variant p.Trp1578Cys in the NC1 domain of the COL4A1 gene. Although both tryptophan and cysteine are non-polar, tryptophan is the amino acid with the largest side chain and a variant at this

position is likely to affect the three-dimensional conformation of the protein. Furthermore, the NC1 domain is involved in the assembly of the collagen heterotrimer, and cysteine residues in this domain are highly conserved as they serve to cross-link triple-helical molecules via disulphide bonds.[266] The introduction of an additional cysteine residue may impact the correct formation of such bonds and thus lead to dysfunctional collagen molecules.

The proband was also found to carry a frameshift variant, p.Val41LeufsTer23 in the gene RNASEH2B. RNASEH2B encodes for the Subunit B protein, a component of ribonuclease H2 which serves to cleave ribonucleotides at sites of DNA replication and repair.[310] Homozygous and compound heterozygous mutations in RNASEH2A, RNASEH2B and RNASEH2C, which make up the ribonuclease H2 complex, as well as both recessive and dominant mutations in TREX1, SAMHD1, IFIHD1 and ADAR have been associated with the Aicardi-Goutières syndrome (AGS). AGS is a paediatric-onset disease characterised by encephalopathy, basal ganglia calcification and white matter abnormalities.[311] Although the variant in this case is heterozygous, it is not possible to exclude a haploinsufficiency or dominant negative effect resulting in a less severe, adult-onset form of Aicardi-Goutières syndrome as seen in TREX1, or a combined effect of both the COL4A1 and RNASEH2B variants.

Determining which, or if both or neither of these variants are causative for disease in such a pedigree is particularly challenging. It may help to sequence and perform brain imaging in the other relatives in this pedigree, although none of those surviving are known to be symptomatic, and many are below the age at which they would be expected to show changes on MRI. The presence of the same variant(s) in other affected pedigrees with similar phenotypes could serve as additional evidence for their pathogenicity.

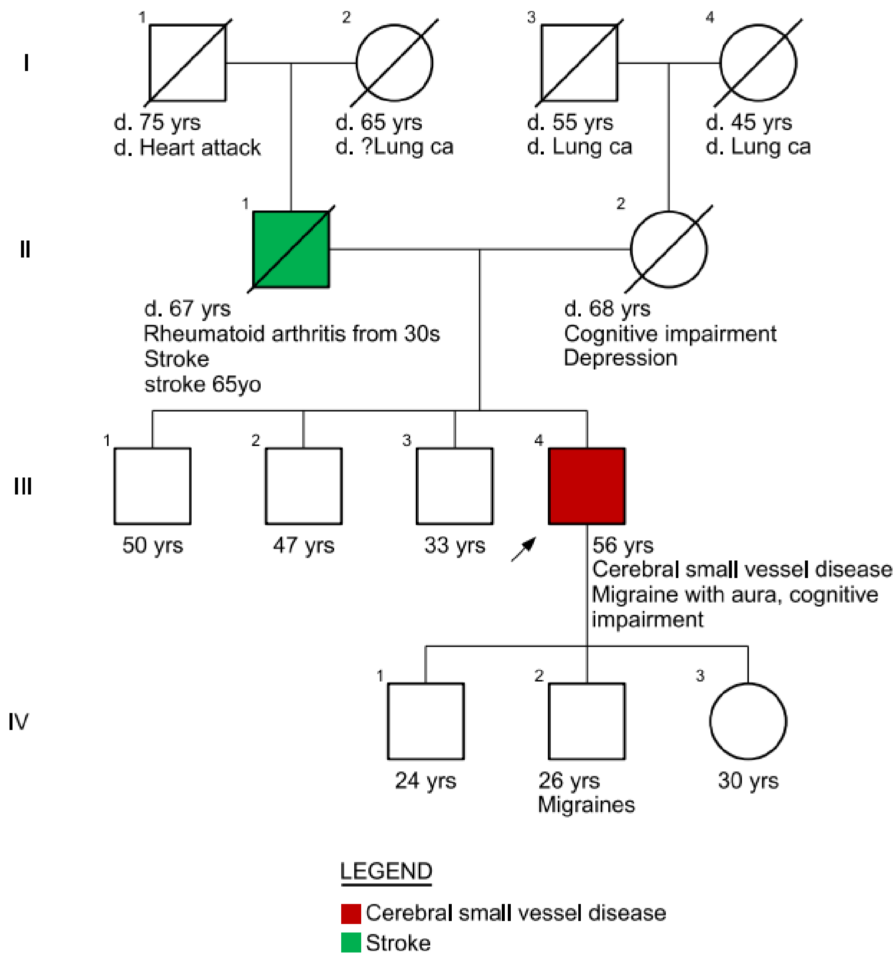


***Figure 5-20**: Pedigree F103*

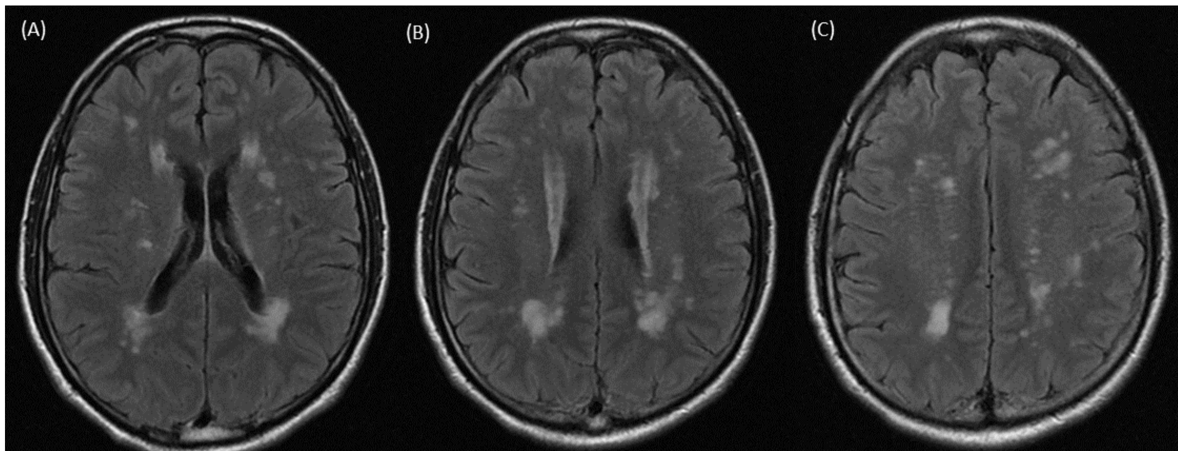


***Figure 5-21 (A-C)**: MRI T2-FLAIR images of the proband from Pedigree F103, showing multiple areas of white matter hyperintensities.*

Table 5-8: *Coding (missense, nonsense, frameshift, inframe insertion and inframe deletion) variants passing filters in all other genes, which have not been excluded based on segregation analysis. Coding variants in NOTCH3, HTRA1, CTSA, COL4A1 and COL4A2 have been described earlier. All variants reported here were not previously reported in mutation databases (HGMD, ClinVar and LOVD).*

	Gene	Associated phenotype	Pedigree	Variant identified	Variant type	Sex	Clinical features	MRI brain findings	Segregation/ remarks	Frequency in gnomAD
Monogenic SVD	None in addition to those described in NOTCH3, HTRA1, CTSA, COL4A1, COL4A2.									
FHM	CACNA1A	FHM	F08	p.Gly2002Ser	Missense	F	Cognitive impairment Dizziness Sib: TIAs	WMH involving anterior temporal poles. Sib: WMH	Affected sib-pair both carry variant	-
Amyloid	GSN	Amyloidosis	F51	p.Gln528Pro	Missense	M	SV ischaemic stroke Depression Migraine without aura	WMH, multiple LIs	Singleton	8.2 x 10^{-6}
CTD	COL3A1	Ehlers Danlos Type IV	F98	p.Pro137Ser	Missense	F	SV ischaemic stroke Migraine with aura		Singleton	-

	Gene	Associated phenotype	Pedigree	Variant identified	Variant type	Sex	Clinical features	MRI brain findings	Segregation/ remarks	Frequency in gnomAD
	ABCC6	Pseudoxanthoma Elasticum (recessive)	F91	p.Gly992Arg heterozygous	Missense	M	Vascular dementia Epilepsy Progressive spasticity	WMH, multiple LIs	Singleton	2.5×10^{-5}
Leukodystrophy	EIF2B1	Vanishing white matter disease (recessive)	F33	p.Val24Met heterozygous	Missense	F	Migraine with aura Aura without headache	WMH	Parent (unknown status) does not carry variant	-
Leukodystrophy	GFAP	Alexander disease	F82	p.Arg29Cys	Missense	M	Migraine with aura Aura without headache Depression SV ischaemic stroke	WMH, LIs	Singleton	4.5×10^{-5}
Leukodystrophy	LMNB1	Adult-onset leukodystrophy	F113	p.Glu559Ala	Missense	M	SV ischaemic stroke Migraine without aura Depression Cognitive impairment	WMH	Singleton	3.2×10^{-5}

	Gene	Associated phenotype	Pedigree	Variant identified	Variant type	Sex	Clinical features	MRI brain findings	Segregation/ remarks	Frequency in gnomAD
Sporadic SVD	CARF	-	F97	p.Thr657Ile	Missense	M	Migraine with aura Depression	WMH, LI	Singleton	2.8×10^{-5}
	FOXC2	Lymphedema-distichiasis syndrome	F18	p.Val224Leu	Missense	F	None – incidental MRI findings Sib: SV ischaemic stroke, depression, cognitive impairment, Meniere's syndrome (dizziness, tinnitus, deafness)	WMH Sib: WMH, LIs	Affected sib-pair both carry variant	-
	MLC1	Megalencephalic leukoencephalopathy with subcortical cysts	F65 (also has CADASIL-causing mutation)	p.Val354_Ala355insGlyGlyLeuTrpGlyGlyGlyAlaGlyGluVal	Inframe insertion	M	SV ischaemic stroke Migraine without aura	WMH involving left temporal pole	Singleton 11aa insertion at position 351-352 reported as polymoprhism[312]	-
	RNASEH2A	Aicardi-Goutières syndrome	F74 (also carries CADASIL-causing mutation)	p.Gln126His	Missense	M	SV ischaemic stroke	WMH involving anterior temporal poles, multiple LIs	Singleton	1.2×10^{-5}

Gene	Associated phenotype	Pedigree	Variant identified	Variant type	Sex	Clinical features	MRI brain findings	Segregation/ remarks	Frequency in gnomAD
RNASEH2B	Aicardi-Goutières syndrome	F103	p.Val41LeufsTer23	Frameshift	M	Migraine with aura Depression Cognitive impairment Dizziness	WMH, LIs	Singleton. Also carries COL4A1 missense variant	-
SAMHD1	Aicardi-Goutières syndrome	F64	p.Leu493Pro	Missense	M	SV ischaemic stroke	WMH, LI	Singleton	-
ZNF474	-	F29 (also carries COL4A1 intron variant)	p.Arg102Ter	Nonsense	M	SV haemorrhages Depression Anxiety Cognitive impairment	Superficial siderosis Cortical and subcortical microbleeds WMH	Singleton	2.5×10^{-5}

WMH: white matter hyperintensities, LI: lacunar infarct, SV: small vessel

5.4 Conclusions

In this study of 118 pedigrees with suspected monogenic SVD, we have screened for the presence of rare variants across 73 candidate genes associated with both familial and sporadic SVD, associated phenotypes such as familial hemiplegic migraine, amyloidosis, connective tissue disease and leukodystrophy. This cohort of individuals has a broad range of disease severity both between and within families, with some index cases having SVD diagnosed only as an incidental finding on MR imaging.

Three cysteine-altering variants in exons 3 and 4 of NOTCH3 were identified in patients previously screened for NOTCH3 mutations on older technology such as DHPLC, and Sanger sequencing which is currently the gold standard for clinical genetic testing. Although these may be false positive calls on whole genome sequencing, the validation of one of these variants in subsequent testing in a clinically accredited lab makes this less likely. These findings highlight the limitations of existing technology used in the clinic, and demonstrates the benefits of using HTS which allows better sensitivity in detecting mutations.

We have described a patient with two non-cysteine changing variants in cis on the NOTCH3 gene, and were unable to demonstrate a clear impact of these variants on disulphide bond formation using molecular modelling. In line with the current consensus on NOTCH3 variant interpretation, our discovery of only two pedigrees with non-cysteine altering variants suggest that these variants in NOTCH3 are indeed unlikely to cause familial SVD. However, this study was limited by the lack of MRI-confirmed phenotyping of relatives and limited in-silico modelling of the variants, making it is difficult to be certain whether these variants are segregating with disease.

Heterozygous HTRA1 variants appear to be the next most common cause of familial SVD after CADASIL, with missense and nonsense variants occurring in 5.1% of index cases and 6.3% of affected cases (including relatives). This is similar to frequencies seen in other cohorts of NOTCH3-negative familial SVD patients in Italy (6.3% of all affected cases),[265] France (4.97% of index cases)[67] and Japan (5.3% of index cases).[68] Consistent with published literature, these families show a milder phenotype than that seen in CARASIL, and also significant heterogeneity, with some patients with SVD showing minor or non-specific symptoms such as migraine which may not typically lead to presentation to the stroke service. Although only

missense, and one nonsense (p.Arg302Ter) mutations were thought to cause disease through a dominant negative effect, we have shown here three individuals with SVD carrying the p.Arg370Ter variant, suggesting that haploinsufficiency due to p.Arg370Ter may also be sufficient to cause a milder form of CARASIL.

Fewer variants in COL4A1/A2 were identified. These patients had phenotypes that largely fit with the phenotype of published cases, with the predominant clinical features being a combination of ischaemic and haemorrhagic strokes and subcortical microbleeds. A spectrum of phenotypes was also seen with variant carriers, with some patients having predominant cognitive impairment and no strokes. The majority of variants identified were non-coding, and these may potentially be pathogenic variants which create new splice sites which alter the Gly-X-Y helical repeat region, as recently demonstrated in COL6A1, associated with COL6-related dystrophy.[284] The impact of the non-coding variants on the resulting mRNA may be demonstrated on RNA sequencing.

Both HTRA1 and CTSA are conventionally associated with early onset autosomal recessive disease. HTRA1 was first described in early onset SVD in CARASIL, with patients in their 20s developing strokes. CTSA was first described in galactosialidosis, a disease with paediatric onset, and was only recently linked with autosomal dominant SVD. Our findings in HTRA1 and CTSA support recent reports that heterozygous variants in these genes are associated with adult-onset familial SVD. Despite this being a hereditary disease, its relatively late onset may have obscured the uncovering of its genetic basis. It appears that a single copy of mutations in genes previously known to cause early-onset autosomal recessive diseases may be sufficient to allow relatively normal development, but may cause cerebrovascular disease in the later stages of life. Beyond looking for novel genes in undiagnosed rare diseases, it is thus also worth re-examining the functions and involvement of genes previously implicated only in recessive disorders.

We have shown that multiple pedigrees were found to have variants in more than one candidate gene. In this chapter we have only screened 73 genes – on a genome-wide scale it is likely that every individual has hundreds or thousands of possible variants. This highlights a key challenge in the clinical interpretation of results from whole genome sequencing. In the absence of MR imaging of all adult members of each pedigree it is difficult to determine if one or multiple variants are disease-causing. In addition, majority of these variants are novel non-

coding variants, for which there is limited to no functional data available, and these may not be present in existing control databases. Hence even with clear segregation demonstrated it would be premature to conclude on the pathogenicity of variants without showing the functional impact or true population allele frequency of these variants.

The utility of in silico predictive tools such as the CADD score, and in vivo and in vitro functional studies is also limited. As demonstrated in one of our pedigrees where a 3'UTR variant has been shown to segregate with disease in a separate study,[209] but did not pass filters in our study, arbitrary cut-off values for predicted pathogenicity and minor allele frequency may result in the exclusion of disease-causing variants. Furthermore, diseases may not necessarily result from loss- or gain-of-function mutations as seen in CADASIL (Chapter 1). A clear understanding of the mechanisms of disease at the molecular and histopathological level is essential to confirm a genotype-phenotype association.

The role of monogenic disease genes in sporadic SVD has been suggested in recent years, and it is possible that the reverse is also true. In this study we have proposed a possible role for rare variants in familial hemiplegic migraine genes (CACNA1A) and genes in loci associated with sporadic SVD (CARF, FOXC2, MLC1, RNASEH2A, RNASEH2B, SAMHD1, ZNF474). We have demonstrated two pedigrees each with two affected individuals carrying rare novel variants in CACNA1A (Pedigree F08) or FOXC2 (Pedigree F18). Although CACNA1A has not previously been associated with white matter abnormalities, and FOXC2 has only purported roles in the developing vasculature and glucose metabolism, it is possible that these have pleiotropic effects. Rather than a dichotomous distinction between monogenic and sporadic SVD or familial hemiplegic migraine and familial SVD with complicated migraines, it may be more likely that patients lie on a spectrum between the two, with each individual's specific environment (in this context, cardiovascular risk factors) influencing the age at onset of disease.

Further work for this data set would require phenotyping of all sequenced symptomatic and asymptomatic adult relatives using MR imaging. This would allow the use of linkage analysis to narrow down gene loci segregating with disease, as well as filtering within trios in cases of suspected de novo or recessive disease. This would allow a hypothesis-free approach rather than limiting the search to known genes associated with neurovascular or white matter disease. Other candidate genes, such as those encoding proteins in the extracellular matrix of

the brain vasculature, may also have a role in disease, and a targeted screen of these genes may also reveal novel genes involved in the disease process. Finally, gaining a better understanding of the non-coding space, such as by using RNA-sequencing analysis to determine the impact of intron variation, may potentially reveal many more disease-causing variants.

5.5 Appendix

Table 5-9: *Summary of variants identified in all 118 pedigrees. Variants which do not segregate with the pattern of inheritance of disease in the pedigree have been excluded from this list.*

Pedigree	Candidate gene	SNV	SNV type	CNV within gene	CNV type	Remarks
F01	APP CACNA1A CECR1	c.1458+5711_1458+5722delTTTTTTTTTTTT c.2172+998_2172+1001delTTTC c.753+2659delT	Intron Intron Intron	RP11-521D12.2 RP11-83M16.6 DNAH8 PBX4	Deletion Deletion Deletion Deletion	
F02	APP	c.58-10602_58-10576delATTTTATTAAAAT CTTCTAATAAAAAA	Intron	AC062021.1 DLAT RP11-551L14.1	Deletion Deletion Deletion	
F03	COL4A2 SAMHD1 PRDM16	c.1596+636_1596+637dupGG c.696+584_696+585dupTT c.439-33525_439-33522dupCCCC	Intron Intron Intron	RP11-2E17.1 RP4-568F9.6	Deletion Deletion	
F04	ATP1A3 CACNA1A NBEAL1	19:42501940_C/CCCCCCACG c.4591-732_4591-730delAAG c.6508+413_6508+417dupGCCAG	Upstream Intron Intron	CYP20A1 LAT2 RNA5SP347 ZSCAN9, ZKSCAN4	Deletion Deletion Deletion Deletion	
F05	HTRA1 APP	p.Pro331Leu c.58-11781_58-11778delAACA	Missense Intron			
F07				EPCAM, MIR559	Deletion	
F08	CACNA1A FOXC2 MIR216A	p.Gly2002Ser 16:86600205_A/G 2:56216414_G/A	Missense Upstream Upstream			

Pedigree	Candidate gene	SNV	SNV type	CNV within gene	CNV type	Remarks
F09	ATP1A2 CACNA1A	1:160084457_AG/A c.978+1429_978+1431dupATG	Intron Intron	RP11-402J7.2	Deletion	
F10	COL4A1	c.144+1918_144+1919insAG	Intron			
F11	ATP1A2 PRDM16 NBEAL1	1:160085222_G/GCCACATGGT c.38-23254_38-23253insGACGTGTGCT CTTCCGATCTTGCGCAGACGCCCCTGGCT CTAGAGGGC c.51+12044_51+12045delTT	Upstream Intron Intron	TEX14 AC111155.1	Deletion Deletion	
F12	COL3A1	2:189879981_A/AT	Downstream	CTD-2269F5.1 EXD1, RN7SL497P	Deletion Deletion	
F13	APP PRDM16 SH3PXD2A SH3PXD2A EIF2B2 MIR216A / MIR217 MIR216B	c.1458+5147_1458+5148insCCTTTCCTTTCC TTTCCTTTCCTTTCCTTT 1:3359482_GTGT/G c.72+26561G>T c.473-1239_473-1231delGCCCCCCCC c.434-46_434-32delTTTTTTTTTTTTTTTinsATATATATATATA 2:56212347_C/T 2:56229175_C/T	Intron Downstream Intron Intron Intron Downstream Downstream	OR5C1 RP11-430B1.2, MYO5C	Deletion Deletion	
F15	PRDM16 ZCCHC14	c.38-4765_38-4758dupGGTCTTAC c.*3566T>C	Intron 3'UTR	NMRK2 SLC39A11, SCARNA24	Deletion Deletion	
F16	COL4A1 KIF18B PRDM16 TUBB4A	c.144+6701_144+6705delCTTTA c.-15+4779_-15+4781delTGT c.3285-2204_3285-2198delCCTGCCC c.278-2010delT	Intron Intron Intron Intron	RALGDS, GBGT1	Deletion	
F17	EARS2	c.295+1153G>C	Intron	AC114765.3	Deletion	
F18	FOXC2 APP	p.Val224Leu c.226-9646_226-9643delTTAG	Missense Intron	DNAH2	Deletion	

Pedigree	Candidate gene	SNV	SNV type	CNV within gene	CNV type	Remarks
F19				Y_RNA	Deletion	
F20				MYOM2	Deletion	
F21				PKD1L2 MMP9 PTGIS	Deletion Deletion Deletion	
F23	COL4A1 FAM126A SH3PXD2A	c.1897+559_1897+561delCCT c.626+3795G>A c.473-1228_473-1227insCG	Intron Intron Intron	HYAL1 WDR49 TRMT44 ZNF682	Deletion Deletion Deletion Deletion	
F24				SLC3A1,PREPL,CAMKMT AC005176.2, AC005176.3, LGALS16, LGALS17A, AC093063.3, AC093063.2, AC006133.3, LGALS14, CLC, LEUTX RERGL RP11-331H13.1, SCAPER, MIR3713	Deletion Deletion Deletion Deletion	
F25	GSN	9:124098555_GC/G	Downstream			
F26	KIF18B	17:43029291_CTCTC/ATATATATATATA	Upstream	AC002485.1, RP11-24C14.1 NCAM2	Deletion Deletion	
F27	EARS2	c.1353-1591delT	Intron	NDUFAF2, AC008498.1 VPS37A, MTMR7 LINC00423, TOMM22P3, KL, AL161898.1, STARD13, STARD13-IT1, STARD13-AS, RP11-141M1.1, AL138999.1, RP11-141M1.3, RP11-37L2.1, RP11-179A7.2	Deletion Deletion Deletion	
F28	HTRA1 CACNA1A TUBB4A ICA1L	p.Val175Leu c.540-5082_540-5081delTA c.*128delC c.-8+10250_-8+10251insTTT	Missense Intron 3’UTR Intron	GRB10	Deletion	

Pedigree	Candidate gene	SNV	SNV type	CNV within gene	CNV type	Remarks
F29	COL4A1 ZNF474 GSN	c.145-5665_145-5634delCACATATAT ATATATATACATATATATATATAinsTATT p.Arg102Ter 9:124099535_C/G	Intron Nonsense Downstream	L3HYPDH, RP11-701B16.2	Deletion	
F31	SNORD118	17:8076200_GA/G	Downstream	NLGN1-AS1	Deletion	
F32	CACNA1A CACNA1A CTSA ZNF474	c.4866+727_4866+728delCT c.1669-1930_1669-1926delTTTTG p.Arg325Cys c.-213+1131_-213+1134delTCAT	Intron Intron Missense Intron			CARASAL-causing mutation
F33	EIF2B1 FAM126A FAM126A EFEMP1	p.Val24Met c.743+407A>C c.626+3782G>A c.517+7277C>T	Missense Intron Intron Intron	TBC1D2B, SH2D7	Deletion	
F34	HTRA1 CSF1R SH3PXD2A	p.Arg370Ter c.49+1988_49+1989insCTTCCTTCTTTCTTTCTT TC c.*5178_*5187dupGTGTCTGTGT	Nonsense Intron 3’UTR	RPL7P9, RN7SKP270, PTBP2, AL592205.1, AL592205.2, DPYD, DPYD-AS1 LINC00877	Deletion Deletion	
F35	COL4A2 FOXF1 LMNB1	c.2425+1717_2425+1725delCATACACAC 16:86541560_T/TG c.359+13169_359+13172dupTAAA	Intron Upstream Intron			
F37	NOTCH3 CACNA1A NBEAL1	p.Gly52Arg p.Asp45His c.6780+227_6780+228insCCCCCC c.293+15236_293+15243delTTTCCTTT c.516-1069_516-1067delACA	Missense Intron Intron Intron			NOTCH3 variants in cis. Neither carried by unaffected relative and relatives of unknown disease status.
F39 (=F54)	HTRA1 ATP1A2	p.Pro285Leu c.1828-236_1828-231delCACACA	Missense Intron			

Pedigree	Candidate gene	SNV	SNV type	CNV within gene	CNV type	Remarks
	ABCC6	c.3634-43_3634-39delAGTGG	Intron			
F40	HTRA1 CACNA1A SAMHD1 SAMHD1	c.777+6599delG c.1346-1598_1346-1594dupAAGAT c.1062+550delCinsTTTTTTTTTTTT c.209-2090_209-2079delTTTTTTTCTTTT	Intron Intron Intron Intron	GCNT2 TMEM135 ITCH	Deletion Deletion Deletion	
F42	ATP1A3 SH3PXD2A	19:42502584_G/A c.1225-2233delAinsGTAAACTTTTTTTTTTTTTTTTT	Upstream Intron			
F43	COL4A2 APP TREX1 CARF WDR12	c.99+16231_99+16232delTG c.58-20144delC 3:48512824_TACAAAAAT/C c.1494+251dupA c.741+866_741+867dupAA	Intron Intron Downstream Intron Intron	ARL2BPP7 MYH1 EFCAB13	Deletion Deletion Deletion	
F44	NOTCH3 ATP1A2 EVL EVL PRDM16 HEPACAM EFEMP1	p.Cys212Arg 1:160118195_TAGAG/T 14:100528561_A/ATG c.180+5977_180+5980dupATCC c.38-9424_38-9423delTGinsCGTCCATCTATCCA c.86-4128delT c.517+12656_517+12657delTG	Missense Downstream Upstream Intron Intron Intron Intron			CADASIL-causing mutation
F45	CACNA1A NBEAL1	c.1255+465dupT c.1098+4385dupT	Intron Intron			
F47	ABCC6 DEGS2 EVL FAM126A	c.601-361delC c.826-752_826-751insGGGGC 14:100613995_G/GGCCCC c.-28-8949_-28-8947delTTA	Intron Intron Downstream Intron	RAB11FIP5, RP11-44N22.3 RFPL4AL1, RFPL4AP1	Deletion Deletion	

Pedigree	Candidate gene	SNV	SNV type	CNV within gene	CNV type	Remarks
F48	CACNA1A LOX	c.539+22891_539+22912delACCCTACT CCACACTCCCCACTinsCCCCCCCC c.539+22882delAinsCCCCCCCCCCCCCCCCCTC c.1036-95A>G	Intron Intron Intron			
F50	APP PRDM16	c.1091-1210_1091-1207delATCA c.57+24642_57+24645delTAAA c.38-4431_38-4419delTCTTACACACGGTinsC	Intron Intron Intron	TRAJ46, TRAJ45, TRAJ44, TRAJ43, TRAJ42, TRAJ41, TRAJ40, TRAJ39, TRAJ38 DLL3, SELV	Deletion Deletion	
F51	GSN EIF2B4	p.Gln528Pro 2:27584647_TGAG/T	Missense Downstream	WDR63 RP11-448N11.3 UBBP5, LIN28AP2, RP11-342C20.2, MIR4500HG, MIR4500, SLITRK5, LINC00397, RP11-471M2.3, TET1P1, RP11-545P6.2, AL354896.1, RPL29P29, LINC00433, LINC00560, RP11-29P20.1	Deletion Deletion Deletion	
F52	ZCCHC14 SCN1A	16:87529390_T/TCCC c.3429+3845_3429+3846insGTATTT	Upstream Intron	CCT3 MARVELD1, ZFYVE27, SFRP5 RNU6-457P RP11-659F24.1, RP11-526H11.1	Deletion Deletion Deletion Deletion	
F53	CACNA1A CARF WDR12	c.5625+280G>A c.-211+43G>A 2:203777090_G/A	Intron Intron Upstream	AC016907.3	Deletion	
F54	SAMHD1 ICA1L	c.348+745_348+760delAAACAATGTATAGTAT c.866+1447_866+1448insTA	Intron Intron	RP11-600K15.1	Deletion	
F55	SH3PXD2A DARS2 NBEAL1	c.399-166C>A 1:173789430_T/A c.7420-1162dupG	Intron Upstream Intron	MYO9B, CTD-2528A14.5, CTD-3032J10.2	Deletion	
F57	APP JPH3	c.865+10820_865+10821dupGA 16:87631934_GGC/G	Intron Downstream			

Pedigree	Candidate gene	SNV	SNV type	CNV within gene	CNV type	Remarks
F58	PRDM16 EIF2B4	c.38-5052_38-5039delCGGTCTTACACACG 2:27586203_C/CT	Intron Intron	ADCY2 SAV1, RN7SL452P	Deletion Deletion	
F59	PMF1	c.162-5888_162-5883delCCCTGC	Intron	C1orf222 TTC30A RP11-195M16.1	Deletion Deletion Deletion	
F60	PRDM16	c.438+25539_438+25568delTAGG ACAGTTGGGAGAAGGACAGTCAGGAG	Intron			
F62	EVL SAMHD1	c.359-9474C>T c.1746+1359delT	Intron Intron	SLC22A23	Deletion	
F63	COL4A2 PRDM16	c.100-22218_100-22217dupGC c.439-5180_439-5153delACACACT CATCCCAACACAGCCCTCCCA	Intron Intron	SLC3A1, PREPL RP11-83M16.1 HEXA	Deletion Deletion Deletion	
F64	SAMHD1 ATP1A3 CACNA1A BGLAP PMF1	p.Leu493Pro 19:42501943_C/CCTGGG c.978+10925delT 1:156214645_G/GGCTGGT 1:156214645_G/GGCTGGT	Missense Upstream Intron Downstream Downstream	DNAJB8, DNAJB8-AS1 FANCD2P2 C9orf66, DOCK8 MIR548AP CRISPLD2	Deletion Deletion Deletion Deletion Deletion	
F65	NOTCH3 MLC1	p.Arg169Cys p.Val354_Ala355insGlyGlyLeuTrpGly GlyGlyAlaGlyGluVal	Missense Inframe insertion			CADASIL-causing mutation
F67	GFAP LMNB1	17:42994270_G/A 5:126175766_AAGG/A	Upstream Upstream	LARGE	Deletion	
F68	FOXF2 PRDM16	6:1388250_A/ACCCCCCCC c.388-5040_388-5038delGTT	Upstream Intron	ALDH8A1 EFCAB6	Deletion Deletion	
F69	APP KIF18B	c.58-23994_58-23992delTTCinsCTTTTCT TTTCTTTTTTTTTTTTTTTT c.801G>A	Intron Synonymous			

Pedigree	Candidate gene	SNV	SNV type	CNV within gene	CNV type	Remarks
F71	SH3PXD2A	c.399-8662_399-8621delAAATGATGGATGGAT GGATGGCTGGATGAATGATGGATGGAT	Intron	U3, ZIM3	Deletion	
F72	ATP1A2 CACNA1A JPH3 SH3PXD2A DARS2 NBEAL1 NBEAL1	1:160085150_GCT/G c.1669-1930_1669-1926delTTTTG c.1160+1415_1160+1417delAGC c.428-1296_428-1295dupAA c.840+208delG c.4761+2621_4761+2622insATTTTTT c.6015+74T>C	Upstream Intron Intron Intron Intron Intron Intron	CTC-340A15.2 SERHL, RRP7A	Deletion Deletion	
F73	KIF18B	17:43029250_ACCATGT/A	Upstream	CLTA	Deletion	
F74	NOTCH3 RNASEH2A CACNA1A PRDM16	p.Cys76Tyr p.Gln126His c.1987-1448dupT c.438+25676_438+25703delAATGACA GGGAGGAGGACAGTCCCAGAG	Missense Missense Intron Intron	WWOX L3MBTL4	Deletion Deletion	CADASIL-causing mutation
	NBEAL1	c.2420-1059_2420-1058delAT	Intron			
F75				RP11-55L3.1 DOK3	Deletion Deletion	
F76	ATP1A3	19:42502544_C/G	Upstream	PARD3B	Deletion	
F77				HMGN2P18	Deletion	
F78	NOTCH3 ICA1L SCN1A	p.Thr328Ile c.1243+420_1243+430delTCCTGTGTCCA 2:166844424_AT/A	Missense Intron Downstream			
F79	COL4A2	c.181-19325_181-19312delGTGGGT	Intron	AC097506.1, AC009499.1	Deletion	

Pedigree	Candidate gene	SNV	SNV type	CNV within gene	CNV type	Remarks
		CCCGGTGTinsCACC				
	CSF1R	c.1627-2825G>A	Intron			
	ATP1A2	c.2841-812delT	Intron			
F80	NOTCH3	c.118+1244_118+1247delCACA	Intron	MICU1	Deletion	
	CSF1R	5:149493575_A/T	Upstream	RP11-553E24.2	Deletion	
	ITM2B	13:48802664_AGTAT/A	Upstream	CTD-3037G24.4, CTD-3037G24.3, SNX29	Deletion	
	DEGS2	c.83-680_83-679insAG	Intron	ZNF566	Deletion	
	SH3PXD2A	c.604+10390delA	Intron			
	ICA1L	c.985+3728_985+3733delTTTTTA	Intron			
F81				RP4-568F9.6	Deletion	
F82	GFAP	p. Arg29Cys	Missense			
	LMNB1	c.1386+396_1386+398delGAG	Intron			
	MLC1	22:50495579_T/TGGGAGGGGGGGTCGGGGTG	Downstream			
F88	APP	c.225+7943delG	Intron			COL4A1 3’UTR variant segregating with disease – not passing filters
F90	SAMHD1	c.209-2107_209-2094delTTTCTTCTTTTCTTinsC	Intron	SPICE1	Deletion	
				KHDRBS2	Deletion	
				APOBEC3D	Deletion	
F91	ABCC6	p.Gly992Arg	Missense	AC005150.1	Deletion	
	CARF	c.1559-890_1559-889insATATATAT	Intron			
F92	APP	c.57+4594delG	Intron	MSR1, MRPL49P2	Deletion	
	LMNB1	c.360-9147_360-9136delCATTTGTTTGTT	Intron	ZNF816-ZNF321P, ZNF816	Deletion	
F93	HTRA1	c.472+12941C>T	Intron	RP11-78F17.1, COX6CP3	Deletion	

Pedigree	Candidate gene	SNV	SNV type	CNV within gene	CNV type	Remarks
	APP CACNA1A SH3PXD2A GFAP	c.1225-177_1225-174delATTA c.5528+525dupC c.398+9147delC 17:42996750_A/AT	Intron Intron Intron Upstream	PKIB, RP11-95G17.2	Deletion	
F94				RP11-568J23.4	Deletion	
F95	HTRA1 ATP1A2	p.Val175Met 1:160081647_CGG/C	Missense Upstream			
F96	CACNA1A EIF2B3	c.399+970delC 19:13620885_AA/GGGAGGGGG c.295-734_295-713delTATATATATATATAT ATATATA	Intron Upstream Intron	BIRC6	Deletion	
F97	CARF APP PRDM16 SH3PXD2A DARS2 NBEAL1	p.Thr657Ile c.57+12441delA c.439-62964_439-62963delTG c.604+3299delC c.1191+28delT c.52-12093delA	Missense Intron Intron Intron Intron Intron	CTD-2377D24.8	Deletion	
F98	COL3A1 PRDM16	p.Pro137Ser c.38-10111_38-10101delTCCC CTTATAAinsGCCATCCCCCCATCC	Missense Intron	LANCL3	Deletion	
F99	ATP1A2 EIF2B3	1:160083751_G/GT c.294+8349C>T	Upstream Intron			
F100	APP JPH3	c.57+22922A>G c.1161-4888dupT	Intron Intron			
F101			Missense	POT1-AS1, EEF1GP1 CTD-2135J3.3	Deletion Deletion	
F102	DARS2 SH3PXD2A	c.128-614_128-613delAA c.399-3755A>G	Intron Intron	TBC1D7	Deletion	

Pedigree	Candidate gene	SNV	SNV type	CNV within gene	CNV type	Remarks
F103	COL4A1 RNASEH2B	p.Trp1578Cys p.Val41LeufsTer23	Missense Frameshift	PRTFDC1 ADAMTS18 RNA5SP443 PTH2,GFY	Deletion Deletion Deletion Deletion	
F104	COL4A2 ATP1A3 PRDM16 DARS2	c.100-1916_100-1877delAGGGCTG TACTAGCTCACCTGTGTGGGGGGAGGGCAGTA CinsGCGGGCGGGA 19:42501941_C/CCCCAGCGG c.38-23264_38-23263insACGGGCTAGAG GGCGCTGGGAGGGGGGAGGTGCGCAGAA GCCCCCGGC c.1129-575delA	Intron Upstream Intron Intron	RP11-165H23.1	Deletion	
F105	GLA GSN ZCCHC14 EIF2B3	X:100649780_TTTTTTTT/ATATTTA c.504+652C>T c.284-9451delT c.1203-7409_1203-7406delAAGAinsTTTTT TTTTTTTAAGT	Downstream Intron Intron Intron	AL031176.1	Deletion	
F107	JPH3	c.383-9601delC	Intron	FECHP1 ATG10, AC114969.1	Deletion Deletion	
F108	EIF2B3	c.1203-7409delAinsTTTTTTTTTTTT	Intron	GATS	Deletion	
F109	CECR1 FAM126A EIF2B3	c.973-422_973-421insATTTTTT c.-29+8308delG c.149-158_149-157insAAAAAAA	Intron Intron Intron			
F110	HTRA1 COL4A2 PRDM16 PRDM16 EIF2B2	c.777+4939G>A p.Pro41Ser c.38-31675delC c.38-4986_38-4985insCCTTACA CACGCGGTCTTACACGCGG c.1008C>T	Intron Missense Intron Intron Synonymous	GORAB RP11-510I6.3 RP11-152P17.2, Y_RNA AP002387.1 AHNAK2	Deletion Deletion Deletion Deletion Deletion	

Pedigree	Candidate gene	SNV	SNV type	CNV within gene	CNV type	Remarks
	LMNB1	c.360-5990_360-5989dupCG	Intron	PLIN4 CTD-3020H12.3	Deletion Deletion	
F111	NOTCH3	19:15267550_T/G	Downstream	RP11-653B10.1	Deletion	
F113	LMNB1 ATP1A3 ATP1A3 PRDM16 RNASEH2A	p.Glu559Ala c.2133+278_2133+279dupGG c.7-2227_7-2224delAGAG c.37+41258_37+41296delTGTGGGGACAC AGTGTCTGTGGCTCCCCCATGTCCTGGG 19:12912601_G/A	Missense Intron Intron Intron Upstream			
F115	ATP1A3	19:42501935_T/TCCCCCCC	Upstream	NTNG1 RP11-428P16.2 RP11-174M13.1 ZNF431	Deletion Deletion Deletion Deletion	
F116	ZNF474	c.-212-2270delG	Intron	AC092155.4 AC083884.8, STAG3L2	Deletion Deletion	
F117	CACNA1A CSF1R EVL	c.979-10759_979-10726delAATATATATATATATATATATAT ATATATATATA c.890-822_890-821insCA c.488-59delA	Intron Intron Intron	C16orf45 CNTLN	Deletion Deletion	
F118	LMNB1 ICA1L	c.-349A>G c.911-2725delA	5'UTR Intron	RP11-91I11.1 DCTPP1	Deletion Deletion	
F120	ITM2B EVL	c.117+1548_117+1549delTA c.181-4540delG	Intron Intron	GBP7,GBP4	Deletion	
F121	NOTCH3 CACNA1A	c.4736+1035_4736+1036insGGGGGGG GGGGG 19:13617511_C/T	Intron Upstream	RP11-193H5.5, MTND6P15, MTND5P18	Deletion	

Pedigree	Candidate gene	SNV	SNV type	CNV within gene	CNV type	Remarks
F122	HTRA1	p.Arg370Ter	Nonsense			
	MLC1	c.894+580_894+603delCCTGATGGGGGGATGGAGTCACTG	Intron			
F123	COL4A2	c.180+7416_180+7417delCT	Intron	AGO1	Deletion	
	PRDM16	c.388-3993_388-3992delAG	Intron	RP11-13E5.2, DCAF12L2	Deletion	
	LMNB1	c.359+9667dupA	Intron			
F124	CACNA1A	c.978+6492dupA	Intron	GNG12-AS1, WLS	Deletion	
	SCN1A	c.2416-314_2416-313dupGA	Intron	RP11-703H8.9	Deletion	
				SMG6, AC090617.1, AC130689.5, AL450226.2, SRR, HNRNPA1P16, TSR1	Deletion	
F125	PRDM16	c.439-68874_439-68873insGGGGGGG	Intron	FAM135B	Deletion	
F126	COL4A1	p. Gly1369ArgfsTer33	Frameshift	RP11-124B13.1	Deletion	
	COL4A2	c.181-30810delG	Intron			
	ADAR					
	CSF1R	5:149493458_CAGAT/C	Upstream			
	CACNA1A	c.4950+621_4950+632delACCCTCTCCTTG	Intron			
	ADAR	1:154583318_GACAAAATTG/AAAAAAAAAAAAAAAT	Upstream			
	NBEAL1	c.6508+418_6508+419insTTTT	Intron			
F128				SMG1	Deletion	
F132	APP	c.57+12572A>C	Intron			
	EARS2	c.486-2202_486-2199dupAATG	Intron			
	LMNB1	c.360-3828_360-3827insATATATAT	Intron			
F133	APP	c.226-9792_226-9790delTAC	Intron			
	JPH3	16:87734234_GGAGGGAGAACGCTGGAGGGCA/G	Downstream			
F134	APP	c.662+4475dupT	Intron			

Pedigree	Candidate gene	SNV	SNV type	CNV within gene	CNV type	Remarks
	CTSA DARS2 SH3PXD2A	20:44516873_T/A c.664-415_664-414delAA c.605-8032_605-8015delAGCCAGCCAGAGAACAGC	Upstream Intron Intron			
F137	CACNA1A	c.978+509delC	Intron			
F139	ATP1A3 PRDM16 SH3PXD2A	19:42501343_AG/A 1:3358271_C/GCCGGTTCCT c.154-10399_154-10398delCT	Upstream Downstream Intron			
F141	FAM126A PRDM16 CARF	c.744-1447dupG c.439-62996_439-62982delACTGTGTGTGTGTGC c.1495-1182_1495-1174delACATACATAinsCCATCCATCCATC	Intron Intron Intron			
F142	APP SH3PXD2A	c.1687+9174_1687+9181dupCTACCATT c.230-14080_230-14060delCTGGCGCTGGCTGCTGGCAGA	Intron Intron			

Table 5-10*: Non-coding SNVs passing filters identified in 118 probands. Variants present in patients with a known pathogenic mutation in another gene, or variants present in relatives but absent in the proband in a pedigree, or variants not segregating with disease within the pedigree are excluded.*

	Gene	Variant identified	Variant type
Monogenic SVD	CECR1	c.753+2659delT	Intron
		c.973-422_973-421insATTTTTT	Intron
	CTSA	20:44516873_T/A	Upstream
	FOXC1	None	
	GLA	X:100649780_TTTTTTTT/ATATTTA	Downstream
		X:100648388_AAAG/A	Downstream
	HTRA1	c.472+12941C>T	Intron
		c.777+4939G>A	Intron
		c.777+6599delG	Intron
	NOTCH3	19:15267550_T/G	Downstream
		c.4736+1035_4736+1036insGGGGGGGGGGG	Intron
		c.118+1244_118+1247delCACA	Intron
	COL4A1	c.144+1918_144+1919insAG	Intron
		c.144+6701_144+6705delCTTTA	Intron
		c.1897+559_1897+561delCCT	Intron
		c.145-5665_145-5634delCACATATATATATATATACATATATATATATAinsTATT	Intron
	COL4A2	c.100-1916_100-1877delAGGGCTGTACTAGCTCACCTGTGTGGGGGGAGGGCAGTACinsGCGGGCGGGA	Intron
		c.100-22218_100-22217dupGC	Intron
		c.1596+636_1596+637dupGG	Intron
		c.180+7416_180+7417delCT	Intron
		c.181-19325_181-19312delGTGGGTCCCGGTGTinsCACC	Intron
		c.181-30810delG	Intron
		c.181-32518_181-32515delGTCA	Intron
		c.2425+1717_2425+1725delCATACACAC	Intron
		c.99+16231_99+16232delTG	Intron

	Gene	Variant identified	Variant type
	SNORD118	17:8076200_GA/G	Downstream
	TREX1	3:48512824_TACAAAAAT/C	Downstream
FHM	SCN1A	2:166844424_AT/A	Upstream
		c.3429+3845_3429+3846insGTATTT	Intron
		c.2416-314_2416-313dupGA	Intron
	ATP1A2	1:160083751_G/GT	Upstream
		1:160084457_AG/A	Upstream
		1:160085150_GCT/G	Upstream
		1:160085222_G/GCCACATGGT	Upstream
		c.1828-236_1828-231delCACACA	Intron
		c.2841-812delT	Intron
		1:160118195_TAGAG/T	Downstream
	ATP1A3	19:42501940_C/CCCCCCACG	Upstream
		19:42501941_C/CCCCAGCGG	Upstream
		c.2133+278_2133+279dupGG	Intron
		c.7-2227_7-2224delAGAG	Intron
		19:42501935_T/TCCCCCCC	Upstream
		19:42501343_AG/A	Upstream
		19:42502584_G/A	Upstream
		19:42501943_C/CCTGGG	Upstream
		19:42502544_C/G	Upstream
	CACNA1A	c.6780+227_6780+228insCCCCCC	Intron
		c.5625+1155_5625+1156insCGCC	Intron
		c.5625+280G>A	Intron
		c.5528+525dupC	Intron
		c.4950+621_4950+632delACCCTCTCCTTG	Intron
		c.4591-732_4591-730delAAG	Intron
		c.2172+998_2172+1001delTTTC	Intron
		c.1987-1448dupT	Intron
		c.1669-1930_1669-1926delTTTTG	Intron
		c.1346-1598_1346-1594dupAAGAT	Intron
		c.1255+465dupT	Intron

	Gene	Variant identified	Variant type
		c.979-10759_979-10726delAATATATATATATATATATATATATATATATATA	Intron
		c.978+10925delT	Intron
		c.978+6492dupA	Intron
		c.978+6492dupA	Intron
		c.978+509delC	Intron
		c.540-5082_540-5081delTA	Intron
		c.539+22891_539+22912delACCCTACTCCACACTCCCCACT insCCCCCCCC	Intron
	PRRT2	None	
Amyloidosis	APP	None	
	ITM2B	c.117+1548_117+1549delTA	Intron
		13:48802664_AGTAT/A	Upstream
	GSN	9:124098555_GC/G	Downstream
		9:124099535_C/G	Downstream
		c.504+652C>T	Intron
	ADAR	1:154583318_GACAAAATTG/AAAAAAAAAAAAAAAT	Upstream
Leukodystrophy	CSF1R	c.1627-2825G>A	Intron
		c.890-822_890-821insCA	Intron
		5:149493458_CAGAT/C	Upstream
		c.49+1988_49+1989insCTTCCTTCTTTCTTTCTTTC	Intron
		5:149493575_A/T	Upstream
	DARS2	c.664-415_664-414delAA	Intron
		c.1129-575delA	Intron
		c.840+208delG	Intron
		c.1129-2107delC	Intron
		c.1191+28delT	Intron
		1:173789430_T/A	Upstream
	EARS2	c.1353-1591delT	Intron
		c.486-2202_486-2199dupAATG	Intron
		c.295+1153G>C	Intron
	EIF2B1	None	

Gene	Variant identified	Variant type
EIF2B2	c.434-46_434-32delTTTTTTTTTTTTTTTinsATATATATATATA	Synonymous
	14:75475843_C/T	Intron
EIF2B3	c.1203-7409_1203-7406delAAGAinsTTTTTTTTTTTTAAGT	Intron
	c.1203-7409delAinsTTTTTTTTTTTT	Intron
	c.295-734_295-713delTATATATATATATATATATATA	Intron
	c.294+8349C>T	Intron
	c.149-158_149-157insAAAAAAA	Intron
EIF2B4	2:27584647_TGAG/T	Downstream
	2:27586203_C/CT	Downstream
FAM126A	c.744-1447dupG	Intron
	c.743+407A>C	Intron
	c.626+3795G>A	Intron
	c.-28-8949_-28-8947delTTA	Intron
	c.-29+8308delG	Intron
GFAP	17:42994270_G/A	Upstream
	17:42996750_A/AT	Upstream
HEPACAM	None	
LMNB1	c.-349A>G	5’UTR
	c.359+9667dupA	Intron
	c.359+13169_359+13172dupTAAA	Intron
	c.360-9147_360-9136delCATTTGTTTGTT	Intron
	c.360-5990_360-5989dupCG	Intron
	c.360-3828_360-3827insATATATAT	Intron
	c.1386+396_1386+398delGAG	Intron
	c.1612-1870dupT	Intron
	5:126175766_AAGG/A	Downstream
MLC1	c.894+580_894+603delCCTGATGGGGGGATGGAGTCACTG	Intron
	22:50495579_T/TGGGAGGGGGGTCGGGGTG	Downstream
RNASEH2A	19:12912601_G/A	Upstream
RNASEH2B	None	
SAMHD1	c.1746+1359delT	Intron

	Gene	Variant identified	Variant type
		c.1062+550delCinsTTTTTTTTTTTT	Intron
		c.696+584_696+585dupTT	Intron
		c.348+745_348+760delAAACAATGTATAGTAT	Intron
		c.209-2090_209-2079delTTTTTTTCTTTT	Intron
		c.209-2107_209-2094delTTTCTTCTTTTCTTinsC	Intron
	TUBB4A	c.*128delC	3'UTR
		c.278-2010delT	Intron
CTD	COL3A1	2:189879981_A/AT	Downstream
	ABCC6	c.3634-43_3634-39delAGTGG	Intron
		c.601-361delC	Intron
Sporadic	CARF	c.1495-1182_1495-1174delACATACATAinsCCATCCATCCATC	Intron
		c.1494+251dupA	Intron
		c.-211+43G>A	Intron
		c.1559-890_1559-889insATATATAT	Intron
	BLAP	1:156214645_G/GGCTGGT	Downstream
	DEGS2	c.826-752_826-751insGGGGC	Intron
		c.83-680_83-679insAG	Intron
	EFEMP1	c.517+7277C>T	Intron
	EVL	c.181-4540delG	Intron
		c.359-9474C>T	Intron
		c.488-59delA	Intron
		14:100613995_G/GGCCCC	Downstream
	FOXC2	16:86600205_A/G	Downstream
	FOXF1	16:86541560_T/TG	Upstream
	FOXF2	6:1388250_A/ACCCCCCCC	Upstream
	FOXJ1	None	
	FOXL1		
	FOXQ1		
	HAAO		
	ICA1L	c.1243+420_1243+430delTCCTGTGTCCA	Intron
		c.985+3728_985+3733delTTTTTA	Intron

Gene	Variant identified	Variant type
	c.911-2725delA	Intron
	c.866+1447_866+1448insTA	Intron
	c.-8+10250_-8+10251insTTT	Intron
JPH3	c.1161-4888dupT	Intron
	c.383-9601delC	Intron
	16:87734234_GGAGGGAGAACGCTGGAGGGCA/G	Downstream
	16:87631934_GGC/G	Upstream
	c.1160+1415_1160+1417delAGC	Intron
KIF18B	c.801G>A	Synonymous
	c.-15+4779_-15+4781delTGT	Intron
	17:43029250_ACCATGT/A	Upstream
	17:43029291_CTCTC/ATATATATATATA	Upstream
LOX	c.1036-95A>G	Intron
MIR216A	2:56212347_C/T	Downstream
MIR217	2:56216414_G/A	Upstream
MIR216B	2:56229175_C/T	Upstream
NBEAL1	c.51+12044_51+12045delTT	Intron
	c.52-12093delA	Intron
	c.516-1069_516-1067delACA	Intron
	c.1098+4385dupT	Intron
	c.2420-1059_2420-1058delAT	Intron
	c.4761+2621_4761+2622insATTTTTT	Intron
	c.6015+74T>C	Intron
	c.6508+413_6508+417dupGCCAG	Intron
	c.6508+418_6508+419insTTTT	Intron
	c.7420-1162dupG	Intron
OXER1	None	
PMF1	c.162-5888_162-5883delCCCTGC	Intron
	1:156214645_G/GGCTGGT	Downstream
PRDM16	c.37+41258_37+41296delTGTGGGGACACAGTGTCTGTGG CTCCCCCATGTCCTGGG	Intron
	c.38-31675delC	Intron

Gene	Variant identified	Variant type
	c.38-23264_38-23263insACGGGCTAGAGGGCGCTGGGAGGGGGGAGGTGCGCAGAAGCCCCCGGC	Intron
	c.38-10111_38-10101delTCCCCTTATAAinsGCCATCCCCCCATCC	Intron
	c.38-5052_38-5039delCGGTCTTACqACACG	Intron
	c.38-4986_38-4985insCCTTACACACGCGGTCTTACACGCGG	Intron
	c.38-4431_38-4419delTCTTACACACGGTinsC	Intron
	c.388-5040_388-5038delGTT	Intron
	c.388-3993_388-3992delAG	Intron
	c.438+25539_438+25568delTAGGACAGTTGGGAGAAGGACAGTCAGGAG	Intron
	c.438+25676_438+25703delAATGACAGGGAGGAGGACAGTCCCAGAG	Intron
SH3PXD2A	c.*5178_*5187dupGTGTCTGTGT	3’ UTR
	c.1225-2233delAinsGTAAACTTTTTTTTTTTTTTTTTT	Intron
	c.836+244C>T	Intron
	c.605-8032_605-8015delAGCCAGCCAGAGAACAGC	Intron
	c.604+10390delA	Intron
	c.604+3299delC	Intron
	c.473-1228_473-1227insCG	Intron
	c.473-1239_473-1231delGCCCCCCCC	Intron
	c.428-1296_428-1295dupAA	Intron
	c.399-166C>A	Intron
	c.399-3755A>G	Intron
	c.399-8662_399-8621delAAATGATGGATGGATGGATGGCTGGATGAATGATGGATGGAT	Intron
	c.398+9147delC	Intron
	c.230-14080_230-14060delCTGGCGCTGGCTGCTGGCAGA	Intron
	c.154-10399_154-10398delCT	Intron
	c.72+26561G>T	Intron

Gene	Variant identified	Variant type
WDR12	c.741+866_741+867dupAA	Intron
	2:203777090_G/A	Upstream
ZCCHC14	c.*3566T>C	3’ UTR
	c.284-9451delT	Intron
ZNF474	c.-213+1131_-213+1134delTCAT	Intron
	c.-212-2270delG	Intron

Chapter 6: The frequency of known mutations in the general population

Outline

6.1 Introduction

6.2 Methods

- 6.2.1 Known disease-causing variants
- 6.2.2 Presence of known disease-causing variants in controls

6.3 Results and discussion

- 6.3.1 NOTCH3
- 6.3.2 HTRA1
- 6.3.3 COL4A1
- 6.3.4 COL4A2

6.4 Conclusions

- 6.4.1 Milder forms of familial SVD
- 6.4.2 Evaluating the strength of evidence for pathogenicity: segregation in reported cases
- 6.4.3 In vitro functional analyses alone are insufficient to demonstrate pathogenicity
- 6.4.4 Moving forward

6.5 Appendix

List of tables

Table 6-1 Search terms used for the systematic review of monogenic SVD genes.

Table 6-2 Number of variants reported as disease-causing in NOTCH3, HTRA1, COL4A1 and COL4A2

Table 6-3 Frequency of cysteine-altering variants in exons 2 to 24 of NOTCH3 identified in unrelated controls in BRIDGE.

Table 6-4 Frequency of cysteine-sparing NOTCH3 variants previously associated with familial SVD in control datasets.

Table 6-5 Frequency of CARASIL and autosomal dominant SVD-causing mutations in the HTRA1 reported in literature, in unrelated BRIDGE controls

Table 6-6 Frequency of COL4A1 variants associated with familial SVD in unrelated BRIDGE controls

Table 6-7 Frequency of COL4A2 variants associated with familial SVD in unrelated BRIDGE controls.

(Appendix)

Table 6-8 Cysteine-altering NOTCH3 variants reported as disease-causing, or identified in patients with CADASIL presenting to the UK National Referral Clinic for CADASIL

Table 6-9 Cysteine-sparing NOTCH3 variants previously associated with CADASIL

Table 6-10 HTRA1 variants reported in CARASIL or autosomal dominant SVD

Table 6-11 COL4A1 variants reported in familial disease

Table 6-12 COL4A2 variants reported in familial disease

6. The frequency of known mutations in the general population

6.1 Introduction

High throughput sequencing (HTS) techniques, as described in Chapter 2, have significantly contributed to our understanding of the genetic basis of rare Mendelian diseases.[184] Earlier genetic studies relied on the genotyping of extensive pedigrees to identify disease-associated loci by linkage analysis. This, coupled with Sanger sequencing led to the identification of novel variants and disease-associated genes, although not always with success. More recently, the use of HTS techniques sometimes coupled with linkage analyses have led to an exponential growth in gene and variant discovery in the diagnosis of rare diseases.[184,313]

A pitfall of the use of HTS lies in the fact that while it allows the screening of multiple genes simultaneously, providing a high yield of variants in each patient, it may not necessarily lead to any clinical benefit in the search for a genetic diagnosis. The discovery of novel variants with no prior association with disease or functional analyses is of limited utility as these are classed as 'variants of uncertain clinical significance (VUCS)'. These variants are understandably either left out of clinical genetics reports, or are returned to the clinician with a note of caution. This means that the endeavour often raises more questions than answers, and may not lead to a genetic diagnosis and the provision of appropriate genetic counselling.

One strategy used to sieve out potentially disease-causing variants is an arbitrary threshold for minor allele frequency (MAF) of variants in the general population. Highly penetrant heterozygous mutations known to cause rare Mendelian diseases are expected to be rare in the population, in line with the relatively low prevalence of these diseases. For example, the population prevalence of CADASIL is estimated at 2 to 4 per 100,000 in the UK.[40,41] An example is shown in Chapters 4 and 5 where we have used minor allele frequency as a filter to identify potentially disease-causing variants. The assumption in such studies is that highly penetrant mutations associated with rare Mendelian diseases are likely to also be rare in the population, and that a single mutation alone carries sufficient impact to account for all features of the disease.

However, recent evidence suggests higher-than-expected frequencies of disease causing mutations in control populations. For example, a recent study in the Exome Aggregation Consortium (ExAC) database showed that CADASIL-causing variants were present in up to 3.4 per 1000 individuals in the control dataset, and suggested that some of these variants may result in a milder, later-onset phenotype.[59] A study of known mutations in epilepsy-associated genes and their frequency in ExAC has also shown that some of the known disease-causing variants may be present in a much higher frequency than would be expected for the disease, particularly if there is insufficient segregation data for the variant.[314] While control databases such as ExAC are by no means repositories of disease-free, well-phenotyped 'healthy' individuals, these frequencies were nevertheless beyond the estimated population prevalence of the disease.

Understanding the frequency of VUCS in the general population is valuable in aiding our interpretation of these variants in the clinical context. Highly penetrant variants which are thought to be the causative mutations in our cohort of families with suspected familial SVD, are theoretically unlikely to be present in high frequencies in the general population since the disease is known to be rare. In this study, we aim to provide an estimate of the allele frequency of known disease-causing mutations in known familial SVD-associated genes in the BRIDGE whole genome sequencing population, as well as in two control databases, the Exome Aggregation Consortium (ExAC) and genome aggregation databases (gnomAD).

6.2 Methods

6.2.1 Known disease-causing variants

A literature search for known genes related to familial SVD was performed on PubMed for publications dating from 1993 to April 2017. A comprehensive search strategy was developed using the search terms detailed in **Table 6-1**. Key words were combined, such as 'cerebral small vessel disease' and 'genetics', using 'AND'. Known SVD-associated genes (NOTCH3, HTRA1, COL4A1, COL4A2) were included in the search terms. All abstracts in English were reviewed, and the full texts of relevant systematic reviews, case reports, prevalence or observational and candidate gene studies were reviewed to create a database of known disease-causing variants.

Table 6-1: *Search terms used for the systematic review of monogenic SVD genes.*

Disease	Features on imaging	Monogenic diseases or genes
• Cerebral small vessel disease • Binswanger disease • Cerebral microangiopathy • Lacunar stroke • Lacunar infarct • Subcortical stroke • Vascular dementia • Vascular cognitive impairment • Subcortical dementia	• Leukoaraiosis • Leukoencephalopathy • White matter disease	• CADASIL, NOTCH3 • CARASIL, HTRA1 • COL4A1 • COL4A2

Published disease-causing variants in each gene were identified from peer-reviewed literature and mutation databases. These include the Leiden Open Variation Database (LOVD, http://www.lovd.nl/3.0/home), and the Human Gene Mutation Database (HGMD, http://www.hgmd.cf.ac.uk). Although these variants have previously been reported in disease it is important to note that not all have the same level of evidence for pathogenicity. Some variants may have only been identified in single cases, without evidence of segregation with disease in the pedigree, or in-vitro evidence to illustrate the impact of the mutation on protein function.

NOTCH3 mutations identified in patients attending the UK National CADASIL Clinic, but had not previously been reported in literature, were also included in this analysis. The total number of variants included is summarised in **Table 6-2**, and the full list of variants identified is provided in the appendix.

Cysteine-altering variants affecting the epidermal growth factor-like repeat (EGFR) regions, encoded by exons 2 to 24 of the NOTCH3 gene, are known to cause CADASIL. These variants are neither gain- or loss-of-function variants, but are instead thought to lead to the abnormal cleavage of the NOTCH3 ectodomain, setting off a cascade of aggregation of proteins around the smooth muscle cells of the small vessels,[119] as described in Chapter 1.

Cysteine-sparing variants, though previously reported as being associated with CADASIL,[262,315,316] are no longer thought to be disease-causing in majority of cases due to the lack of sufficient segregation evidence or absence of full analysis of all 23 exons encoding the

EGFR regions.[317–320] However, a few of these variants have been identified in families with complete screening of all 23 variants, segregation of the variant in the pedigree or have granular osmiophilic material, and as a result there has been a recent review suggesting the possible pathogenicity of at least a few of these variants.[321] For completeness, all cysteine-sparing variants have been summarised in a separate table in the appendix, and have also been queried in control databases as part of this study.

Table 6-2: *Number of variants reported as disease-causing in NOTCH3, HTRA1, COL4A1 and COL4A2*

Gene	Total no. of mutations reported	No. of mutation reported in heterozygous cases	No. of mutations reported in homozygous cases	Table in appendix
NOTCH3*	278	273	8	Table 6-8
HTRA1	33	20	13	Table 6-10
COL4A1	80	80	0	Table 6-11
COL4A2	10	10	0	Table 6-12

*Only cysteine-changing variants are included in this table.

6.2.2 Presence of known disease-causing variants in controls

Patients recruited to the BRIDGE study under rare diseases other than SVD served as presumed SVD-free controls in this study. These individuals were recruited as either affected or disease-free individuals in different rare disease categories, as described in Chapters 2 and 5. Whole genome sequencing was performed in 7348 unrelated non-SVD individuals. The MAF of known disease-causing mutations as listed in the appendix was calculated in this set of controls. Data on the age, sex, disease category and disease status for patients carrying SVD-associated disease-causing variants was collected.

The frequencies of known disease-causing variants were also queried in two publicly accessible genomic databases: the ExAC v0.3 database, and genome Aggregation Database (gnomAD). The ExAC database is a repository of exome sequencing data from 60706 individuals, curated by the Broad Institute (http://exac.broadinstitute.org/downloads). GnomAD is a database of 123,136 whole exome sequences and 15,496 whole genome sequences, a catalogue developing from ExAC, also curated by the Broad Institute

(http://gnomad.broadinstitute.org/). Combined Annotation Dependent Depletion (CADD) scores, computed using an in silico predictive tool which integrates multiple annotations to provide an estimate of deleteriousness (http://cadd.gs.washington.edu), were also queried.

The term 'diagnostic yield' is used here to describe the proportion of unrelated individuals carrying a variant or group of variants, while the term 'allele frequency' is used here to describe the proportion of alleles sequenced in controls (assuming two alleles sequenced per individual).

6.3 Results and Discussion

Previously reported disease-causing mutations in HTRA1, COL4A1 and COL4A2, and cysteine-changing mutations in NOTCH3 were identified in 214 out of 7348 unrelated controls. All except one mutation (in COL4A2) were identified in the heterozygous state. None of the controls carried more than one mutation.

All mutations except one (in COL4A1) had CADD scores above 15, a threshold used in Chapter 5 as a filter for rare variants in individuals with familial cerebral SVD. Cysteine-sparing variants in NOTCH3 have not been included in the above summary.

6.3.1 NOTCH3

Nine cysteine-altering NOTCH3 variants in exons 2 to 24 of the gene were identified in 19 unrelated controls (diagnostic yield of 0.26%). The CADD scores for all nine variants ranged from 22.8 to 30.0, and six of these variants were novel. Each variant was found in 1 to 6 unrelated controls, with these controls being recruited under various rare diseases, with a broad age range of 5 to 80.

Although the age at onset of symptoms in CADASIL can vary greatly,[43] MRI changes tend to be present by around the age of 35 in CADASIL mutation carriers, as shown in earlier studies in small cohorts of pre-symptomatic and symptomatic individuals.[217,322] Excluding individuals where the age is not known or below 35, there remain 9 unrelated controls carrying cysteine-altering NOTCH3 variants which are presumed to be pathogenic. These results are summarised in Table 6-3.

Electron microscopy of a kidney biopsy from a ten-year-old p.Cys522Ser variant carrier with membranoproliferative glomerulonephritis was performed, and this showed characteristic subendothelial dense deposits typical of membranoproliferative glomerulonephritis. Granular osmiophilic material (GOM, see Chapter 1) was not specifically looked for, and images were not available for re-evaluation. Evidence of GOM has been documented in patients with CADASIL as young as 20 years of age. [323]

Table 6-3: *Frequency of cysteine-altering variants in exons 2 to 24 of NOTCH3 identified in unrelated controls in BRIDGE. All variants identified were in a heterozygous state.*

Variant	Exon	CADD score	No. of unrelated controls	Allele frequency in controls	Sex of controls	Ages of controls at recruitment	Disease phenotype	Frequency in ExAC	Frequency in gnomAD	Reference	Remarks
p.Phe270Cys	6 of 33	25.6	1	6.8×10^{-5}	F	65	Multiple primary malignant tumours	-	-	Novel	
p.Cys522Ser	10 of 33	28.6	1	6.8×10^{-5}	M	10	Membranoproliferative glomerulonephritis	-	-	Novel	GOM not described on EM of kidney biopsy
p.Gly667Cys	10 of 33	26.0	1	6.8×10^{-5}	F	68.7	Bleeding disorder	-	4.1×10^{-6}	LOVD	
p.Arg728Cys	14 of 33	26.4	1	6.8×10^{-5}	F	65.7	Pulmonary arterial hypertension	-	8.4×10^{-6}	[57,324]	
p.Arg785Cys	15 of 33	25.9	1	6.8×10^{-5}	M	Not known	Unaffected, recruited under Leber Hereditary Optic Neuropathy	-	3.2×10^{-5}	Novel	
p.Tyr916Cys	15 of 33	22.8	1	6.8×10^{-5}	F	45	Bleeding disorder	-	-	Novel	
p.Cys1222Gly		25.6	5	3.4×10^{-4}	F	44	Bleeding disorder	7.4×10^{-5}	1.1×10^{-4}	[256]	

Variant	Exon	CADD score	No. of unrelated controls	Allele frequency in controls	Sex of controls	Ages of controls at recruitment	Disease phenotype	Frequency in ExAC	Frequency in gnomAD	Reference	Remarks
	21 of 33										
					M	Not known	Not known – Genomics England Pilot				
					F	80	Multiple primary malignant tumours				
					M	46	Primary immune disorder				
					F	4	Paediatric neurology				
p.Arg1231Cys	22 of 33	30	6	4.1×10^{-4}	F	36	Pulmonary arterial hypertension	9.9×10^{-4}	7.8×10^{-4}	[57,58,176,245,325,326] Internal database	
					M	Not known	Pulmonary arterial hypertension				
					M	36	Primary immune disorder				
					M	33	Unaffected, recruited under primary immune disorder				Related
					F	5					
					F	12	Rod-cone dystrophy				
					M	10	Paediatric neurology				
p.Cys1324Trp	24 of 33	24.7	1	6.8×10^{-5}	M	26	Factor VII deficiency	-	-	Novel	

GOM: Granular osmiophilic material. EM: electron microscopy

The two most commonly occurring variants were p.Arg1231Cys, with an allele frequency of 4.1×10^{-4}, and p.Cys1222Gly, with a frequency of 3.4×10^{-4}. Both of these variants have previously been associated with patients suspected of having CADASIL.[43,57,58,245,256,325,326] These variants are also relatively common in ExAC and gnomAD, with p.Arg1231Cys occurring in nearly 1 in 1000 individuals. Although a definite genotype-phenotype correlation has not been proven in CADASIL,[43,58,258] cysteine-altering variants in epidermal growth factor-repeat like regions (EGFr) outside of the first six repeats have recently been suggested to be associated with a milder form of disease.[59] As p.Arg1231Cys and p.Cys1222Gly are found in the 31st EGFr repeat, the presence of these variants in a significant number of controls who may still be pre-symptomatic serves to support this hypothesis. p.Cys1222Gly was also a variant that did not pass stringent filtering in our study on sequencing in presumed-sporadic, early-onset SVD cases (Chapter 4).

Cysteine-sparing variants previously associated with CADASIL were also queried in the three control datasets. These variants are not thought to be pathogenic, and have not been included in the total of 214 unrelated controls found to carry disease-causing variants. These variants were found in relatively higher frequencies in this study. Of 16 reported variants, 5 were identified in 49 unrelated controls across the disease phenotypes, giving an overall diagnostic yield of 0.67%. The most commonly occurring variant was p.His170Arg, found in 35 controls, giving an allele frequency of 0.24%. This is comparable to the frequency of the same variant in ExAC (0.18%) and gnomAD (0.17%). Our findings are summarised in **Table 6-4**.

Table 6-4: *Frequency of cysteine-sparing NOTCH3 variants previously associated with familial SVD in control datasets. These variants are no longer thought to be disease causing.*

Variant	Exon	CADD Score	No. of unrelated controls	Frequency in controls	Sex of controls	Ages of controls at recruitment	Disease phenotype	ExAC	gnomAD	Ref	Remarks
p.Arg61Trp	2 of 33	32	3	2.1 x 10^{-4}	2 F 1M	Unknown	Not known – Genomics England pilot (n=2) Pulmonary arterial hypertension (n=1)	8.3 x 10^{-5}	1.4 x 10^{-4}	[327]	
p.His170Arg	4 of 33	23.8	35	0.0024	20 F 15 M	10 unknown Range 13.3 - 85.0 Mean ± SD 49.1 ± 20.9	Bleeding and platelet disorder (n=5) Hypertrophic cardiomyopathy (n=3) Leber's hereditary optic neuropathy (n=1) Multiple primary malignant tumours (n=4) Neuropathic pain disorder (n=1) Pulmonary arterial hypertension (n=6) Primary immune disorder (n=6) Primary membranoproliferative glomerulonephritis (n=1) Paediatric neurology (n=4) Not known – Genomics England pilot (n=6)	0.0018	0.0017	[263]	
p.Thr577Ala	11 of 33	13.9	2	1.4 x 10^{-4}	2 F	40.0 35.0	Primary immune disorder (n=2)	1.7 x 10^{-5}	2.2 x 10^{-5}	[328]	
p.Ser978Arg	18 of 33	13	7	4.3 x 10^{-4}	3 F 4 M	Unknown	Multiple primary malignant tumours (n=3) Stem cell and myeloid disorders (n=1) Not known – Genomics England pilot (n=3)	2.9 x 10^{-4}	2.4 x 10^{-4}	[328]	

Variant	Exon	CADD Score	No. of unrelated controls	Frequency in controls	Sex of controls	Ages of controls at recruitment	Disease phenotype	ExAC	gnomAD	Ref	Remarks
p.Gly1347Arg	24 of 33	25	2	1.4×10^{-4}	1 F 1 M	10.0 3.0	Paediatric neurology (n=2)	-	1.7×10^{-4}	[329]	

6.3.2 HTRA1

Two out of 33 reported HTRA1 mutations were identified in nine controls, giving an overall diagnostic yield of 0.12%. Both variants have CADD scores above 30. These findings are summarised in **Table 6-5**.

p.Gln151Lys, which was previously described only in autosomal dominant SVD, was identified in 8 unrelated individuals, although the ages of majority of these controls were not known at the time of writing. This variant is also relatively common in gnomAD, with a frequency of 1.6 x 10^{-4}. p.Gln151Lys is located in the Kazal-like domain of the HTRA1 protein, and its impact on the function of the HTRA1 enzyme is not known. This variant was previously reported in a singleton with ataxic gait, dysarthria, dysphagia and cognitive decline from the age of 59. This patient also had two affected relatives, and his brain MRI showed white matter hyperintensities, microbleeds and diffuse atrophy. Genetic analysis in the relatives, and functional assays of this mutation were not performed.[330]

The nonsense mutation p.Arg302Ter, identified in only 1 control recruited at age 22, has previously been described in both CARASIL and autosomal dominant disease. The evidence for the pathogenicity of the p.Arg302Ter variant is more conclusive than for p.Gln151Lys. p.Arg302Ter was first identified through linkage analysis in a patient from a consanguineous family with autosomal recessive disease (CARASIL). The proband had white matter hyperintensities on brain MRI, alopecia, spondylosis, stroke and dementia in the fourth decade of life. This variant has been shown in vitro to result in nonsense-mediated mRNA decay, and functional analyses demonstrated the resulting haploinsufficiency in the form of reduced enzyme activity, and failure of the formation of a covalent complex with α_1-antitrypsin as compared to the wild-type protein. The variant was also shown to result in the increased synthesis of extracellular matrix proteins, which theoretically results in vascular fibrosis.[61] This variant was subsequently reported in the heterozygous state in a 63-year-old patient with multiple lacunar infarcts, cognitive impairment and gait disturbance. This patient had white matter hyperintensities and multiple microbleeds on his brain MRI, however imaging and genetic analyses were not performed in any of his relatives.[264]

Evidence for the pathogenicity of different HTRA1 variants thus varies between reports. The presence of p.Gln151Lys in 0.1% of controls raises more questions rather than answers. This

variant may in fact have incomplete penetrance, or these carriers in the general population may instead be pre-symptomatic undiagnosed cases. It is also possible that the variant was falsely associated with disease to begin with. Similarly, the presence of p.Arg302Ter in a 22-year-old could be explained by the fact that autosomal dominant SVD is not expected to present at this age, and that this individual may in fact be a pre-symptomatic case.

Table 6-5*: Frequency of CARASIL and autosomal dominant SVD-causing mutations in the HTRA1 reported in literature, in unrelated BRIDGE controls*

Variant	Exon	CADD score	No. of unrelated controls	Frequency in controls	Sex of controls	Ages of controls at recruitment	Disease phenotype	Frequency in ExAC	Frequency in gnomAD	Reference	Remarks
p.Gln151Lys	1 of 9	31.0	8	5.4×10^{-4}	F	Not known	Bleeding and platelet disorder	5.3×10^{-5}	1.6×10^{-4}	[330]	
					M	Not known	Bleeding and platelet disorder				
					M	Not known	Bleeding and platelet disorder				
					F	Not known	Bleeding and platelet disorder				
					F	Not known	Not known				
					F	40	Intrahepatic cholestasis of pregnancy				
					M	34	Partial rod-cone dystrophy				
					M	Not known	Paediatric neurology				(Related individuals)
					M	Not known	Unaffected relative, recruited under paediatric neurology				
p.Arg302Ter	4 of 9	45.0	1	6.8×10^{-5}	F	22	Bleeding and platelet disorder	-	1.8×10^{-5}	[61,264]	

6.3.3 COL4A1

Two out of 37 reported COL4A1 mutations were identified in 11 unrelated controls, giving an overall diagnostic yield of 0.15%. p.Gln1316Glu, which has a CADD score of 11.3, was present in 8 individuals with an age range of 15 to 75 (allele frequency 5.4 x 10^{-4}). p.Arg1238Cys, which has a CADD score of 34, was found in a similar allele frequency (2.0 x 10^{-4}) to that in the ExAC database (2.1 x 10^{-4}), but is not found in gnomAD. (**Table 6-6**)

The heterozygous p.Gln1316Glu mutation was previously described in an infant with hydrocephalus, seizures, hypotonia and ocular abnormalities. MR imaging of the brain showed pronounced hydrocephalus, gyral abnormalities, thinning of the corpus callosum and pontine hypoplasia, and was diagnosed as being on the spectrum between Walker-Warburg Syndrome and Muscle-Eye-Brain disease (MEB/WWS).[268] MEB/WES is known in some cases to result from autosomal recessive mutations in genes encoding glycosyltransferases, however many other patients do not have mutations in these genes. A relationship between these syndromes and the SVD spectrum has not been suggested in literature.

The same variant, present in a low-complexity region of the protein (i.e. not affecting the Gly-X-Y repeat of the triple helix) was not present in 286 control chromosomes in the same study.[268] This variant thus has not specifically been described in cerebral small vessel disease manifesting as white matter hyperintensities, but has been identified in a related phenotype. As COL4A1 mutations are associated with familial porencephaly,[331] anterior segment dysgenesis and other ocular abnormalities,[332] muscle and renal disease,[165] it is possible that this is one example of the disease spectrum seen in COL4A1-related disease.

The association between the p.Arg1238Cys variant and familial SVD is also tenuous. This variant, which is also present in a low-complexity region of the protein, was reported in a child with congenital cataracts, epileptic seizures and spastic diplegia, showing bilateral white matter hyperintensities, a periventricular cyst and a focus of susceptibility on MRI. The child also had muscle cramps with raised creatine phosphokinase levels, and microhaematuria without proteinuria. The variant was inherited from the father, who had muscle cramps and Raynaud's phenomenon, but no ocular or brain abnormalities apart from a dermoid cyst. The child also carried a de novo splice site variant (c.2458 + 1G>A) of the COL4A1 gene, which is predicted to result in the skipping of exon 31 from the resulting mRNA transcript as confirmed

on real time PCR on the patient's skin fibroblasts. The authors hypothesized that the proband was a compound heterozygote, with each variant contributing to part of the phenotype as suggested by the presence of muscle cramps in both parent and child. No functional analyses were performed to investigate each variant in isolation,[267] and it is not known if these controls also had muscle cramps.

__Table 6-6:__ Frequency of COL4A1 variants associated with familial SVD in unrelated BRIDGE controls

Variant	Exon	CADD Score	No. of unrelated controls	Frequency in controls	Sex of controls	Ages of controls at recruitment	Disease phenotype	Frequency in ExAC	Frequency in gnomAD	Reference	Remarks
p.Gln1316Glu	44 of 52	11.3	8	5.4×10^{-4}	F	74.9	Pulmonary arterial hypertension	9.1×10^{-5}	1.3×10^{-4}	[268]	
					F	Not known	Not known				
					F	Not known	Not known				
					M	38.0	Pulmonary arterial hypertension				
					M	14.9	Paediatric neurology				
					F	64.6	Bleeding disorder				
					F	35.4	Thrombocytopenia				
					F	70.3	Membranoproliferative Glomerulonephritis				
p.Arg1238Cys	42 of 52	34.0	3	2.0×10^{-4}	M	44.3	Antibody defect	2.1×10^{-4}	-	[267]	Variant reported in an individual with one novel mutation, and one paternally inherited COL4A1 mutation. The father did not have white matter lesions on MRI
					F	37.4	Pulmonary arterial hypertension				
					F	Not known	Pulmonary arterial hypertension				

6.3.4 COL4A2

Of 10 reported COL4A2 mutations, three were identified in 175 unrelated controls, giving an overall diagnostic yield of 2.4%. The most common of these, p.Glu1123Gly, was identified in 152 controls (allele frequency 1.1%), of which 4 were homozygous for this variant (diagnostic yield 2.1%). This variant was also present in 4 individuals with familial SVD. The allele frequency of this variant was similar in other control databases (0.78% in ExAC, 0.95% in gnomAD). Similarly, p.Gln1150Lys was present in a similar frequency in this control dataset when compared to ExAC and gnomAD, with an allele frequency of 0.14%. (**Table 6-7**)

All 3 variants were first described in four singletons in a study of 96 patients with adult-onset intracerebral haemorrhage, and were absent in a dataset of 144 haemorrhage-free controls.[159] None of these individuals were reported to have a family history of haemorrhage or stroke, and genetic analysis was not performed on any relatives. Two of these individuals carried the p.Glu1123Gly variant. p.Glu1123Gly and p.Gln1150Lys affect the 'X' residue in the triple-helix Gly-X-Y repeat region of COL4A2, while p.Ala1690Thr is found in the NC1 domain of the protein. The NC1 domain is thought to have roles in stabilising the collagen trimer using methionine-lysine covalent bonds. In vitro functional assays demonstrated that these three variants impair the secretion of both COL4A1 and COL4A2 from the endoplasmic reticulum and trigger endoplasmic reticulum stress.[159]

Although p.Glu1123Gly was subsequently identified in a six-month-old infant with porencephaly and focal cortical dysplasia, segregation was again not demonstrated. The variant was present in their asymptomatic father, in whom an MRI had not been performed. Hence, the evidence for the pathogenicity of all three variants is limited, and it is possible that these are either variants with low penetrance, or polymorphisms which have been falsely associated with disease.

__Table 6-7:__ Frequency of COL4A2 variants associated with familial SVD in unrelated BRIDGE controls. p.Glu1123Gly is also present in 2 CSVD index cases.

Variant	Exon	CADD Score	No. of unrelated controls	Frequency in controls	Sex of controls	Ages of controls at recruitment	Disease phenotype	ExAC	gnomAD	Ref	Remarks
p.Glu1123Gly	37 of 48	25.1	152 (4 homozygous)	0.010	61 M 76 F 15 not known	Range 1.8 – 80.8 Mean ± SD 42.7 ± 20.2	Bleeding and platelet disorders (n=22) Hypertrophic cardiomyopathy (1) Intrahepatic cholestasis of pregnancy (2) Multiple primary malignant tumours (5) Pulmonary arterial hypertension (18) Primary immune disorder (18) Primary membranoproliferative glomerulonephritis (3) Paediatric neurology (21) Steroid resistant nephrotic syndrome (1) Not known (61)	0.0078	0.0095	[159,333]	Also in 2 unrelated CSVD index cases
p.Gln1150Lys	37 of 48	20	20	0.0014	12 M 11 F	Range 6.6 – 74.4 Mean ± SD 32.2 ± 22	Bleeding and platelet disorders (3) Primary immune disorder (4, 2 affected) Paediatric neurology (9) Steroid resistant nephrotic syndrome (3) Not known (1)	0.0019	0.0026	[159]	
p.Ala1690Thr	48 of 48	23	3	2.0×10^{-4}	F	44.1	Bleeding and platelet disorder	5.4×10^{-4}	6.3×10^{-4}	[159]	
					M	10.6	Steroid resistant nephrotic syndrome				
					M	Not known	Neuropathic pain disorder				

6.4 Conclusions

In this study we have investigated the frequency of known SVD-causing mutations in a control dataset of 7348 unrelated individuals, and demonstrated varying frequencies across different mutations, with some reaching levels seen in polymorphisms. Interpreted in the context of varying levels of evidence available for the pathogenicity of each variant, it is possible that some of these have been erroneously associated with disease. This study has also allowed us to glean several insights on the interpretation of known 'disease-causing' variants in the clinic.

6.4.1 Milder forms of familial SVD

We have shown the presence of mutations in a higher-than-expected frequency overall, with 0.26% of controls carrying CADASIL-causing mutations, where the population prevalence of CADASIL is estimated at 2 to 4 per 100,000 population in the UK.[40,41,334] Most of these cysteine-altering variants are also rare (less than 1 in 10 000) in ExAC and gnomAD. These variants may indeed be rare mutations which exist with a low background carrier rate in the control population, in which there may exist a number of undiagnosed individuals with CADASIL. These variants may also be incompletely penetrant, where another genetic or environmental factor, or both, are pre-requisites for the manifestation of early-onset disease.

Two of these variants (p.Cys1222Gly and p.Arg1231Cys) are present in a greater number of controls in this study, as well as in ExAC and gnomAD, suggesting that these variants may indeed contribute to a milder disease phenotype as has previously been suggested of variants in this portion of the protein.[59] These individuals may eventually have a later onset of symptoms, or milder symptoms that would not have warranted investigation for a familial disease. In a disease like CADASIL where the earliest feature (migraine with aura) is common in the population, and its more severe neurological features such as stroke and cognitive impairment arise relatively late (in the fourth and fifth decades) (Chapter 3), it would be reasonable to expect that such variants are not subject to the same selection pressures as those which result in diseases affecting reproductive ability, and thus may possibly be more common in the population than previously predicted. It is also possible that different variants in the gene have differing degrees of penetrance, and thus have different background carrier rates.

These findings thus call into question the thresholds of allele frequency currently used as a filter in the diagnosis of rare disease in the era of high throughput sequencing. Disease causing variants within a single gene may exist at varying frequencies in the population, and at higher frequencies than expected. Using arbitrary thresholds as a filtering strategy brings with it the risk of eliminating potentially causative mutations, or being over-inclusive resulting in large numbers of variants of uncertain significance.

6.4.2 Evaluating the strength of evidence for pathogenicity: segregation in reported cases

As illustrated with the example of two HTRA1 variants, there is significant variation in the quality of evidence available for different reported variants in the same gene. Initial associations of variants with disease were established through linkage analysis, demonstrating segregation of the variant with disease, and in vitro functional analyses showing a deleterious effect of the variant on normal gene function. Subsequent reports of variants in the same gene, identified through large screening studies of singletons with the same phenotype, may have been reported and accepted as disease-causing without being subjected to the same rigour. Variants reported in clinical genetic testing may have been based on some of these variants falsely associated with the disease, resulting in an erroneous genetic diagnosis being returned in the clinic.

6.4.3 In vitro functional analyses alone are insufficient to demonstrate pathogenicity

Despite the utility of in vitro assays of each variant, these may potentially also be misleading. As seen in CADASIL, loss- or gain-of-function mutations in the NOTCH3 gene are not associated with disease, and knock-outs of the gene do not result in CADASIL.[335] Assays of protein function may thus show a deleterious effect of a mutation, but would not recapitulate the entire disease mechanism. In this study, we have shown that the p.Gln1123Gly variant in COL4A2 is present in up to 2.1% of controls. This variant was previously reported in singletons with no clear evidence of familial disease, and functional studies suggested that this variant impairs the secretion of both COL4A1 and COL4A2 from the endoplasmic reticulum.[159] The

conclusions from these assays were thus incongruent with what would be expected for a common variant.

6.4.4 Moving forward

This study is limited by the fact that controls were from disease cohorts which may be related to SVD, such as paediatric neurology, and many were young patients who may not be disease-free, but are instead pre-symptomatic affected cases. An ideal control dataset would comprise of age-matched, MRI-phenotyped, disease-free controls. This would allow us to determine the true background carrier rate of mutations, and help to determine a threshold which can serve as an effective filter when sifting through large numbers of variants identified on sequencing. However, a true disease-free dataset does not exist, and phenotyping controls to determine whether they have SVD remains a challenge in the absence of routine MR imaging for large numbers of potential controls.

The high frequency of some known variants also calls into question the validity of previous reports of pathogenicity. Some of these variants have only been identified in singletons without segregation analysis, but have been reported as pathogenic by virtue of the fact that the same gene has previously been associated with the disease. The strength of evidence for each variant reported in literature should thus be weighed prior to their addition to clinical reports.

It is possible that some rare variants may alone result in mild, late-onset disease which would normally be classified as sporadic disease, by contributing to risk in a complex, non-Mendelian pattern of inheritance. While studying extremes in phenotype is an approach to boost our understanding of the mechanisms of disease, our study has highlighted the possibility that the genetic basis of SVD is more likely a continuum between, rather than a dichotomy of monogenic and sporadic disease.

Beyond isolated affected cases and functional assays, a more robust method of ascribing the pathogenicity of variants is to identify different pedigrees with the same phenotype and variant. This has been demonstrated in recent sequencing studies such as the Diagnosing Developmental Disorders project, which identified causative de novo variants for developmental disorders by identifying pedigrees with similar phenotypes.[252] Web-based

repositories of genotypic and phenotypic information such as Matchmaker Exchange (http://www.matchmakerexchange.org/) serve this purpose of helping clinicians compare cases.

6.5 Appendix

Table 6-8: *Cysteine-altering NOTCH3 variants reported as disease-causing, or identified in patients with CADASIL presenting to the UK National Referral Clinic for CADASIL.*

Coding variant	Protein variant	Exon	Reported in recessive cases	References
c.127C>T	p.Cys43Gly	2		[257,258]
c.128G>T	p.Cys43Phe	2		[336]
c.128G>C	p.Cys42Ser	2		[337]
c.145T>C	p.Cys49Arg	2		[259]
c.145T>G	p.Cys49Gly	2		[338]
c.146G>A	p.Cys49Tyr	2		[339]
c.146G>T	p.Cys49Phe	2		[257–259]
c.157G>T	p.Gly53Cys	2		[150]
c.160C>T	p.Arg54Cys	2		[259,340,341] Internal database
c.179C>G	p.Ser60Cys	2		[257,258]
c.193T>G	p.Cys65Gly	2		[342]
c.194G>A	p.Cys65Tyr	2		[343]
c.194G>C	p.Cys65Ser	2		[257,258]
c.199T>A	p.Cys67Ser	3		[344]
c.200G>A	p.Cys67Tyr	3		[345]
c.213G>T	p.Trp71Cys	3		[57]
c.226T>C	p.Cys76Arg	3		[245] Internal database
c.227G>A	p.Cys76Tyr	3		[59]
c.228T>G	p.Cys76Trp	3		[257,258]
c.226_234del	p.Cys76_Leu78del	3		[346]
c.231_248del	p.Gln77_Cys82del	3		[258]
c.239_253del	p.Asp80_Ser84del	3		[347]
c.259T>C	p.Cys87Arg	3		[257,258]
c.260G>A	p.Cys87Tyr	3		[258]
c.265G>T	p.Gly89Cys	3		[348,349]
c.268C>T	p.Arg90Cys	3		[57,259] Internal database
c.277_279dup	p.Cys93dup	3		[350]
c.278G>A	p.Cys93Tyr	3		[351]
c.278G>T	p.Cys93Phe	3		[44]
c.316T>C	p.Cys106Arg	3		[352]
c.318C>G	p.Cys106Trp	3		[258]
c.322T>C	p.Cys108Arg	3		[150]
c.323G>A	p.Cys108Tyr	3		[257,258]
c.323G>C	p.Cys108Ser	3		[353]
c.324C>G	p.Cys108Trp	3		[354]

Coding variant	Protein variant	Exon	Reported in recessive cases	References
c.328C>T	p.Arg110Cys	3		[124,244] Internal database
c.341-2A>G	p.Gly114_Pro120del	Intron 3		[355]
c.349T>C	p.Cys117Phe	4		[150,356]
c.350G>A	p.Cys117Tyr	4		[263]
c.350G>C	p.Cys117Ser	4		[357]
c.350G>T	p.Cys117Arg	4		[44]
c.353C>G	p.Ser118Cys	4		[259,358]
c.368G>A	p.Cys123Tyr	4		[340]
c.368G>T	p.Cys123Phe	4		[44]
c.382T>G	p.Cys128Gly	4		[359]
c.383G>A	p.Cys128Tyr	4		[351]
c.383G>T	p.Cys128Phe	4		LOVD Internal database
c.391G>T	p.Gly131Cys	4		[326]
c.397C>T	p.Arg133Cys	4	Yes	[57,360,361] Internal database
c.401G>A	p.Cys134Tyr	4		Internal database
c.402C>G	p.Cys134Trp	4		[39]
c.421C>T	p.Arg141Cys	4		[39] Internal database
c.425T>G	p.Phe142Cys	4		[244,259,351]
c.431G>A	p.Cys144Tyr	4		[44] Internal database
c.431G>C	p.Cys144Ser	4		[44]
c.431G>T	p.Cys144Phe	4		[362]
c.434C>G	p.Ser145Cys	4		[258]
c.436T>C	p.Cys146Arg	4		[57]
c.437G>A	p.Cys146Tyr	4		[363]
c.437G>T	p.Cys146Phe	4		[57]
c.445G>T	p.Gly149Cys	4		[258,364]
c.449A>G	p.Tyr150Cys	4		[44]
c.457C>T	p.Arg153Cys	4		[57,244,259,365] Internal database
c.459_467del	p.Arg153_Cys155del	4		[347]
c.460A>T	p.Ser154Cys	4		Internal database
c.463T>A	p.Cys155Ser	4		[263]
c.464G>A	p.Cys155Tyr	4		LOVD
c.464G>C	p.Cys155Ser	4		[258]
c.484T>A	p.Cys162Ser	4		[340]
c.484T>C	p.Cys162Arg	4		[366]
c.485G>A	p.Cys162Tyr	4		[356]
c.486C>G	p.Cys162Trp	4		[336]
c.493G>T	p.Gly165Cys	4		[263]

Coding variant	Protein variant	Exon	Reported in recessive cases	References
c.505C>T	p.Arg169Cys	4		[57,259] Internal database
c.511G>T	p.Gly171Cys	4		[57]
c.520T>A	p.Cys174Ser	4		[344]
c.520T>C	p.Cys174Arg	4		[367]
c.521_522delinsTG	p.Cys174Leu	4		[368]
c.521G>A	p.Cys174Tyr	4		[44]
c.521G>T	p.Cys174Phe	4		[369]
c.539C>G	p.Ser180Cys	4		[340]
c.542T>G	p.Phe181Cys	4		[370]
c.544C>T	p.Arg182Cys	4		[57,244] Internal database
c.547T>A	p.Cys183Ser	4	Yes	[44,371]
c.547T>C	p.Cys183Arg	4		[44,372] Internal database
c.548G>T	p.Cys183Phe	4		[257,258]
c.553T>A	p.Cys185Ser	4		[43] Internal database
c.553T>C	p.Cys185Arg	4		[57]
c.553T>G	p.Cys185Gly	4		[339]
c.566A>G	p.Tyr189Cys	4		[244,336]
c.580T>A	p.Cys194Ser	4		[245] Internal database
c.580T>C	p.Cys194Arg	4		[351] Internal database
c.581G>A	p.Cys194Tyr	4		[340]
c.581G>C	p.Cys194Ser	4		[43] Internal database
c.581G>T	p.Cys194Phe	4		[44]
c.601T>C	p.Cys201Arg	4		[373]
c.602G>A	p.Cys201Tyr	4		[258]
c.616T>C	p.Cys206Arg	4		[374]
c.617G>A	p.Cys206Tyr	4		[340]
c.619C>T	p.Arg207Cys	4		[336]
c.634T>A	p.Cys212Ser	4		[244,356,375]
c.634T>C	p.Cys212Arg	4		Internal database
c.635G>A	p.Cys212Tyr	4		[356,376]
c.636C>G	p.Cys212Trp	4		[357]
c.659A>G	p.Tyr220Cys	4		[354]
c.664T>G	p.Cys222Gly	4		[57]
c.665G>A	p.Cys222Tyr	4		[351]
c.665G>C	p.Cys222Ser	4		[150,259]
c.670T>C	p.Cys224Arg	4		[259]
c.671G>A	p.Cys224Thr	4		Internal database

Coding variant	Protein variant	Exon	Reported in recessive cases	References
c.671G>A	p.Cys224Tyr	4		[57]
	p.Cys194Gly	4		[244]
	p.Ser126Cys	4		Internal database
c.697T>A	p.Cys233Ser	5		[339]
c.697T>C	p.Cys233Arg	5		[43] Internal database
c.698G>A	p.Cys233Tyr	5		[258,364]
c.699T>G	p.Cys233Trp	5		[260]
c.714_758del	p.Asp239_Asp253del	5		[377]
c.719G>C	p.Cys240Ser	5		[258]
c.733T>A	p.Cys245Ser	5		[40] Internal database
c.733T>C	p.Cys245Arg	5		[258,364]
c.751T>A	p.Cys251Ser	5		[260]
c.751T>C	p.Cys251Arg	5		[245] Internal database
c.751T>G	p.Cys251Gly	5		[378]
c.752G>A	p.Cys251Tyr	5		[344]
c.773A>G	p.Tyr258Cys	5		[57,244]
c.778T>G	p.Cys260Gly	5		[379]
c.779G>A	p.Cys260Tyr	5		[257,258]
c.1004C>G	p.Ser335Cys	6		[257,258]
c.1010A>G	p.Tyr337Cys	6		[260]
c.1012T>C	p.Cys338Arg	6		[380]
c.812G>T	p.Cys271Phe	6		[381]
c.886G>T	p.Gly296Cys	6		[382]
c.895A>T	p.Ser299Cys	6		[383]
c.955_956delGCinsTG	p.Ala319Cys	6		[258]
c.967C>A	p.Cys323Ser	6		Internal database
c.994C>T	p.Arg332Cys	6		[259,384–386] Internal database
c.1013G>T	p.Cys338Phe	6		[244]
c.1057_1071dup	p.Asp353_Ser357dup	7		[387]
c.1064G>C	p.Cys355Ser	7		Internal database
c.1078T>C	p.Cys360Arg	7		[388]
c.1096T>C	p.Cys366Arg	7		HGMD
c.1098T>G	p.Cys366Trp	7		[389]
c.1135T>C	p.Cys379Arg	7		[390]
c.1136G>C	p.Cys379Ser	7		[257,258]
c.1136G>T	p.Cys379Phe	7		Internal database
c.1144G>T	p.Gly382Cys	7		[260]
c.1163G>A	p.Cys388Tyr	7		[261]
c.1183T>C	p.Cys395Arg	7		[257,258]
c.1187C>G	p.Ser396Cys	7		[353]

Coding variant	Protein variant	Exon	Reported in recessive cases	References
c.1222T>G	p.Cys408Gly	8		Internal database
c.1223G>A	p.Cys408Tyr	8		[259]
c.1241C>G	p.Ser414Cys	8		[262]
c.1257C>G	p.Cys419Trp	8		[259]
c.1258G>T	p.Gly420Cys	8		[339]
c.1261C>T	p.Arg421Cys	8		[257,258]
c.1279C>T	p.Arg427Cys	8		[259] Internal database
c.1282T>C	p.Cys428Arg	8		[380] Internal database
c.1283G>A	p.Cys428Tyr	8		[257,258]
c.1283G>C	p.Cys428Ser	8		[339]
c.1300_1308dup	p.Glu434_Leu436dup	8		[344]
c.1303T>C	p.Cys435Arg	8		[260]
c.1318T>A	p.Cys440Ser	8		[391]
c.1318T>C	p.Cys440Arg	8		[257,258]
c.1318T>G	p.Cys440Gly	8		[245] Internal database
c.1337G>C	p.Cys446Ser	8		[258]
c.1337G>T	p.Cys446Phe	8		[257]
c.1345C>T	p.Arg449Cys	8		[257,258]
c.1363T>C	p.Cys455Arg	8		[392]
c.1364G>A	p.Cys455Tyr	8		[262]
c.1364G>T	p.Cys455Phe	8		ClinVar
c.1370G>C	p.Cys457Ser	8		[43] Internal database
c.1394A>G	p.Tyr465Cys	9		[260,372]
c.1397G>A	p.Cys466Tyr	9		[244,393] Internal database
c.1426A>T	p.Ser476Cys	9		Internal database
c.1433G>A	p.Cys478Tyr	9		[394]
c.1450T>G	p.Cys484Gly	9		[244]
c.1451G>A	p.Cys484Tyr	9		[258]
c.1451G>T	p.Cys484Phe	9		[257]
c.1484G>A	p.Cys495Tyr	9		[257,258]
c.1510T>C	p.Cys504Arg	10		[395]
c.1531T>C	p.Cys511Arg	10		[257,258]
c.1532G>A	p.Cys511Tyr	10		[396]
c.1532G>T	p.Cys511Phe	10		[397]
c.1582G>T	p.Gly528Cys	10	Yes	[380,398]
c.1592G>C	p.Cys531Ser	10		[399]
c.1594C>T	p.Arg532Cys	10		[400]
c.1624T>C	p.Cys542Arg	10		[262]
c.1625G>A	p.Cys542Tyr	11		[57]
c.1630C>T	p.Arg544Cys	11	Yes	[259,260,401,402]

Coding variant	Protein variant	Exon	Reported in recessive cases	References
c.1645T>C	p.Cys549Arg	11		[375]
c.1646G>A	p.Cys549Tyr	11		[257,258]
c.1672C>T	p.Arg558Cys	11		[57]
c.1703G>A	p.Cys568Tyr	11		[328]
c.1721A>G	p.Tyr574Cys	11		[399]
c.1732C>T	p.Arg578Cys	11	Yes	[57,403]
c.1735T>C	p.Cys579Arg	11		[404]
c.1759C>T	p.Arg587Cys	11		[259,262]
c.1771T>C	p.Cys591Arg	11		LOVD
c.1774C>T	p.Arg592Cys	11		[405] Internal database
c.1789C>T	p.Cys597Arg	11		Internal database
c.1790G>C	p.Cys597Ser	11		[406]
c.1791C>G	p.Cys597Trp	11		[372]
c.1816T>C	p.Cys606Arg	11		[353]
c.1819C>T	p.Arg607Cys	11		[340] Internal
c.1918C>T	p.Arg640Cys	12		LOVD Internal database
c.1999G>T	p.Gly667Cys	13		LOVD
c.2038C>T	p.Arg680Cys	13		[407]
c.2129A>G	p.Tyr710Cys	13		[408]
c.2149C>T	p.Arg717Cys	14		[259]
c.2182C>T	p.Arg728Cys	14		[57]
c.2324G>C	p.Cys775Ser	15		[257]
c.2411-1G>T	p.Gly804_Asn856delinsAsp	Intron 15		[409]
c.2815T>C	p.Cys939Arg	18		[353]
c.2857G>T	p.Gly953Cys	18		[245] Internal database
c.2898C>A	p.Cys966Ter	18	Yes	[410]
c.2923G>T	p.Gly975Cys	18		[369]
c.2929T>A	p.Cys977Ser	18		[259,395]
c.2951T>G	p.Phe984Cys	18		[244,340]
c.2953C>T	p.Arg985Cys	18		[57] Internal database
c.2963G>A	p.Cys988Tyr	18		[262]
c.2989T>G	p.Cys997Gly	18		[326,372]
c.3011G>A	p.Cys1004Tyr	19		[54]
c.3016C>T	p.Arg1006Cys	19		[57,411]
c.3037G>T	p.Gly1013Cys	19		[353]
c.3043T>C	p.Cys1015Arg	19		[53]
c.3062A>G	p.Tyr1021Cys	19		[351]
c.3065G>T	p.Cys1022Phe	19		[412]
c.3084G>T	p.Trp1028Cys	19		[413]

Coding variant	Protein variant	Exon	Reported in recessive cases	References
c.3091C>T	p.Arg1031Cys	19		[57]
c.3172G>T	p.Gly1058Cys	20		[351]
c.3182G>A	p.Cys1061Tyr	20		LOVD
c.3206A>G	p.Tyr1069Cys	20		[344]
c.3226C>T	p.Arg1076Cys	20		[53,259]
c.3296G>A	p.Cys1099Tyr	20		[328]
c.3317A>G	p.Tyr1106Cys	20		[372]
c.3356G>A	p.Cys1119Thr	20		Internal database
c.3356G>A	p.Cys1119Tyr	20		Internal database
c.3393C>G	p.Cys1131Trp	21		[414]
c.3427C>T	p.Arg1143Cys	21		HGMD
c.3471C>G	p.Cys1157Trp	21		LOVD
c.3664 T>G	p.Cys1222Gly	22		[256]
c.3691C>T	p.Arg1231Cys	22	Yes	[57,58,176,245,325,326] Internal database
c.3750C>G	p.Cys1250Trp	23		[390]
c.3748T>G	p.Cys1250Gly	23		Internal database
c.3781T>C	p.Cys1261Arg	23		[39]
c.3782G>A	p.Cys1261Tyr	23		[257,258]
c.3893G>T	p.Cys1298Phe	24		[415]
c.3944G>A	p.Cys1315Tyr	24		[416]

LOVD: Leiden Open Variation Database (www.lovd.nl).

HGMD: The Human Gene Mutation Database (http://www.hgmd.cf.ac.uk)

Table 6-*9:** Cysteine-sparing NOTCH3 variants previously associated with CADASIL. These variants do not lead to a gain or loss of a cysteine residue in the epidermal growth factor-like repeat region encoded by exons 2 to 24 of the gene, and are no longer considered pathogenic.*

Coding variant	Protein variant	Exon	Reference
c.259C>T	p.Arg61Trp	2	[327]
c.224G>C	p.Arg75Pro	3	[417]
	p.Asp80Gly	3	[315,319,320]
c.451C>G	p.Gln151Glu	4	[263]
c.509A>G	P.His170Arg	4	[263]
c.605C>T	P.Ala202Val	4	[418]
c.638G>A	p.Arg213Lys	4	[369]
c.709G>A	p.Val237Met	5	[419]
	p.Val252Met	5	[420]
	p.Val322Met	6	[420]
c.1729A>G	p.Thr577Ala	11	[328]
c.2932A>C	p.Ser978Arg	18	[328]
c.3058G>C	p.Ala1020Pro	19	[316]
c.3292A>T	p.Tyr1098Ser	20	[346]
	p.Gly1298Thr	24	[421]
c.4038G>C	p.Gly1347Arg	24	[329]

Table 6-10: *HTRA1 variants reported in CARASIL or autosomal dominant SVD.*

Coding variant	Protein variant	CARASIL	Autosomal dominant SVD	References
c.126delG	p.Gly42fs	Yes	No	[422]
c.161_162insAG	p.Gly56Alafs*160	Yes	No	[423]
c.361A.C	p.Ser121Arg	No	Yes	[67]
c.367G>T	p.Ala123Ser	No	Yes	[67]
c.397C>G	p.Arg133Gly	No	Yes	[67]
	p.Gln151Lys	No	Yes	[265]
	p.Ser163Gly	No	Yes	[265]
c.496C>T	p.Arg166Cys	Yes	No	[66]
c.497G>T	p.Arg166Leu	No	Yes	[67]
c.517G>C	p.Ala173Pro	No	Yes	[67]
c.517G>A	p.Ala173Thr	Yes	No	[424]
	p.Val175Met	No	Yes	[265]
	p.Gly206Glu	No	Yes	[265]
c.616G>A	p.Gly206Arg	Yes	No	[425]
c.754G>A	p.Ala252Thr	Yes	No	[61]
c.821G>A	p.Arg274Gln	No	Yes	[426]
	p.Gly283Glu	No	Yes	[174]
c.850A>G	p.Ser284Gly	No	Yes	[67]
c.852C>A	p.Ser284Arg	No	Yes	[67]
c.854C>A	p.Pro285Gln	No	Yes	[67]
c.854C>T	p.Pro285Leu	No	Yes	[293]
c.856T>G	p.Phe286Val	No	Yes	[67]
c.883G>A	p.Gly295Arg	Yes	No	[65]
c.889G>A	p.Val297Met	Yes	No	[61]
	p.Arg302Gln	No	Yes	[174]
c.904C>T	p.Arg302Ter	Yes	Yes	[61,264]
	p.Thr319Ile	No	Yes	[174]
c.961G>A	p.Ala321Thr	Yes	No	[422]
c.973-1G>A	p.Tyr325_Leu335del	No	Yes	[67]
c.1091T>C	p.Leu364Pro	Yes	No	[427]
c.1108C>T	p.Arg370Ter	Yes	No (see chapter 5)	[61]
c.1348G>C	p.Asp450His	No	Yes	[67]

Table 6-11: *COL4A1 variants reported in familial disease, including paediatric porencephalyor schizencephaly, anterior segment dysgenesis, ophthalmological and renal abnormalities. All variants reported were in a heterozygous state.*

Coding variant	Protein variant	References
c.*31G>T	3'UTR	[208]
c.*32G>T	3'UTR	[208]
c.*32G>A	3'UTR	[209]
c.*32G>T	3'UTR	[208]
c.*35C>A	3'UTR	[208]
c.1A>T	p.Met1Leu	[162]
c.502G>A	p.Gly168Arg	[428]
	p.Gly188Glu	[429]
c.1055C>T	p.Pro352Leu	[156]
c.1249G>C	p.Gly417Arg	[430]
c.1493G>A	p.Gly498Asp	[165]
c.1493G>T	p.Gly498Val	[431,432]
c.1528G>A	p.Gly510Arg	[165,433]
c.1537G>A	p.Gly513Ser	[434]
c.1555G>A	p.Gly519Arg	[85,435]
c.1573_1574GG>TT	p.Gly525Leu	[165]
c.1583G>A	p.Gly528Glu	[85,435]
c.1612C>G	p.Arg538Gly	[156]
c.1685G>A	p.Gly562Glu	[83,84]
c.1835G>A	p.Gly612Asp	[436]
c.1870G>T	p.Gly624Ter	[269]
c.1963G>A	p.Gly655Arg	[436]
c.1964G>A	p.Gly655Glu	[269]
c.1973G>A	p.Gly658Asp	[437]
c.2008G>A	p.Gly670Arg	[269]
c.2063G>A	p.Gly688Asp	[438]
c.2078G>A	p.Gly693Glu	[437]
c.2085del	p.Gly696fs	[294]
c.2086G>A	p.Gly696Ser	[439]
c.2122G>A	p.Gly708Arg	[440]
c.2123G>T	p.Gly708Val	[441]
c.2159G>A	p.Gly720Asp	[442–444]
c.2245G>A	p.Gly749Ser	[331]
c.2317G>A	p.Gly773Arg	[440,445,446]
c.2263G>A	p.Gly755Arg	[443,445,447,448]
c.2345G>C	p.Gly782Ala	[449]
	p.Gly785Glu	[450]
c.2413G>A	p.Gly805Arg	[451]
c.2423G>T	p.Gly808Val	[452]

Coding variant	Protein variant	References
c.2424delT	p.Pro810fs	[269]
c.2581G>C	p.Gly861Ser	[269]
c.2608G>A	p.Gly870Arg	[436]
c.2636G>A	p.Gly879Glu	[269]
c.2645G>A	p.Gly882Asp	[445]
c.2662G>A	p.Gly888Arg	[269,430]
c.2689G>A	p.Gly897Ser	[436]
c.2903G>A	p.Arg968Gln	Internal database
c.2931dupT	p.Gly978TrpfsTer15	[436]
c.2969G>A	p.Gly990Glu	[437]
c.2969G>T	p.Gly990Val	[453]
c.3005G>A	p.Gly1002Asp	[446]
c.3022G>A	p.Gly1008Arg	[452]
c.3046A>G	p.Met1016Val	[268]
c.3104G>T	p.Gly1035Val	[454]
c.3122G>A	p.Gly1041Glu	[436]
c.3130G>C	p.Gly1044Arg	[452]
c.3200G>C	p.Gly1067Ala	[269]
c.3245G>A	p.Gly1082Glu	[436]
c.3280G>A	p.Gly1094Arg	[269]
	p.Gly1103Arg	[455]
c.3389G>A	p.Gly1130Asp	[162]
c.3706G>A	p.Gly1236Arg	[331,456]
c.3712C>A	p.Arg1238Cys	[267]
c.3715G>A	p.Gly1239Arg	[457]
c.3770G>A	p.Gly1257Glu	[269]
c.3796G>C	p.Gly1266Arg	[445]
c.3941G>T	p.Gly1314Val	[437]
c.3946C>G	p.Gln1316Glu	[268]
c.3976G>A	p.Gly1326Arg	[436]
c.3995G>A	p.Gly1332Asp	[436]
c.4031G>C	p.Gly1344Ala	[458]
c.4133G>A	p.Gly1378Asp	[437]
c.4150G>A	p.Gly1384Ser	[437]
c.4267G>C	p.Gly1423Arg	[162]
c.4582-4586dupCCCATG	p.Pro1530_Met1531dup	[459]
c.4611_4612insG	p.Thr1537fs	[255]
c.4738G>C	p.Gly1580Arg	[296]
c.4738G>A	p.Gly1580Ser	[460]
c.4739G>C	p.Gly1580Ala	[269]
c.4814_4816insGGG	p.ins1605Gly	[461]
c.4843G>A	p.Glu1615Lys	[436]
c.4881C>G	p.Asn1627Lys	[462]

Coding variant	Protein variant	References
c.4887C>A	p.Tyr1629Ter	[436]

Table 6-12: *COL4A2 variants reported in familial disease, including paediatric porencephaly or schizencephaly, anterior segment dysgenesis, ophthalmological and renal abnormalities. All variants reported were in a heterozygous state.*

Coding variant	Protein variant	Reference
c.2105G>A	p.Gly702Asp	[164]
c.2399G>A	p.Gly800Glu	[463]
c.2821G>A	p.Gly941Arg	[464]
c.3110G>A	p.Gly1037Glu	[465]
c.3206delC	p.Arg1069fs	[86]
c.3368A>G	p.Glu1123Gly	[159,333]
c.3448C>A	p.Gln1150Lys	[159]
c.3455G>A	p.Gly1152Asp	[465]
c.4165G>A	p.Gly1389Arg	[86]
c.5068G>A	p.Ala1690Thr	[159]

Chapter 7: Conclusions and future directions

Outline

7.1 Key findings

- 7.1.1 CADASIL: Migraine, encephalopathy and strokes (Chapter 3)
- 7.1.2 A candidate gene study in presumed sporadic SVD (Chapter 4)
- 7.1.3 Rare variants in suspected familial SVD: a candidate gene approach using whole genome sequencing (Chapter 5)
- 7.1.4 The frequency of known mutations in the general population (Chapter 6)

7.2 Moving forward: suggestions for future work

- 7.2.1 Familial disease: linkage analysis using whole genome sequencing data
- 7.2.2 Familial disease: comparison of genotyped and phenotyped cases
- 7.2.3 Familial disease: other candidate gene approaches
- 7.2.4 Familial disease: non-coding variation
- 7.2.5 Sporadic disease: rare variant, common disease hypothesis
- 7.2.6 Discussion
- 7.2.7 SVD as a spectrum – fluid definitions of 'monogenic' or 'sporadic' disease
- 7.2.8 Variant interpretation in the age of sequencing
- 7.2.9 Controls, allele frequency and functional analyses

7.3 Concluding Remarks

7. Conclusions and future directions

Recent advances in elucidating disease pathways in monogenic forms of SVD have transformed our understanding of the mechanisms of SVD in general (Chapter 1). In line with these findings, this thesis aimed to explore the genetic aetiology of cerebral SVD primarily from a monogenic disease perspective. Approaching SVD from both ends of the disease spectrum, we endeavoured to dissect the role of monogenic disease genes in both presumed familial and sporadic forms of SVD, and to explore the role of genes associated with sporadic SVD in suspected familial cases.

We began by illustrating the variability of phenotypes and ages at onset of familial SVD (Chapter 1), particularly with CADASIL as an example (Chapter 3). We next investigated the prevalence of known and novel variants in the coding regions of candidate genes associated with monogenic disease in a cohort of patients with presumed sporadic SVD using a high throughput sequencing (HTS) gene panel (Chapter 4). In addition, a similar candidate gene approach was taken in a cohort of presumed familial SVD patients using whole genome sequencing data (Chapter 5). Finally, we examined the presence of previously reported disease-causing variants in a cohort of whole-genome sequenced controls (Chapter 6).

Through this body of work, we have examined the phenotypes seen in familial SVD, and have explored not only the wealth of information provided by performing high throughput sequencing on families, but also the limitations faced when interpreting this data. This chapter will briefly highlight key findings from each chapter of the thesis and discuss future directions.

7.1 Key findings

7.1.1 CADASIL: Migraine, encephalopathy and stroke (Chapter 3)

CADASIL is the most well-described form of monogenic SVD. In a cohort of 300 symptomatic CADASIL patients we illustrated the variability in the ages at onset, and the broad range of presentations of the disease. We speculated a possible relationship between migraine with aura and a lower incidence of stroke, as well as between migraine with confusional aura and encephalopathy in CADASIL. The medical management of CADASIL symptoms has not been extensively investigated, hence we have also retrospectively reviewed patients' responses to different medications used in the prevention of and relief of migraines in CADASIL. As a relatively well-understood prototype of monogenic SVD, we have used CADASIL to illustrate the heterogeneity of presentations of disease both within and between families, despite the simplicity of its pattern of inheritance.

7.1.2 A candidate gene study in presumed sporadic SVD (Chapter 4)

Using the ThromboGenomics sub-exome panel sequencing platform, we performed high throughput sequencing (HTS) of the coding regions of 15 candidate genes associated with familial SVD or its related phenotypes in 993 individuals with presumed early-onset (≤70 years) sporadic SVD. With this panel we achieved high coverage of the coding regions of each candidate gene.

We show that a large number of rare variants can be identified using such a HTS technique, and also demonstrate that 1.1% of this cohort were found to have CADASIL-causing mutations, of which mutations in two individuals had not been identified on a previous screen using denaturing high-performance liquid chromatography (DHPLC) and Sanger sequencing. This suggests that a multi-gene HTS panel may be more sensitive and efficient in screening multiple genes simultaneously in patients with suspected familial SVD.

We found that known disease-causing and novel variants in NOTCH3, HTRA1, FOXC1, TREX1, COL4A1 and COL4A2 may account for a few of these presumed-sporadic SVD cases, suggesting that this panel may also be useful for screening patients where there is a low index of suspicion for a single-gene disease. However, clinical interpretation of the role of both known and novel variants remains a key challenge.

The presence of known disease-causing variants in this cohort suggests that there may be a continuum between the between the phenotypes seen in presumed familial or sporadic disease, rather than a clear distinction. In this small sample of patients and relatively narrow selection of candidate genes we also found that each individual may have multiple variants identified in different genes, depending on the stringency of the filters used. This highlights the complexities of interpreting data produced using such techniques.

7.1.3 Rare variants in suspected familial SVD: a candidate gene approach using whole genome sequencing (Chapter 5)

We performed whole genome sequencing in 118 index cases with suspected familial SVD who did not have CADASIL-causing mutations identified on clinical genetic testing. As seen in Chapter 4 we similarly identified cases with CADASIL-causing mutations that had not previously been identified on Sanger sequencing or DHPLC.

Using a candidate gene approach, we show that HTRA1 is the next most common gene affected in familial SVD, with rates in the UK being similar to those seen in Italy, France and Japan. We propose that some of the variants previously only seen in recessive disease (CARASIL) may in fact also be pathogenic in heterozygous individuals. We also report a third family with the only known mutation in the CTSA gene causing CARASAL.[299] We also suggest a potential role for the CACNA1A gene, which has been implicated in multiple neurological diseases, in SVD. In several pedigrees we also found that genes associated with sporadic SVD may also act in a Mendelian fashion to cause familial disease.

Finally, we demonstrate that the majority of patients have a large number of potentially causative variants identified through the filters we have used, which further illustrates how high throughput sequencing on the background of our current understanding of disease and genetics is a double-edged sword: while we can rapidly screen multiple genes simultaneously, we are also left with a deluge of data to interpret in a clinical context. This highlights the importance of developing additional strategies beyond predictive tools, in vitro functional analyses and minor allele frequency to establish the true causative mutation in each case.

7.1.4 The frequency of known mutations in the general population (Chapter 6)

As a critical evaluation of existing methods in literature, which we have employed when sifting potentially disease-causing mutations from other variants identified on HTS, we examined the

frequency of known mutations in the general population – both in patients recruited in non-SVD rare diseases, and in established exome and whole genome sequencing databases.

We show that some variants exist in high frequencies in the population and may in fact be polymorphisms that have been erroneously reported as disease-causing due to the relatively small sample size of the control populations used in earlier studies. This highlights the importance of using large age- and ethnicity matched, well-phenotyped datasets in establishing the true background frequency of a variant, and hence its likelihood of pathogenicity.

At the same time, however, our findings also cast doubt on the utility of an arbitrarily determined minor allele frequency cut-off as a filter for pathogenic variants. As the 'general population' or non-SVD patients are by no means disease-free controls, there may be milder forms of SVD that are undiagnosed in the population, thus resulting in the premature exclusion of potentially causative mutations. Truly pathogenic variants may in fact cause a spectrum of severity and symptoms as we have seen in CADASIL (Chapters 1 and 3).

7.2 Moving forward: suggestions for future work

7.2.1 Familial disease: linkage analysis using whole genome sequencing data

Although linkage analysis was the main statistical genetic approach used in mapping Mendelian diseases, they were not always successful and were often limited by the size of families, with larger, cross-generational families and distantly related affected individuals being more likely to provide statistical power. On the other hand, genome-wide studies provided statistical power for studying complex traits on a population basis, but were found to have limited utility for Mendelian diseases. HTS approaches now offer cost- and time-effective identification of rare variants in both complex and monogenic disease, however these techniques when applied in the rare disease context usually lack statistical power, and filtering approaches within families are likely to lead to multiple hits segregating with disease.

A fundamental lesson learnt through our work is that rather than having hundreds of index cases which would never reach statistical power for proving causality in multiple rare variants through conventional Bonferroni-corrected testing, it may be more useful to have multiple affected and unaffected individuals across generations in a single family. In a pedigree of 4 individuals (2 affected, 2 unaffected) a 3' untranslated variant in COL4A1 segregating with

disease failed to pass our filters and was relegated to having a status as a 'variant of uncertain significance', whereas the variant was found to segregate in the same family when it was studied with 10 affected and 10 unaffected individuals. A genetic diagnosis was thus achieved through the combination of exome sequencing and linkage analysis.

Methods in monogenic disease genetics have now, in some respects, come full circle. The combination of HTS and linkage analysis brings together the advantages from both approaches, and has led to the successful identification of disease-causing variants within families with suspected Mendelian inheritance of disease. Some other examples have been seen in a large pedigrees with SVD.[208,209] Such a pedigree-based approach in the context of new sequencing technologies may well help us achieve genetic diagnoses in the remainder of unsolved cases. Ongoing work extending from that reported in this thesis involves the recruitment of additional relatives from each pedigree recruited, and the MR phenotyping of both symptomatic and asymptomatic relatives. This will provide us with the data required to perform robust linkage analyses.

7.2.2 Familial disease: comparison of genotyped and phenotyped cases

Recruiting multiple affected individuals across generations and branches of large pedigrees can be a challenge in genetic diseases, particularly in one with a late age at onset and with varying severity even within families. An alternative to validating the genotype for cases is to find similar pedigrees with the same phenotype and genotype. With deep phenotyping there may be a role for the use of online clinical genomics databases such as the DECIPHER project (https://decipher.sanger.ac.uk), where more than 18,000 cases have been uploaded with phenotypic and genotypic information, and with an option to contact referring clinicians for more information about reported cases.[252]

Similarly, Matchmaker Exchange (http://www.matchmakerexchange.org/) is a web-based platform which serves as a bridge between clinicians who have cases with similar phenotypic or genotypic profiles. This may help to interpret variants of uncertain significance uncovered in patients with rare diseases, and thus promote the discovery of the aetiology behind these cases.[466]

7.2.3 Familial disease: other candidate gene approaches

In recent years the matrisome has taken centrestage in our growing understanding of the mechanisms of monogenic SVD.[118] There are more than three hundred different proteins in the matrisome,[467] and aberrations of each may well be associated with disease. This has been discussed in depth in Chapter 1. As part of an initial screen for mutations it would be worth considering genes encoding for key matrisome proteins. Variants arising in the same gene across multiple patients with similar phenotypes may point towards a novel disease gene, and by extension a novel disease pathway. Planned extensions of our whole genome sequencing work involves the screening of genes encoding matrisome proteins involved in the cerebral microvasculature.[118,468]

As many of the variants identified in existing candidate genes were of uncertain significance, it may also be worth exploring the option of looking for enrichment of variants in a cohort of affected individuals as compared to controls. Again, this would require a large, well-phenotyped and ethnicity-matched control cohort in order to provide any robust evidence.

7.2.4 Familial disease: non-coding variation

In this thesis we have focused mainly on the coding spaces, which at 2 to 3% of the genome is sufficiently large to raise more variants of uncertain significance than we can manage with our present knowledge. However, we have seen from COL4A1 that variants in the 3' untranslated region can have an impact on the binding site of microRNAs, and that these variants are associated with familial SVD.[208,295] Studies in other diseases have also illustrated that non-coding variants may introduce a new splice site, and thus affect the resulting protein.[284]

Candidate gene approaches may serve us well in the short term to exclude mutations in known genes, however with larger pedigrees and better curated control databases it will become necessary and fruitful to explore the non-coding genome. Functional analyses can also be performed, such as using RNA sequencing to establish the impact of a non-coding variant on the resulting transcript.

7.2.5 Sporadic disease: rare variant, common disease hypothesis

GWAS studies capture the contribution of common variants to complex diseases such as SVD.[29,469] However, there remains a proportion of missing heritability which may likely be explained by rare variants. These are only detectable by HTS techniques, and achieving a statistically significant result is challenging in the absence of a massive, robust set of controls. The definition of a 'rare' variant is also arbitrary and widely differs between studies.[470]

One possible extension of GWAS studies is to deep-sequence the gene loci identified, as the lead single nucleotide polymorphism (SNP) may be tagging the causative rare variant in the same region. The downside of this is that variants may exert their effects in cis or trans, and a linear capture-sequencing approach that does not take into account the three-dimensional folding of chromosomes may potentially be an expensive gamble. A possible compromise would be to also consider expression quantitative trait loci (eQTL) and the folding of chromosomes in order to select regions to sequence.

One other option is to perform burden association tests with a robust control data set. This essentially asks the question of whether there is a difference between the load of variants in cases as opposed to controls. Planned future work currently involves combining our data set of 993 individuals with that from other populations in Europe, which may lead to greater statistical power. A fundamental component of such an analysis is a well-phenotyped, ethnicity-matched set of controls.

7.3 Discussion

7.3.1 SVD as a spectrum – fluid definitions of 'monogenic' or 'sporadic' disease

Determining the true pathogenicity and penetrance of a variant requires a sound definition of the disease phenotype, which may be complicated in cases where patients do not have syndromic features of Mendelian disease. While earlier descriptions of familial disease are likely to have captured the most severe forms of the disease, it is probable that there exist milder forms in the population that have gone undiagnosed – as we have seen in Chapter 4.

For example, classifying a 55-year-old patient with a history of hypertension, lacunar stroke and features of SVD seen on imaging, as well as a family history of stroke and dementia, as a

suspected familial or sporadic case can be challenging. Is there a single gene mutation with low penetrance causing a milder form of disease that has presented relatively late in life? Or could multiple low impact variants in the context of a shared environment and/or existing cardiovascular risk factors be sufficient to explain the symptoms?

In the same vein, assigning disease status to an individual in the context of a pedigree for segregation analysis is challenging. Phenotyping is often based on clinical judgment and can be subject to variation between clinicians. Is a relative with mild features of SVD on MRI and coexisting cardiovascular risk factors who becomes symptomatic at a later age affected or unaffected?

Our definitions of disease and phenotyping of types of SVD are certainly called into question, and the decision remains a clinical one that will inevitably result in inter-rater variability. It may be the case that the phenotype assigned to each patient needs to be fluid between different studies involving the same patients, or that only the very extremes in phenotype such as those with syndromic features are studied.

7.3.2 Variant interpretation in the age of sequencing

In the context of HTS technology we are now burdened with more questions than answers. An important question to grapple with is whether SVD is in fact a spectrum, with familial disease at one end and hypertensive sporadic disease at the other. We have attempted to suggest a possible role for both ’Mendelian’ and ‘sporadic’ disease genes in the development of disease anywhere along this spectrum, where variants in either of these categories may have varying degrees of penetrance.

As the cost of sequencing continues to fall and turnaround times shorten, it is important that we address the key challenge of interpreting the data generated. Predictive tools such as the Variant Effect Predictor or the CADD score can help to estimate likely impact of variants. For example, loss-of-function variants introducing a premature termination codon are predicted to be of ‘high’ impact, while missense variants are thought to result in ‘moderate’ impact. Hence, such tools can help us give greater weight to nonsense and frameshift mutations which result in a significant change in protein function, as opposed to a synonymous variant. Yet, a key lesson learnt from studies in the pathogenesis of CADASIL is that disease is likely to result from variants altering cleavage and deposition of domains of the protein – while

variants resulting in a loss- or gain-of-function of the protein are not associated with disease.[125] Furthermore, nonsense variants may not always result in a loss of function: in some proteins these may still permit the production of functional protein.[471] Isolating a single causative genetic mutation where a particular inheritance pattern has been hypothesized is thus difficult when multiple variants have been identified, especially since a loss- or gain-of-function variant may not necessarily result in disease.

A sound understanding of the disease mechanism is thus inherent in our prediction of the impact of variants subsequently identified in a known disease-associated gene. It is of critical importance that subsequent variants reported in a known disease gene are subject to the same standards as those initially identified. In order to recognise a variant as disease-causing, it would be preferable to demonstrate segregation of the variant with disease in a large pedigree, alongside multiple functional in-vitro analyses and animal models to serve as convincing evidence of pathogenicity.

It is also important to recognise that rare variants may not necessarily have high impact and common variants may not always have low impact. The context of an individual's genetic and environmental factors, and the fact that rare forms of SVD do not affect survival before the reproductive years unfortunately muddy the waters, clouding our interpretation of the true impact of each variant.

Other models of disease must also be considered: patients may require a 'double hit', or more than one mutation for the disease to develop. Patients may also have other rare variants that are 'protective' and prevent development of disease in context of presence of a disease-causing mutation, further complicating our interpretation of the genetic picture. Mutations may not be as penetrant as expected, and individuals may require other genetic or environmental modulators in addition to the mutation in order to develop the disease.

7.3.3 Controls, allele frequency and functional analyses

In this thesis we have demonstrated the limitations of existing control databases. We have used patients with other rare diseases- which may or may not be related to one of these monogenic forms of SVD, and have not been MRI-phenotyped to verify the absence of any cerebral disease.

A truly 'disease-free' control cohort does not exist, and in the absence of a large set of age- and ethnicity-matched controls we run the risk of including pre-symptomatic individuals who are in fact affected. In Chapter 6 we also demonstrate how the 'mutations' shown to have deleterious effects on function in vitro may in fact simply be polymorphisms. Minor allele frequency or in-vitro functional studies alone are less helpful in the context of a rare disease with a later age at onset.

In Chapters 4 and 6 we have also demonstrated that known disease-causing mutations may in fact be present in higher-than-expected frequencies in the general population. This highlights the importance of using different filtering strategies in addition to the minor allele frequency, which may inadvertently exclude disease-associated variants. As a corollary to this, we will also find ourselves reclassifying mutations as variants in light of more robust databases and functional analyses.

We have also seen the limitations of in-vitro functional analyses used to illustrate the pathogenicity of variants identified in diseased individuals. Studies have shown clear evidence that the p.Gln1123Gly variant in COL4A2 is associated with impaired secretion of both COL4A1 and COL4A2 from the endoplasmic reticulum, as well as endoplasmic reticulum stress.[159] With the availability of larger control databases this variant is now appearing to be a polymorphism rather than a rare, disease-causing mutation.

7.4 Concluding Remarks

In rare diseases we endeavour to discover a single rare variant to account for all of the features of disease, while GWAS studies examine the roles of common variants in common disease. It is likely that there also exist rare variants contributing to the heritability of common diseases, as well as variants with intermediate effect on disease at both ends of the spectrum.

The involvement of genes such as COL4A1 and COL4A2 in both monogenic and sporadic SVD,[157,166] and the presence of variants in genes associated with sporadic disease in pedigrees with familial disease highlights the possibility that the disease is not dichotomous. Instead, the lines between familial and sporadic disease are increasingly blurred and they may in fact belong to two ends of a single disease spectrum. This is in line with recent advancements in our understanding of the genetic basis of monogenic SVD, which points towards the convergence of pathways between the different diseases. Similarly, SVD and its associated

phenotypes such as familial hemiplegic migraine may simply be different axes on a plane, along which patients lie at different positions manifesting each phenotype to different extents, depending on their genetic background and cardiovascular risk factors. It is likely that multiple distinct pathways lead to different manifestations SVD in different groups of patients.[29]

In view of this changing landscape, a more flexible approach to phenotyping variants and sifting out variants identified on sequencing could be taken. For example, a family with 'familial disease' may need to be assumed to have a multiple-gene disorder, while a middle-aged patient with hypertension may in fact also have a monogenic disease. Examining the same pedigrees with different analytical approaches may thus allow us to better understand the true genetic basis of SVD. Arbitrary thresholds used in filtering large numbers of variants called may need to be adjusted in order not to be over-inclusive, or exclude true disease-causing mutations. The same pedigree may thus need to be explored under multiple models or hypotheses.

Despite the growing availability and falling costs of rapid sequencing approaches, our ability to make sense of the data produced has not kept up with the pace with which these technologies are becoming available to us in the clinic. Different approaches to phenotyping, sequencing, variant interpretation and validation, as well as the matching of phenotyped and genotyped cases from across the globe are needed in order for us to make further progress in our understanding of the genetic basis of SVD.

References

1. Sacco RL, Kasner SE, Broderick JP, Caplan LR, Connors JJB, Culebras A, et al. An updated definition of stroke for the 21st century: a statement for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke.* 2013;44:2064–89. doi: 10.1161/STR.0b013e318296aeca

2. Flossmann E, Schulz U, Rothwell P. Systematic review of methods and results of studies of the genetic epidemiology of ischemic stroke. *Stroke*. 2004;35:212–227. doi: 10.1161/01.STR.0000107187.84390.AA

3. Meschia JF. Ischaemic stroke: one or several complex genetic disorders? *Lancet. Neurol.* 2003;2:459.

4. Pantoni L. Cerebral Small Vessel Disease Https://books.google.com/books?id=z6lkAwAAQBAJ&pgis=1. Cambridge University Press; 2014.

5. Wardlaw JM, Smith EE, Biessels GJ, Cordonnier C, Fazekas F, Frayne R, et al. Neuroimaging standards for research into small vessel disease and its contribution to ageing and neurodegeneration. *Lancet Neurol.* 2013;12:822–838. doi: 10.1016/S1474-4422(13)70124-8

6. Rosenblum WI. Fibrinoid necrosis of small brain arteries and arterioles and miliary aneurysms as causes of hypertensive hemorrhage: a critical reappraisal. *Acta Neuropathol.* 2008;116:361–9. doi: 10.1007/s00401-008-0416-9

7. Markus HS, Lythgoe DJ, Ostegaard L, O'Sullivan M, Williams SC. Reduced cerebral blood flow in white matter in ischaemic leukoaraiosis demonstrated using quantitative exogenous contrast based perfusion MRI. *J. Neurol. Neurosurg. Psychiatry*. 2000;69:48–53.

8. Markus HS, Hunt B, Palmer K, Enzinger C, Schmidt H, Schmidt R. Markers of endothelial and hemostatic activation and progression of cerebral white matter hyperintensities: longitudinal results of the Austrian Stroke Prevention Study. *Stroke.* 2005;36:1410–4. doi: 10.1161/01.STR.0000169924.60783.d4

9. Wardlaw JM, Smith C, Dichgans M. Mechanisms of sporadic cerebral small vessel disease: insights from neuroimaging. *Lancet. Neurol.* 2013;12:483–97. doi: 10.1016/S1474-4422(13)70060-7

10. Wardlaw JM, Sandercock PAG, Dennis MS, Starr J. Is Breakdown of the Blood-Brain Barrier Responsible for Lacunar Stroke, Leukoaraiosis, and Dementia? *Stroke*. 2003;34.

11. Alafuzoff I, Adolfsson R, Grundke-Iqbal I, Winblad B. Perivascular deposits of serum proteins in cerebral cortex in vascular dementia. *Acta Neuropathol.* 1985;66:292–98.

12. Poggesi A, Pasi M, Pescini F, Pantoni L, Inzitari D. Circulating biologic markers of endothelial dysfunction in cerebral small vessel disease: A review. *J. Cereb. Blood Flow Metab.* 2016;36:72–94. doi: 10.1038/jcbfm.2015.116

13. Strand T, Alling C, Karlsson B, Karlsson I, Winblad B. Brain and plasma proteins in spinal fluid as markers for brain damage and severity of stroke. *Stroke.* 15:138–44.

14. Heye AK, Culling RD, Valdés Hernández MDC, Thrippleton MJ, Wardlaw JM. Assessment of blood-brain barrier disruption using dynamic contrast-enhanced MRI. A systematic review. *NeuroImage. Clin.* 2014;6:262–74. doi: 10.1016/j.nicl.2014.09.002

15. Revesz T, Holton JL, Lashley T, Plant G, Frangione B, Rostagno A, et al. Genetics and molecular pathogenesis of sporadic and hereditary cerebral amyloid angiopathies. *Acta Neuropathol.* 2009;118:115–30. doi: 10.1007/s00401-009-0501-8

16. Thanprasertsuk S, Martinez-Ramirez S, Pontes-Neto OM, Ni J, Ayres A, Reed A, et al. Posterior white matter disease distribution as a predictor of amyloid angiopathy. *Neurology*. 2014;83:794–800. doi: 10.1212/WNL.0000000000000732

17. Linn J, Halpin A, Demaerel P, Ruhland J, Giese AD, Dichgans M, et al. Prevalence of superficial siderosis in patients with cerebral amyloid angiopathy. *Neurology*. 2010;74:1346–50. doi: 10.1212/WNL.0b013e3181dad605

18. Rodrigues MA, Samarasekera N, Lerpiniere C, Humphreys C, McCarron MO, White PM, et al. The Edinburgh CT and genetic diagnostic criteria for lobar intracerebral haemorrhage associated with cerebral amyloid angiopathy: model development and diagnostic test accuracy study. *Lancet. Neurol.* 2018;0. doi: 10.1016/S1474-4422(18)30006-1

19. Rubattu S, Volpe M, Kreutz R, Ganten U, Ganten D, Lindpaintner K. Chromosomal mapping of quantitative trait loci contributing to stroke in a rat model of complex human disease. *Nat. Genet.* 1996;13:429–34. doi: 10.1038/ng0896-429

20. Bak S, Gaist D, Sindrup SH, Skytthe A, Christensen K. Genetic liability in stroke: a long-term follow-up study of Danish twins. *Stroke.* 2002;33:769–74.

21. Brass LM, Isaacsohn JL, Merikangas KR, Robinette CD. A study of twins and stroke. *Stroke.* 1992;23:221–3.

22. Kiely DK, Wolf P a, Cupples L a, Beiser a S, Myers RH. Familial aggregation of stroke. The Framingham Study. *Stroke.* 1993;24:1366–1371. doi: 10.1161/01.STR.24.9.1366

23. Bellenguez C, Bevan S, Gschwendtner A, Spencer CCA, Burgess AI, Pirinen M, et al. Genome-wide association study identifies a variant in HDAC9 associated with large vessel ischemic stroke. *Nat. Genet.* 2012;44:328–33. doi: 10.1038/ng.1081

24. Traylor M, Farrall M, Holliday EG, Sudlow C, Hopewell JC, Cheng Y-C, et al. Genetic risk factors for ischaemic stroke and its subtypes (the METASTROKE collaboration): a meta-analysis of genome-wide association studies. *Lancet. Neurol.* 2012;11:951–62. doi: 10.1016/S1474-4422(12)70234-X

25. Neurology Working Group of the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium, Stroke Genetics Network (SiGN), International Stroke Genetics Consortium (ISGC). Identification of additional risk loci for stroke and small vessel disease: a meta-analysis of genome-wide association studies. *Lancet. Neurol.* 2016;15:695–707. doi: 10.1016/S1474-4422(16)00102-2

26. Traylor M, Mäkelä K-M, Kilarski LL, Holliday EG, Devan WJ, Nalls MA, et al. A Novel MMP12 Locus Is Associated with Large Artery Atherosclerotic Stroke Using a Genome-Wide Age-at-Onset Informed Approach. *PLoS Genet.* 2014;10:e1004469. doi: 10.1371/journal.pgen.1004469

27. Woo D, Falcone GJ, Devan WJ, Brown WM, Biffi A, Howard TD, et al. Meta-analysis of Genome-wide Association Studies Identifies 1q22 as a Susceptibility Locus for Intracerebral Hemorrhage. *Am. J. Hum. Genet.* 2014;94:511–521. doi: 10.1016/j.ajhg.2014.02.012

28. Jerrard-Dunne P, Cloud G, Hassan A, Markus HS. Evaluating the genetic component of ischemic stroke subtypes: A family history study. *Stroke*. 2003;34:1364–1369. doi: 10.1161/01.STR.0000069723.17984.FD

29. Traylor M, Bevan S, Baron J-C, Hassan A, Lewis CM, Markus HS. Genetic Architecture of Lacunar Stroke. *Stroke.* 2015;46:2407–12. doi: 10.1161/STROKEAHA.115.009485

30. Biffi A, Greenberg SM. Cerebral amyloid angiopathy: a systematic review. *J. Clin. Neurol.* 2011;7:1–9. doi: 10.3988/jcn.2011.7.1.1

31. Biffi A, Sonni A, Anderson CD, Kissela B, Jagiella JM, Schmidt H, et al. Variants at APOE influence risk of deep and lobar intracerebral hemorrhage. *Ann. Neurol.* 2010;68:934–43. doi: 10.1002/ana.22134

32. Domenga V, Fardoux P, Lacombe P, Monet M, Maciazek J, Krebs LT, et al. Notch3 is required for arterial identity and maturation of vascular smooth muscle cells. *Genes Dev.* 2004;18:2730–5. doi: 10.1101/gad.308904

33. Oka C, Tsujimoto R, Kajikawa M, Koshiba-Takeuchi K, Ina J, Yano M, et al. HtrA1 serine protease inhibits signaling mediated by Tgfbeta family proteins. *Development*. 2004;131:1041–53. doi: 10.1242/dev.00999

34. Caciotti A, Catarzi S, Tonin R, Lugli L, Perez C, Michelakakis H, et al. Galactosialidosis: review and analysis of CTSA gene mutations. *Orphanet J. Rare Dis.* 2013;8:114. doi: 10.1186/1750-1172-8-114

35. Rhodes JM, Simons M. The extracellular matrix and blood vessel formation: not just a scaffold. *J. Cell. Mol. Med.* 2007;11:176–205. doi: 10.1111/j.1582-4934.2007.00031.x

36. Kavanagh D, Spitzer D, Kothari PH, Shaikh A, Liszewski MK, Richards A, et al. New roles for the major human 3’-5’ exonuclease TREX1 in human disease. *Cell Cycle*. 2008;7:1718–25.

37. Seo S, Singh HP, Lacal PM, Sasman A, Fatima A, Liu T, et al. Forkhead box transcription factor FoxC1 preserves corneal transparency by regulating vascular growth. *Proc. Natl. Acad. Sci. U. S. A.* 2012;109:2015–20. doi: 10.1073/pnas.1109540109

38. Belmonte JCI, Ryan AK, Blumberg B, Rodriguez-Esteban C, Yonei-Tamura S, Tamura K, et al. Pitx2 determines left-right asymmetry of internal organs in vertebrates. *Nature*. 1998;394:545–551. doi: 10.1038/29004

39. Joutel A, Corpechot C, Ducros A, Vahedi K, Chabriat H, Mouton P, et al. Notch3 mutations in CADASIL, a hereditary adult-onset condition causing stroke and

dementia. *Nature*. 1996;383:707–10. doi: 10.1038/383707a0

40. Razvi SSM, Davidson R, Bone I, Muir KW. The prevalence of cerebral autosomal dominant arteriopathy with subcortical infarcts and leucoencephalopathy (CADASIL) in the west of Scotland. *J. Neurol. Neurosurg. Psychiatry*. 2005;76:739–41. doi: 10.1136/jnnp.2004.051847

41. Narayan SK, Gorman G, Kalaria RN, Ford GA, Chinnery PF. The minimum prevalence of CADASIL in northeast England. *Neurology*. 2012;78:1025–7. doi: 10.1212/WNL.0b013e31824d586c

42. Kilarski LL, Rutten-Jacobs LCA, Bevan S, Baker R, Hassan A, Hughes DA, et al. Prevalence of CADASIL and Fabry Disease in a Cohort of MRI Defined Younger Onset Lacunar Stroke. *PLoS One*. 2015;10:e0136352. doi: 10.1371/journal.pone.0136352

43. Adib-Samii P, Brice G, Martin RJ, Markus HS. Clinical spectrum of CADASIL and the effect of cardiovascular risk factors on phenotype: study in 200 consecutively recruited individuals. *Stroke.* 2010;41:630–4. doi: 10.1161/STROKEAHA.109.568402

44. Dichgans M, Mayer M, Uttner I, Brüning R, Müller-Höcker J, Rungger G, et al. The phenotypic spectrum of CADASIL: clinical findings in 102 cases. *Ann. Neurol.* 1998;44:731–9. doi: 10.1002/ana.410440506

45. Desmond DW, Moroney JT, Lynch T, Chan S, Chin SS, Mohr JP. The natural history of CADASIL: a pooled analysis of previously published cases. *Stroke.* 1999;30:1230–3.

46. Roine S, Pöyhönen M, Timonen S, Tuisku S, Marttila R, Sulkava R, et al. Neurologic symptoms are common during gestation and puerperium in CADASIL. *Neurology*. 2005;64:1441–3. doi: 10.1212/01.WNL.0000158655.71323.8A

47. Hinze S, Goonasekera M, Nannucci S, Quaghebeur G, Briley D, Markus HS, et al. Longitudinally extensive spinal cord infarction in CADASIL. *Pract. Neurol.* 2015;15:60–2. doi: 10.1136/practneurol-2014-000870

48. Robinson W, Galetta SL, McCluskey L, Forman MS, Balcer LJ. Retinal findings in cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (cadasil). *Surv. Ophthalmol.* 45:445–8.

49. Roine S, Harju M, Kivelä TT, Pöyhönen M, Nikoskelainen E, Tuisku S, et al. Ophthalmologic findings in cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy: a cross-sectional study. *Ophthalmology*. 2006;113:1411–7. doi: 10.1016/j.ophtha.2006.03.030

50. Haritoglou C, Hoops JP, Stefani FH, Mehraein P, Kampik A, Dichgans M. Histopathological abnormalities in ocular blood vessels of CADASIL patients. *Am. J. Ophthalmol.* 2004;138:302–5. doi: 10.1016/j.ajo.2004.02.073

51. Tournier-Lasserve E, Joutel A, Melki J, Weissenbach J, Lathrop GM, Chabriat H, et al. Cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy maps to chromosome 19q12. *Nat. Genet.* 1993;3:256–9. doi: 10.1038/ng0393-256

52. O'Sullivan M, Jarosz JM, Martin RJ, Deasy N, Powell JF, Markus HS. MRI

hyperintensities of the temporal lobe and external capsule in patients with CADASIL. *Neurology*. 2001;56:628–34.

53. Lesnik Oberstein SA, van den Boom R, van Buchem MA, van Houwelingen HC, Bakker E, Vollebregt E, et al. Cerebral microbleeds in CADASIL. *Neurology*. 2001;57:1066–70.

54. Guerrot D, François A, Boffa J-J, Boulos N, Hanoy M, Legallicier B, et al. Nephroangiosclerosis in Cerebral Autosomal Dominant Arteriopathy With Subcortical Infarcts and Leukoencephalopathy: Is NOTCH3 Mutation the Common Culprit? *Am. J. Kidney Dis.* 2008;52:340–345. doi: 10.1053/j.ajkd.2008.04.017

55. Kusaba T, Hatta T, Kimura T, Sonomura K, Tanda S, Kishimoto N, et al. Renal involvement in cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL). *Clin. Nephrol.* 2007;67:182–7.

56. Morroni M, Marzioni D, Ragno M, Di Bella P, Cartechini E, Pianese L, et al. Role of electron microscopy in the diagnosis of cadasil syndrome: a study of 32 patients. *PLoS One*. 2013;8:e65482. doi: 10.1371/journal.pone.0065482

57. Joutel A, Vahedi K, Corpechot C, Troesch A, Chabriat H, Vayssière C, et al. Strong clustering and stereotyped nature of Notch3 mutations in CADASIL patients. *Lancet*. 1997;350:1511–5. doi: 10.1016/S0140-6736(97)08083-5

58. Singhal S, Bevan S, Barrick T, Rich P, Markus HS. The influence of genetic and cardiovascular risk factors on the CADASIL phenotype. *Brain*. 2004;127:2031–8. doi: 10.1093/brain/awh223

59. Rutten JW, Dauwerse HG, Gravesteijn G, van Belzen MJ, van der Grond J, Polke JM, et al. Archetypal NOTCH3 mutations frequent in public exome: implications for CADASIL. *Ann. Clin. Transl. Neurol.* 2016;3:844–853. doi: 10.1002/acn3.344

60. Opherk C, Peters N, Holtmannspotter M, Gschwendtner A, Muller-Myhsok B, Dichgans M. Heritability of MRI Lesion Volume in CADASIL: Evidence for Genetic Modifiers. *Stroke*. 2006;37:2684–2689. doi: 10.1161/01.STR.0000245084.35575.66

61. Hara K, Shiga A, Fukutake T, Nozaki H, Miyashita A, Yokoseki A, et al. Association of HTRA1 mutations and familial ischemic cerebral small-vessel disease. *N. Engl. J. Med.* 2009;360:1729–39. doi: 10.1056/NEJMoa0801560

62. Fukutake T. Cerebral autosomal recessive arteriopathy with subcortical infarcts and leukoencephalopathy (CARASIL): from discovery to gene identification. *J. Stroke Cerebrovasc. Dis.* 2011;20:85–93. doi: 10.1016/j.jstrokecerebrovasdis.2010.11.008

63. Yanagawa S, Ito N, Arima K, Ikeda S -i. S. Cerebral autosomal recessive arteriopathy with subcortical infarcts and leukoencephalopathy. *Neurology*. 2002;58:817–820. doi: 10.1212/WNL.58.5.817

64. Zheng DM, Xu FF, Gao Y, Zhang H, Han SC, Bi GR. A Chinese pedigree of cerebral autosomal recessive arteriopathy with subcortical infarcts and leukoencephalopathy (CARASIL): clinical and radiological features. *J. Clin. Neurosci.* 2009;16:847–9. doi: 10.1016/j.jocn.2008.08.031

65. Mendioroz M, Fernández-Cadenas I, Del Río-Espinola A, Rovira A, Solé E, Fernández-

Figueras MT, et al. A missense HTRA1 mutation expands CARASIL syndrome to the Caucasian population. *Neurology*. 2010;75:2033–5. doi: 10.1212/WNL.0b013e3181ff96ac

66. Menezes Cordeiro I, Nzwalo H, Sá F, Ferreira RB, Alonso I, Afonso L, et al. Shifting the CARASIL Paradigm: Report of a Non-Asian Family and Literature Review. *Stroke.* 2015;46:1110–2. doi: 10.1161/STROKEAHA.114.006735

67. Verdura E, Hervé D, Scharrer E, Amador MDM, Guyant-Maréchal L, Philippi A, et al. Heterozygous HTRA1 mutations are associated with autosomal dominant cerebral small vessel disease. *Brain*. 2015;awv155-. doi: 10.1093/brain/awv155

68. Nozaki H, Sekine Y, Fukutake T, Nishimoto Y, Shimoe Y, Shirata A, et al. Characteristic features and progression of abnormalities on MRI for CARASIL. *Neurology*. 2015;85:459–63. doi: 10.1212/WNL.0000000000001803

69. Fukutake T, Hirayama K. Familial young-adult-onset arteriosclerotic leukoencephalopathy with alopecia and lumbago without arterial hypertension. *Eur. Neurol.* 1995;35:69–79.

70. Arima K, Yanagawa S, Ito N, Ikeda S. Cerebral arterial pathology of CADASIL and CARASIL (Maeda syndrome). *Neuropathology*. 2003;23:327–34.

71. Oide T, Nakayama H, Yanagawa S, Ito N, Ikeda S-I, Arima K. Extensive loss of arterial medial smooth muscle cells and mural extracellular matrix in cerebral autosomal recessive arteriopathy with subcortical infarcts and leukoencephalopathy (CARASIL). *Neuropathology*. 2008;28:132–42. doi: 10.1111/j.1440-1789.2007.00864.x

72. Yamamura T, Nishimura M, Shirabe T, Fujita M. Subcortical vascular encephalopathy in a normotensive, young adult with premature baldness and spondylitis deformans. A clinicopathological study and review of the literature. *J. Neurol. Sci.* 1987;78:175–87.

73. Richards A, van den Maagdenberg AMJM, Jen JC, Kavanagh D, Bertram P, Spitzer D, et al. C-terminal truncations in human 3’-5’ DNA exonuclease TREX1 cause autosomal dominant retinal vasculopathy with cerebral leukodystrophy. *Nat. Genet.* 2007;39:1068–70. doi: 10.1038/ng2082

74. Grand MG, Kaine J, Fulling K, Atkinson J, Dowton SB, Farber M, et al. Cerebroretinal vasculopathy. A new hereditary syndrome. *Ophthalmology*. 1988;95:649–59.

75. Ophoff RA, DeYoung J, Service SK, Joosse M, Caffo NA, Sandkuijl LA, et al. Hereditary vascular retinopathy, cerebroretinal vasculopathy, and hereditary endotheliopathy with retinopathy, nephropathy, and stroke map to a single locus on chromosome 3p21.1-p21.3. *Am. J. Hum. Genet.* 2001;69:447–53. doi: 10.1086/321975

76. Jen J, Cohen AH, Yue Q, Stout JT, Vinters H V, Nelson S, et al. Hereditary endotheliopathy with retinopathy, nephropathy, and stroke (HERNS). *Neurology*. 1997;49:1322–30.

77. Terwindt GM, Haan J, Ophoff RA, Groenen SM, Storimans CW, Lanser JB, et al. Clinical and genetic analysis of a large Dutch family with autosomal dominant vascular retinopathy, migraine and Raynaud’s phenomenon. *Brain*. 1998;121 (Pt 2:303–16.

78. DiFrancesco JC, Novara F, Zuffardi O, Forlino A, Gioia R, Cossu F, et al. TREX1 C-terminal frameshift mutations in the systemic variant of retinal vasculopathy with cerebral leukodystrophy. *Neurol. Sci.* 2014;doi: 10.1007/s10072-014-1944-9

79. Winkler DT, Lyrer P, Probst A, Devys D, Haufschild T, Haller S, et al. Hereditary systemic angiopathy (HSA) with cerebral calcifications, retinopathy, progressive nephropathy, and hepatopathy. *J. Neurol.* 2008;255:77–88. doi: 10.1007/s00415-008-0675-3

80. Pelzer N, Vries B, Boon EMJ, Kruit MC, Haan J, Ferrari MD, et al. Heterozygous TREX1 mutations in early-onset cerebrovascular disease. *J. Neurol.* 2013;260:2188–2190. doi: 10.1007/s00415-013-7050-8

81. Vahedi K, Alamowitch S. Clinical spectrum of type IV collagen (COL4A1) mutations: a novel genetic multisystem disease. *Curr. Opin. Neurol.* 2011;24:63–8. doi: 10.1097/WCO.0b013e32834232c6

82. Lanfranconi S, Markus HS. COL4A1 mutations as a monogenic cause of cerebral small vessel disease: a systematic review. *Stroke.* 2010;41:e513-8. doi: 10.1161/STROKEAHA.110.581918

83. Gould DB, Phalan FC, van Mil SE, Sundberg JP, Vahedi K, Massin P, et al. Role of COL4A1 in small-vessel disease and hemorrhagic stroke. *N. Engl. J. Med.* 2006;354:1489–96. doi: 10.1056/NEJMoa053727

84. Vahedi K, Boukobza M, Massin P, Gould DB, Tournier-Lasserve E, Bousser M-G. Clinical and brain MRI follow-up study of a family with COL4A1 mutation. *Neurology*. 2007;69:1564–8. doi: 10.1212/01.wnl.0000295994.46586.e7

85. Alamowitch S, Plaisier E, Favrole P, Prost C, Chen Z, Van Agtmael T, et al. Cerebrovascular disease related to COL4A1 mutations in HANAC syndrome. *Neurology*. 2009;73:1873–82. doi: 10.1212/WNL.0b013e3181c3fd12

86. Verbeek E, Meuwissen MEC, Verheijen FW, Govaert PP, Licht DJ, Kuo DS, et al. COL4A2 mutation associated with familial porencephaly and small-vessel disease. *Eur. J. Hum. Genet.* 2012;20:844–51. doi: 10.1038/ejhg.2012.20

87. Renard D, Miné M, Pipiras E, Labauge P, Delahaye A, Benzacken B, et al. Cerebral small-vessel disease associated with COL4A1 and COL4A2 gene duplications. *Neurology*. 2014;83:1029–31. doi: 10.1212/WNL.0000000000000769

88. Tümer Z, Bach-Holm D. Axenfeld-Rieger syndrome and spectrum of PITX2 and FOXC1 mutations. *Eur. J. Hum. Genet.* 2009;17:1527–39. doi: 10.1038/ejhg.2009.93

89. Delahaye A, Khung-Savatovsky S, Aboura A, Guimiot F, Drunat S, Alessandri J-L, et al. Pre- and postnatal phenotype of 6p25 deletions involving the FOXC1 gene. *Am. J. Med. Genet. A*. 2012;158A:2430–8. doi: 10.1002/ajmg.a.35548

90. Cellini E, Disciglio V, Novara F, Barkovich JA, Mencarelli MA, Hayek J, et al. Periventricular heterotopia with white matter abnormalities associated with 6p25 deletion. *Am. J. Med. Genet. A*. 2012;158A:1793–7. doi: 10.1002/ajmg.a.35416

91. French CR, Seshadri S, Destefano AL, Fornage M, Arnold CR, Gage PJ, et al. Mutation

of FOXC1 and PITX2 induces cerebral small-vessel disease. *J. Clin. Invest.* 2014;124:4877–81. doi: 10.1172/JCI75109

92. Zhou Q, Yang D, Ombrello AK, Zavialov A V., Toro C, Zavialov A V., et al. Early-Onset Stroke and Vasculopathy Associated with Mutations in ADA2. *N. Engl. J. Med.* 2014;370:911–920. doi: 10.1056/NEJMoa1307361

93. Mutant ADA2 in Vasculopathies. *N. Engl. J. Med.* 2014;371:478–481. doi: 10.1056/NEJMc1405506

94. Nagae-Poetscher LM, Bibat G, Philippart M, Rosemberg S, Fatemi A, Lacerda MTC, et al. Leukoencephalopathy, cerebral calcifications, and cysts: new observations. *Neurology*. 2004;62:1206–9. doi: 10.1212/01.WNL.0000119341.59445.CF

95. Corboy JR, Gault J, Kleinschmidt-DeMasters BK. An adult case of leukoencephalopathy with intracranial calcifications and cysts. *Neurology*. 2006;67:1890–2. doi: 10.1212/01.wnl.0000244470.06748.18

96. Novo J, Lin D, Shanks M, Kocak M, Arvanitis L. A 55-year-old female with leukoencephalopathy with cerebral calcifications and cysts: Case report and radiopathologic description. *Pathol. - Res. Pract.* 2017;doi: 10.1016/j.prp.2017.07.007

97. Sener U, Zorlu Y, Men S, Bayol U, Zanapalioglu U. Leukoencephalopathy, Cerebral Calcifications, and Cysts. *Am. J. Neuroradiol.* 2006;27.

98. Linnankivi T, Valanne L, Paetau A, Alafuzoff I, Hakumäki JM, Kivelä T, et al. Cerebroretinal microangiopathy with calcifications and cysts. *Neurology*. 2006;67:1437–43. doi: 10.1212/01.wnl.0000236999.63933.b0

99. Livingston J, Mayer J, Jenkinson E, Kasher P, Stivaros S, Berger A, et al. Leukoencephalopathy with Calcifications and Cysts: A Purely Neurological Disorder Distinct from Coats Plus. *Neuropediatrics*. 2014;45:175–182. doi: 10.1055/s-0033-1364180

100. Anderson BH, Kasher PR, Mayer J, Szynkiewicz M, Jenkinson EM, Bhaskar SS, et al. Mutations in CTC1, encoding conserved telomere maintenance component 1, cause Coats plus. *Nat. Genet.* 2012;44:338–342. doi: 10.1038/ng.1084

101. Di Fede G, Giaccone G, Tagliavini F. Hereditary and sporadic beta-amyloidoses. *Front. Biosci. (Landmark Ed.* 2013;18:1202–26.

102. Dermaut B, Kumar-Singh S, De Jonghe C, Cruts M, Löfgren A, Lübke U, et al. Cerebral amyloid angiopathy is a pathogenic lesion in Alzheimer’s disease due to a novel presenilin 1 mutation. *Brain*. 2001;124:2383–2392. doi: 10.1093/brain/124.12.2383

103. Bornebroek M, Haan J, Maat-Schieman ML, Van Duinen SG, Roos RA. Hereditary cerebral hemorrhage with amyloidosis-Dutch type (HCHWA-D): I--A review of clinical, radiologic and genetic aspects. *Brain Pathol.* 1996;6:111–4.

104. Hendriks L, van Duijn CM, Cras P, Cruts M, Van Hul W, van Harskamp F, et al. Presenile dementia and cerebral haemorrhage linked to a mutation at codon 692 of the β–amyloid precursor protein gene. *Nat. Genet.* 1992;1:218–221. doi: 10.1038/ng0692-218

105. Bugiani O, Giaccone G, Rossi G, Mangieri M, Capobianco R, Morbin M, et al. Hereditary Cerebral Hemorrhage With Amyloidosis Associated With the E693K Mutation of APP. *Arch. Neurol.* 2010;67:987–95. doi: 10.1001/archneurol.2010.178

106. Obici L, Demarchi A, de Rosa G, Bellotti V, Marciano S, Donadei S, et al. A novelAβPP mutation exclusively associated with cerebral amyloid angiopathy. *Ann. Neurol.* 2005;58:639–644. doi: 10.1002/ana.20571

107. Grabowski TJ, Cho HS, Vonsattel JP, Rebeck GW, Greenberg SM. Novel amyloid precursor protein mutation in an Iowa family with dementia and severe cerebral amyloid angiopathy. *Ann. Neurol.* 2001;49:697–705.

108. Kamino K, Orr HT, Payami H, Wijsman EM, Alonso ME, Pulst SM, et al. Linkage and mutational analysis of familial Alzheimer disease kindreds for the APP gene region. *Am. J. Hum. Genet.* 1992;51:998–1014.

109. Jensson O, Palsdottir A, Thorsteinsson L, Arnason A. The saga of cystatin C gene mutation causing amyloid angiopathy and brain hemorrhage--clinical genetics in Iceland. *Clin. Genet.* 1989;36:368–77.

110. Vidal R, Frangione B, Rostagno A, Mead S, Révész T, Plant G, et al. A stop-codon mutation in the BRI gene associated with familial British dementia. *Nature*. 1999;399:776–81. doi: 10.1038/21637

111. Vidal R, Revesz T, Rostagno A, Kim E, Holton JL, Bek T, et al. A decamer duplication in the 3' region of the BRI gene originates an amyloid peptide that is associated with dementia in a Danish kindred. *Proc. Natl. Acad. Sci. U. S. A.* 2000;97:4920–5. doi: 10.1073/pnas.080076097

112. Sekijima Y, Yoshida K, Tokuda T, Ikeda S. Familial Transthyretin Amyloidosis Http://www.ncbi.nlm.nih.gov/pubmed/20301373. University of Washington, Seattle; 1993.

113. Kiuru-Enari S, Haltia M. Hereditary gelsolin amyloidosis Http://www.ncbi.nlm.nih.gov/pubmed/23931809. In: Handbook of clinical neurology. 2013. p. 659–681.doi: 10.1016/B978-0-444-52902-2.00039-4

114. Watanabe R, Duchen LW. Cerebral amyloid in human prion disease. *Neuropathol. Appl. Neurobiol.* 1993;19:253–60.

115. Frantz C, Stewart KM, Weaver VM. The extracellular matrix at a glance. *J. Cell Sci.* 2010;123:4195–200. doi: 10.1242/jcs.023820

116. Naba A, Clauser KR, Hoersch S, Liu H, Carr SA, Hynes RO. The matrisome: in silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices. *Mol. Cell. Proteomics*. 2012;11:M111.014647. doi: 10.1074/mcp.M111.014647

117. Xu J, Shi G-P. Vascular wall extracellular matrix proteins and vascular diseases. *Biochim. Biophys. Acta*. 2014;1842:2106–2119. doi: 10.1016/j.bbadis.2014.07.008

118. Joutel A, Haddad I, Ratelade J, Nelson MT. Perturbations of the cerebrovascular matrisome: A convergent mechanism in small vessel disease of the brain? *J. Cereb.*

Blood Flow Metab. 2016;36:143–57. doi: 10.1038/jcbfm.2015.62

119. Monet-Leprêtre M, Haddad I, Baron-Menguy C, Fouillot-Panchal M, Riani M, Domenga-Denier V, et al. Abnormal recruitment of extracellular matrix proteins by excess Notch3 ECD: a new pathomechanism in CADASIL. *Brain*. 2013;136:1830–45. doi: 10.1093/brain/awt092

120. Dabertrand F, Krøigaard C, Bonev AD, Cognat E, Dalsgaard T, Domenga-Denier V, et al. Potassium channelopathy-like defect underlies early-stage cerebrovascular dysfunction in a genetic model of small vessel disease. *Proc. Natl. Acad. Sci. U. S. A.* 2015;112:E796-805. doi: 10.1073/pnas.1420765112

121. Capone C, Dabertrand F, Baron-Menguy C, Chalaris A, Ghezali L, Domenga-Denier V, et al. Mechanistic insights into a TIMP3-sensitive pathway constitutively engaged in the regulation of cerebral hemodynamics. *Elife*. 2016;5. doi: 10.7554/eLife.17536

122. Capone C, Cognat E, Ghezali L, Baron-Menguy C, Aubin D, Mesnard L, et al. Reducing Timp3 or vitronectin ameliorates disease manifestations in CADASIL mice. *Ann. Neurol.* 2016;79:387–403. doi: 10.1002/ana.24573

123. Walsh JS, Perniciaro C, Meschia JF. CADASIL (cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy): diagnostic skin biopsy changes determined by electron microscopy. *J. Am. Acad. Dermatol.* 2000;43:1125–7. doi: 10.1067/mjd.2000.110895

124. Joutel A, Andreux F, Gaulis S, Domenga V, Cecillon M, Battail N, et al. The ectodomain of the Notch3 receptor accumulates within the cerebrovasculature of CADASIL patients. *J. Clin. Invest.* 2000;105:597–605. doi: 10.1172/JCI8047

125. Monet M, Domenga V, Lemaire B, Souilhol C, Langa F, Babinet C, et al. The archetypal R90C CADASIL-NOTCH3 mutation retains NOTCH3 function in vivo. *Hum. Mol. Genet.* 2007;16:982–992. doi: 10.1093/hmg/ddm042

126. Dabertrand F, Krøigaard C, Bonev AD, Cognat E, Dalsgaard T, Domenga-Denier V, et al. Potassium channelopathy-like defect underlies early-stage cerebrovascular dysfunction in a genetic model of small vessel disease. *Proc. Natl. Acad. Sci. U. S. A.* 2015;112:E796-805. doi: 10.1073/pnas.1420765112

127. Joutel A, Monet-Leprêtre M, Gosele C, Baron-Menguy C, Hammes A, Schmidt S, et al. Cerebrovascular dysfunction and microcirculation rarefaction precede white matter lesions in a mouse genetic model of cerebral ischemic small vessel disease. *J. Clin. Invest.* 2010;120:433–45. doi: 10.1172/JCI39733

128. Kast J, Hanecker P, Beaufort N, Giese A, Joutel A, Dichgans M, et al. Sequestration of latent TGF-β binding protein 1 into CADASIL-related Notch3-ECD deposits. *Acta Neuropathol. Commun.* 2014;2:96. doi: 10.1186/s40478-014-0096-8

129. Liem MK, Lesnik Oberstein SAJ, Haan J, Boom R v d, Ferrari MD, Buchem MA v, et al. Cerebrovascular reactivity is a main determinant of white matter hyperintensity progression in CADASIL. *AJNR. Am. J. Neuroradiol.* 2009;30:1244–7. doi: 10.3174/ajnr.A1533

130. Lacombe P, Oligo C, Domenga V, Tournier-Lasserve E, Joutel A. Impaired cerebral

vasoreactivity in a transgenic mouse model of cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy arteriopathy. *Stroke.* 2005;36:1053–8. doi: 10.1161/01.STR.0000163080.82766.eb

131. Meng H, Zhang X, Hankenson KD, Wang MM. Thrombospondin 2 potentiates notch3/jagged1 signaling. *J. Biol. Chem.* 2009;284:7866–74. doi: 10.1074/jbc.M803650200

132. Bornstein P, Kyriakides TR, Yang Z, Armstrong LC, Birk DE. Thrombospondin 2 modulates collagen fibrillogenesis and angiogenesis. *J. Investig. Dermatol. Symp. Proc.* 2000;5:61–6. doi: 10.1046/j.1087-0024.2000.00005.x

133. Calabro NE, Kristofik NJ, Kyriakides TR. Thrombospondin-2 and extracellular matrix assembly. *Biochim. Biophys. Acta*. 2014;1840:2396–402. doi: 10.1016/j.bbagen.2014.01.013

134. Duering M, Karpinska A, Rosner S, Hopfner F, Zechmeister M, Peters N, et al. Co-aggregate formation of CADASIL-mutant NOTCH3: a single-particle analysis. *Hum. Mol. Genet.* 2011;20:3256–65. doi: 10.1093/hmg/ddr237

135. Hyytiäinen M, Penttinen C, Keski-Oja J. Latent TGF-beta binding proteins: extracellular matrix association and roles in TGF-beta activation. *Crit. Rev. Clin. Lab. Sci.* 2004;41:233–64. doi: 10.1080/10408360490460933

136. Beaufort N, Scharrer E, Kremmer E, Lux V, Ehrmann M, Huber R, et al. Cerebral small vessel disease-related protease HtrA1 processes latent TGF-β binding protein 1 and facilitates TGF-β signaling. *Proc. Natl. Acad. Sci. U. S. A.* 2014;111:16496–501. doi: 10.1073/pnas.1418087111

137. Qi JH, Ebrahem Q, Moore N, Murphy G, Claesson-Welsh L, Bond M, et al. A novel function for tissue inhibitor of metalloproteinases-3 (TIMP3): inhibition of angiogenesis by blockage of VEGF binding to VEGF receptor-2. *Nat. Med.* 2003;9:407–415. doi: 10.1038/nm846

138. Preissner KT, Reuning U. Vitronectin in vascular context: facets of a multitalented matricellular protein. *Semin. Thromb. Hemost.* 2011;37:408–24. doi: 10.1055/s-0031-1276590

139. Edwards DR, Handsley MM, Pennington CJ. The ADAM metalloproteinases. *Mol. Aspects Med.* 2008;29:258–89. doi: 10.1016/j.mam.2008.08.001

140. Koide M, Penar PL, Tranmer BI, Wellman GC. Heparin-binding EGF-like growth factor mediates oxyhemoglobin-induced suppression of voltage-dependent potassium channels in rabbit cerebral artery myocytes. *Am. J. Physiol. Heart Circ. Physiol.* 2007;293:H1750-9. doi: 10.1152/ajpheart.00443.2007

141. Jucker M, Walker LC. Self-propagation of pathogenic protein aggregates in neurodegenerative diseases. *Nature*. 2013;501:45–51. doi: 10.1038/nature12481

142. Prusiner SB. Biology and genetics of prions causing neurodegeneration. *Annu. Rev. Genet.* 2013;47:601–23. doi: 10.1146/annurev-genet-110711-155524

143. Schmidt H, Zeginigg M, Wiltgen M, Freudenberger P, Petrovic K, Cavalieri M, et al.

Genetic variants of the NOTCH3 gene in the elderly and magnetic resonance imaging correlates of age-related cerebral small vessel disease. *Brain*. 2011;134:3384–97. doi: 10.1093/brain/awr252

144. Rutten-Jacobs LCA, Traylor M, Adib-Samii P, Thijs V, Sudlow C, Rothwell PM, et al. Common NOTCH3 Variants and Cerebral Small-Vessel Disease. *Stroke.* 2015;46:1482–1487. doi: 10.1161/STROKEAHA.114.008540

145. Malik R, Chauhan G, Traylor M, Sargurupremraj M, Okada Y, Mishra A, et al. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat. Genet.* 2018;doi: 10.1038/s41588-018-0058-3

146. Chong M, O'Donnell M, Thijs V, Dans A, López-Jaramillo P, Gómez-Arbeláez D, et al. Mendelian Genes and Risk of Intracerebral Hemorrhage and Small-Vessel Ischemic Stroke in Sporadic Cases. *Stroke*. 2017;48.

147. Ross OA, Soto-Ortolaza AI, Heckman MG, Verbeeck C, Serie DJ, Rayaprolu S, et al. NOTCH3 variants and risk of ischemic stroke. *PLoS One*. 2013;8:e75035. doi: 10.1371/journal.pone.0075035

148. Dong Y, Hassan A, Zhang Z, Huber D, Dalageorgou C, Markus HS, et al. Yield of screening for CADASIL mutations in lacunar stroke and leukoaraiosis. *Stroke*. 2002;34:203–5. doi: 10.1161/01.STR.0000048162.16852.88

149. Ito D. Notch3 gene polymorphism and ischaemic cerebrovascular disease. *J. Neurol. Neurosurg. Psychiatry*. 2002;72:382–384. doi: 10.1136/jnnp.72.3.382

150. Wang T, Sharma SD, Fox N, Rossor M, Brown MJ, Sharma P. Description of a simple test for CADASIL disease and determination of mutation frequencies in sporadic ischaemic stroke and dementia patients. *J. Neurol. Neurosurg. Psychiatry*. 2000;69:652–4.

151. Kilarski LL, Rutten-Jacobs LCA, Bevan S, Baker R, Hassan A, Hughes DA, et al. Prevalence of CADASIL and Fabry Disease in a Cohort of MRI Defined Younger Onset Lacunar Stroke. *PLoS One*. 2015;10:e0136352. doi: 10.1371/journal.pone.0136352

152. Rolfs A, Böttcher T, Zschiesche M, Morris P, Winchester B, Bauer P, et al. Prevalence of Fabry disease in patients with cryptogenic stroke: a prospective study. *Lancet*. 2005;366:1794–6. doi: 10.1016/S0140-6736(05)67635-0

153. Baptista MV, Ferreira S, Pinho-E-Melo T, Carvalho M, Cruz VT, Carmona C, et al. Mutations of the GLA gene in young patients with stroke: the PORTYSTROKE study--screening genetic conditions in Portuguese young stroke patients. *Stroke.* 2010;41:431–6. doi: 10.1161/STROKEAHA.109.570499

154. Brouns R, Thijs V, Eyskens F, Van den Broeck M, Belachew S, Van Broeckhoven C, et al. Belgian Fabry Study: Prevalence of Fabry Disease in a Cohort of 1000 Young Patients With Cerebrovascular Disease. *Stroke*. 2010;41:863–868. doi: 10.1161/STROKEAHA.110.579409

155. Nakamura K, Sekijima Y, Nakamura K, Hattori K, Nagamatsu K, Shimizu Y, et al. p.E66Q mutation in the *GLA* gene is associated with a high risk of cerebral small-vessel

occlusion in elderly Japanese males. *Eur. J. Neurol.* 2014;21:49–56. doi: 10.1111/ene.12214

156. Weng Y-C, Sonni A, Labelle-Dumais C, de Leau M, Kauffman WB, Jeanne M, et al. COL4A1 mutations in patients with sporadic late-onset intracerebral hemorrhage. *Ann. Neurol.* 2012;71:470–7. doi: 10.1002/ana.22682

157. Rannikmäe K, Davies G, Thomson PA, Bevan S, Devan WJ, Falcone GJ, et al. Common variation in COL4A1/COL4A2 is associated with sporadic cerebral small vessel disease. *Neurology*. 2015;doi: 10.1212/WNL.0000000000001309

158. Traylor M, Zhang CR, Adib-Samii P, Devan WJ, Parsons OE, Lanfranconi S, et al. Genome-wide meta-analysis of cerebral white matter hyperintensities in patients with stroke. *Neurology*. 2016;86:146–53. doi: 10.1212/WNL.0000000000002263

159. Jeanne M, Labelle-Dumais C, Jorgensen J, Kauffman WB, Mancini GM, Favor J, et al. COL4A2 mutations impair COL4A1 and COL4A2 secretion and cause hemorrhagic stroke. *Am. J. Hum. Genet.* 2012;90:91–101. doi: 10.1016/j.ajhg.2011.11.022

160. Pelzer N, de Vries B, Boon EMJ, Kruit MC, Haan J, Ferrari MD, et al. Heterozygous TREX1 mutations in early-onset cerebrovascular disease. *J. Neurol.* 2013;260:2188–90. doi: 10.1007/s00415-013-7050-8

161. Johnston S, Mendis S, Mathers C, Falcone G, Malik R, Dichgans M, et al. Identification of additional risk loci for stroke and small vessel disease: a meta-analysis of genome-wide association studies. *Lancet Neurol.* 2016;15:695–707. doi: 10.1016/S1474-4422(16)00102-2

162. Breedveld G, de Coo IF, Lequin MH, Arts WFM, Heutink P, Gould DB, et al. Novel mutations in three families confirm a major role of COL4A1 in hereditary porencephaly. *J. Med. Genet.* 2006;43:490–5. doi: 10.1136/jmg.2005.035584

163. Gould DB, Marchant JK, Savinova O V, Smith RS, John SWM. Col4a1 mutation causes endoplasmic reticulum stress and genetically modifiable ocular dysgenesis. *Hum. Mol. Genet.* 2007;16:798–807. doi: 10.1093/hmg/ddm024

164. Murray LS, Lu Y, Taggart A, Van Regemorter N, Vilain C, Abramowicz M, et al. Chemical chaperone treatment reduces intracellular accumulation of mutant collagen IV and ameliorates the cellular phenotype of a COL4A2 mutation that causes haemorrhagic stroke. *Hum. Mol. Genet.* 2014;23:283–92. doi: 10.1093/hmg/ddt418

165. Plaisier E, Chen Z, Gekeler F, Benhassine S, Dahan K, Marro B, et al. Novel COL4A1 mutations associated with HANAC syndrome: a role for the triple helical CB3[IV] domain. *Am. J. Med. Genet. A*. 2010;152A:2550–5. doi: 10.1002/ajmg.a.33659

166. Rannikmäe K, Sivakumaran V, Millar H, Malik R, Anderson CD, Chong M, et al. *COL4A2* is associated with lacunar ischemic stroke and deep ICH. *Neurology*. 2017;89:1829–1839. doi: 10.1212/WNL.0000000000004560

167. Siegenthaler JA, Choe Y, Patterson KP, Hsieh I, Li D, Jaminet S-C, et al. Foxc1 is required by pericytes during fetal brain angiogenesis. *Biol. Open*. 2013;2:647–59. doi: 10.1242/bio.20135009

168. Reyahi A, Nik AM, Ghiami M, Gritli-Linde A, Pontén F, Johansson BR, et al. Foxf2 Is Required for Brain Pericyte Differentiation and Development and Maintenance of the Blood-Brain Barrier. *Dev. Cell*. 2015;34:19–32. doi: 10.1016/j.devcel.2015.05.008

169. McKeone R, Vieira H, Gregory-Evans K, Gregory-Evans CY, Denny P. Foxf2: A Novel Locus for Anterior Segment Dysgenesis Adjacent to the Foxc1 Gene. *PLoS One*. 2011;6:e25489. doi: 10.1371/journal.pone.0025489

170. Dziewulska D, Lewandowska E. Pericytes as a new target for pathological processes in CADASIL. *Neuropathology*. 2012;32:515–521. doi: 10.1111/j.1440-1789.2011.01290.x

171. Gu X, Liu X-Y, Fagan A, Gonzalez-Toledo ME, Zhao L-R. Ultrastructural changes in cerebral capillary pericytes in aged Notch3 mutant transgenic mice. *Ultrastruct. Pathol.* 2012;36:48–55. doi: 10.3109/01913123.2011.620220

172. Ghosh M, Balbi M, Hellal F, Dichgans M, Lindauer U, Plesnila N. Pericytes are involved in the pathogenesis of cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy. *Ann. Neurol.* 2015;78:887–900. doi: 10.1002/ana.24512

173. Rustenhoven J, Aalderink M, Scotter EL, Oldfield RL, Bergin PS, Mee EW, et al. TGF-beta1 regulates human brain pericyte inflammatory processes involved in neurovasculature function. *J. Neuroinflammation*. 2016;13:37. doi: 10.1186/s12974-016-0503-0

174. Nozaki H, Kato T, Nihonmatsu M, Saito Y, Mizuta I, Noda T, et al. Distinct molecular mechanisms of HTRA1 mutants in manifesting heterozygotes with CARASIL. *Neurology*. 2016;86:1964–74. doi: 10.1212/WNL.0000000000002694

175. Miyazaki T, Kim Y-S, Yoon J, Wang H, Suzuki T, Morse HC, et al. The 3’-5’ DNA exonuclease TREX1 directly interacts with poly(ADP-ribose) polymerase-1 (PARP1) during the DNA damage response. *J. Biol. Chem.* 2014;289:32548–58. doi: 10.1074/jbc.M114.547331

176. Adib-Samii P, Devan W, Traylor M, Lanfranconi S, Zhang CR, Cloonan L, et al. Genetic architecture of white matter hyperintensities differs in hypertensive and nonhypertensive ischemic stroke. *Stroke.* 2015;46:348–53. doi: 10.1161/STROKEAHA.114.006849

177. Holliday EG, Traylor M, Malik R, Bevan S, Falcone G, Hopewell JC, et al. Genetic overlap between diagnostic subtypes of ischemic stroke. *Stroke.* 2015;46:615–9. doi: 10.1161/STROKEAHA.114.007930

178. Verhaaren BFJ, Debette S, Bis JC, Smith JA, Ikram MK, Adams HH, et al. Multiethnic genome-wide association study of cerebral white matter hyperintensities on MRI. *Circ. Cardiovasc. Genet.* 2015;8:398–409. doi: 10.1161/CIRCGENETICS.114.000858

179. Traylor M, Malik R, Nalls MA, Cotlarciuc I, Radmanesh F, Thorleifsson G, et al. Genetic Variation at 16q24.2 is associated with small vessel stroke. *Ann. Neurol.* 2016;doi: 10.1002/ana.24840

180. Teider N, Scott DK, Neiss A, Weeraratne SD, Amani VM, Wang Y, et al. Neuralized1 causes apoptosis and downregulates Notch target genes in medulloblastoma. *Neuro.*

Oncol. 2010;12:1244–56. doi: 10.1093/neuonc/noq091

181. Koutelou E, Sato S, Tomomori-Sato C, Florens L, Swanson SK, Washburn MP, et al. Neuralized-like 1 (Neurl1) targeted to the plasma membrane by N-myristoylation regulates the Notch ligand Jagged1. *J. Biol. Chem.* 2008;283:3846–53. doi: 10.1074/jbc.M706974200

182. Traylor M, Rutten-Jacobs LCA, Thijs V, Holliday EG, Levi C, Bevan S, et al. Genetic Associations With White Matter Hyperintensities Confer Risk of Lacunar Stroke. *Stroke.* 2016;47:1174–9. doi: 10.1161/STROKEAHA.115.011625

183. Habashi JP, Judge DP, Holm TM, Cohn RD, Loeys BL, Cooper TK, et al. Losartan, an AT1 antagonist, prevents aortic aneurysm in a mouse model of Marfan syndrome. *Science*. 2006;312:117–21. doi: 10.1126/science.1124287

184. Boycott KM, Vanstone MR, Bulman DE, MacKenzie AE. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nat. Rev. Genet.* 2013;14:681–91. doi: 10.1038/nrg3555

185. Pulst SM. Genetic linkage analysis. *Arch. Neurol.* 1999;56:667–72.

186. Kerem B, Rommens JM, Buchanan JA, Markiewicz D, Cox TK, Chakravarti A, et al. Identification of the cystic fibrosis gene: genetic analysis. *Science*. 1989;245:1073–80.

187. MacDonald ME, Novelletto A, Lin C, Tagle D, Barnes G, Bates G, et al. The Huntington's disease candidate region exhibits many different haplotypes. *Nat. Genet.* 1992;1:99–103. doi: 10.1038/ng0592-99

188. Bush WS, Moore JH, Li J, McDonnell S, Rabe K. Chapter 11: Genome-Wide Association Studies. *PLoS Comput. Biol.* 2012;8:e1002822. doi: 10.1371/journal.pcbi.1002822

189. Ng PC, Kirkness EF. Whole Genome Sequencing Http://link.springer.com/10.1007/978-1-60327-367-1_12. Humana Press, Totowa, NJ; 2010. p. 215–226.doi: 10.1007/978-1-60327-367-1_12

190. Bodi K, Perera AG, Adams PS, Bintzler D, Dewar K, Grove DS, et al. Comparison of commercially available target enrichment methods for next-generation sequencing. *J. Biomol. Tech.* 2013;24:73–86. doi: 10.7171/jbt.13-2402-002

191. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 2009;10:57–63. doi: 10.1038/nrg2484

192. Soto J, Rodriguez-Antolin C, Vallespín E, de Castro Carpeño J, Ibanez de Caceres I. The impact of next-generation sequencing on the DNA methylation–based translational cancer research. *Transl. Res.* 2016;169:1–18.e1. doi: 10.1016/j.trsl.2015.11.003

193. Ott J, Wang J, Leal SM. Genetic linkage analysis in the age of whole-genome sequencing. *Nat. Rev. Genet.* 2015;16:275–84. doi: 10.1038/nrg3908

194. Vrijenhoek T, Kraaijeveld K, Elferink M, de Ligt J, Kranendonk E, Santen G, et al. Next-generation sequencing-based genome diagnostics across clinical genetics centers: implementation choices and their effects. *Eur. J. Hum. Genet.* 2015;doi: 10.1038/ejhg.2014.279

195. Network NUGT. Dementia Disorders 27 Gene Panel Https://ukgtn.nhs.uk/find-a-test/search-by-disorder-gene/dementia-disorders-27-gene-panel-891/. 2017;

196. Illumina Inc. Illumina Sequencing Technology Https://www.illumina.com/documents/products/techspotlights/techspotlight_sequencing.pdf. 2010;

197. Simeoni I, Stephens JC, Hu F, Deevi SV V., Megy K, Bariana TK, et al. A high-throughput sequencing test for diagnosing inherited bleeding, thrombotic, and platelet disorders. *Blood*. 2016;127.

198. Guerreiro R, Brás J, Hardy J, Singleton A. Next Generation Sequencing Techniques in Neurological Diseases: Redefining Clinical and Molecular Associations. *Hum. Mol. Genet.* 2014;44:1–7. doi: 10.1093/hmg/ddu203

199. Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, Nickerson DA, et al. Exome sequencing as a tool for Mendelian disease gene discovery. *Nat. Rev. Genet.* 2011;12:745–55. doi: 10.1038/nrg3031

200. Lek M, Karczewski KJ, Minikel E V., Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016;536:285–291. doi: 10.1038/nature19057

201. Karczewski KJ, Francioli L. The genome Aggregation Database (gnomAD) Https://macarthurlab.org/2017/02/27/the-genome-aggregation-database-gnomad/. 2017;

202. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. 2009;doi: 10.1038/nprot.2009.86

203. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nat. Methods*. 2010;7:248–249. doi: 10.1038/nmeth0410-248

204. Flanagan SE, Patch A-M, Ellard S. Using SIFT and PolyPhen to Predict Loss-of-Function and Gain-of-Function Mutations. *Genet. Test. Mol. Biomarkers*. 2010;14:533–537. doi: 10.1089/gtmb.2010.0036

205. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* 2014;46:310–5. doi: 10.1038/ng.2892

206. Pescini F, Nannucci S, Bertaccini B, Salvadori E, Bianchi S, Ragno M, et al. The Cerebral Autosomal-Dominant Arteriopathy With Subcortical Infarcts and Leukoencephalopathy (CADASIL) Scale: a screening tool to select patients for NOTCH3 gene analysis. *Stroke.* 2012;43:2871–6. doi: 10.1161/STROKEAHA.112.665927

207. Nannucci S, Pescini F, Bertaccini B, Bianchi S, Ciolli L, Valenti R, et al. Clinical, familial, and neuroimaging features of CADASIL-like patients. *Acta Neurol. Scand.* 2015;131:30–36. doi: 10.1111/ane.12284

208. Verdura E, Hervé D, Bergametti F, Jacquet C, Morvan T, Prieto-Morin C, et al. Disruption of a miR-29 binding site leading to COL4A1 upregulation causes PADMAL.

Ann. Neurol. 2016;doi: 10.1002/ana.24782

209. Siitonen M, Börjesson-Hanson A, Pöyhönen M, Ora A, Pasanen P, Bras J, et al. Multi-infarct dementia of Swedish type is caused by a 3'UTR mutation of COL4A1. *Brain*. 2017;140:e29–e29. doi: 10.1093/brain/awx062

210. Rothwell P, Coull A, Silver L, Fairhead J, Giles M, Lovelock C, et al. Population-based study of event-rate, incidence, case fatality, and mortality for all acute vascular events in all arterial territories (Oxford Vascular Study). *Lancet*. 2005;366:1773–1783. doi: 10.1016/S0140-6736(05)67702-1

211. Fazekas F, Kleinert R, Offenbacher H, Schmidt R, Kleinert G, Payer F, et al. Pathologic correlates of incidental MRI white matter signal hyperintensities. *Neurology*. 1993;43:1683–9.

212. Foo J-N, Liu J-J, Tan E-K. Whole-genome and whole-exome sequencing in neurological diseases. *Nat. Rev. Neurol.* 2012;8:508–17. doi: 10.1038/nrneurol.2012.148

213. van der Velde KJ, de Boer EN, van Diemen CC, Sikkema-Raddatz B, Abbott KM, Knopperts A, et al. GAVIN: Gene-Aware Variant INterpretation for medical sequencing. *Genome Biol.* 2017;18:6. doi: 10.1186/s13059-016-1141-7

214. Guey S, Mawet J, Hervé D, Duering M, Godin O, Jouvent E, et al. Prevalence and characteristics of migraine in CADASIL. *Cephalalgia*. 2015;doi: 10.1177/0333102415620909

215. Choi JC, Song S-K, Lee JS, Kang S-Y, Kang J-H. Headache among CADASIL patients with R544C mutation: prevalence, characteristics, and associations. *Cephalalgia*. 2014;34:22–8. doi: 10.1177/0333102413497598

216. Liem MK, Oberstein SAL, van der Grond J, Ferrari MD, Haan J. CADASIL and migraine: A narrative review. *Cephalalgia*. 2010;30:1284–1289. doi: 10.1177/0333102410370870

217. Vahedi K, Chabriat H, Levy C, Joutel A, Tournier-Lasserve E, Bousser M-G. Migraine with aura and brain magnetic resonance imaging abnormalities in patients with CADASIL. *Arch. Neurol.* 2004;61:1237–40. doi: 10.1001/archneur.61.8.1237

218. Donnini I, Nannucci S, Valenti R, Pescini F, Bianchi S, Inzitari D, et al. Acetazolamide for the prophylaxis of migraine in CADASIL: a preliminary experience. *J. Headache Pain*. 2012;13:299–302. doi: 10.1007/s10194-012-0426-9

219. Forteza AM, Brozman B, Rabinstein AA, Romano JG, Bradley WG. Acetazolamide for the treatment of migraine with aura in CADASIL. *Neurology*. 2001;57:2144–2145. doi: 10.1212/WNL.57.11.2144

220. Weller M, Dichgans J, Klockgether T. Acetazolamide-responsive migraine in CADASIL. *Neurology*. 1998;50:1505–1505. doi: 10.1212/WNL.50.5.1505

221. Martikainen MH, Roine S. Rapid improvement of a complex migrainous episode with sodium valproate in a patient with CADASIL. *J. Headache Pain*. 2012;13:95–7. doi: 10.1007/s10194-011-0400-y

222. Spector JT, Kahn SR, Jones MR, Jayakumar M, Dalal D, Nazarian S. Migraine headache and ischemic stroke risk: an updated meta-analysis. *Am. J. Med.* 2010;123:612–24. doi: 10.1016/j.amjmed.2009.12.021

223. Kruit MC, van Buchem MA, Launer LJ, Terwindt GM, Ferrari MD. Migraine is associated with an increased risk of deep white matter lesions, subclinical posterior circulation infarcts and brain iron accumulation: the population-based MRI CAMERA study. *Cephalalgia*. 2010;30:129–36. doi: 10.1111/j.1468-2982.2009.01904.x

224. Bousser M-G, Welch KMA. Relation between migraine and stroke. *Lancet. Neurol.* 2005;4:533–42. doi: 10.1016/S1474-4422(05)70164-2

225. Etminan M, Takkouche B, Isorna FC, Samii A. Risk of ischaemic stroke in people with migraine: systematic review and meta-analysis of observational studies. *BMJ*. 2005;330:63. doi: 10.1136/bmj.38302.504063.8F

226. Schon F, Martin RJ, Prevett M, Clough C, Enevoldson TP, Markus HS. "CADASIL coma": an underdiagnosed acute encephalopathy. *J. Neurol. Neurosurg. Psychiatry*. 2003;74:249–252. doi: 10.1136/jnnp.74.2.249

227. Classification Committee of the International Headache Society . The International Classification of Headache Disorders, 3rd edition (beta version). *Cephalalgia*. 2013;33:629–808. doi: 10.1177/0333102413485658

228. Sathe S, DePeralta E, Pastores G, Kolodny EH. Acute confusional migraine may be a presenting feature of CADASIL. *Headache*. 2009;49:590–6. doi: 10.1111/j.1526-4610.2009.01363.x

229. Launer LJ, Terwindt GM, Ferrari MD. The prevalence and characteristics of migraine in a population-based cohort: the GEM study. *Neurology*. 1999;53:537–42.

230. Lipton RB, Stewart WF, Diamond S, Diamond ML, Reed M. Prevalence and Burden of Migraine in the United States: Data From the American Migraine Study II. *Headache J. Head Face Pain*. 2001;41:646–657. doi: 10.1046/j.1526-4610.2001.041007646.x

231. Zivadinov R, Willheim K, Jurjevic A, Sepic-Grahovac D, Bucuk M, Zorzon M. Prevalence of migraine in Croatia: a population-based survey. *Headache*. 2001;41:805–12.

232. Becker C, Brobert GP, Almqvist PM, Johansson S, Jick SS, Meier CR. Migraine incidence, comorbidity and health resource utilization in the UK. *Cephalalgia*. 2008;28:57–64. doi: 10.1111/j.1468-2982.2007.01469.x

233. Stewart WF, Wood C, Reed ML, Roy J, Lipton RB. Cumulative lifetime migraine incidence in women and men. *Cephalalgia*. 2008;28:1170–8. doi: 10.1111/j.1468-2982.2008.01666.x

234. Gunda B, Hervé D, Godin O, Bruno M, Reyes S, Alili N, et al. Effects of gender on the phenotype of CADASIL. *Stroke.* 2012;43:137–41. doi: 10.1161/STROKEAHA.111.631028

235. Wenzel RG, Sarvis CA, Krause ML. Over-the-counter drugs for acute migraine attacks: literature review and recommendations. *Pharmacotherapy*. 2003;23:494–505.

236. Tfelt-Hansen P. A review of evidence-based medicine and meta-analytic reviews in migraine. *Cephalalgia*. 2006;26:1265–74. doi: 10.1111/j.1468-2982.2006.01194.x

237. MHRA. Imigran 100mg TABLETS (sumatriptan succinate) Patient Information Leaflet Http://www.medicines.org.uk/emc/medicine/749. 2013;

238. Hall GC, Brown MM, Mo J, MacRae KD. Triptans in migraine: the risks of stroke, cardiovascular disease, and death in practice. *Neurology*. 2004;62:563–8.

239. Chabriat H, Vahedi K, Iba-Zizen MT, Joutel A, Nibbio A, Nagy TG, et al. Clinical spectrum of CADASIL: a study of 7 families. Cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy. *Lancet*. 1995;346:934–9.

240. Fan Y, McGowan S, Rubeiz H, Wollmann R, Javed A, Mastrianni J. Acute encephalopathy as the initial manifestation of CADASIL. *Neurol. Clin. Pract.* 2012;2:359–361. doi: 10.1212/CPJ.0b013e31826af219

241. Eikermann-Haerter K. Spreading depolarization may link migraine and stroke. *Headache*. 2014;54:1146–57. doi: 10.1111/head.12386

242. Brennan KC, Beltrán-Parrazal L, López-Valdés HE, Theriot J, Toga AW, Charles AC. Distinct vascular conduction with cortical spreading depression. *J. Neurophysiol.* 2007;97:4143–51. doi: 10.1152/jn.00028.2007

243. Liu X, Sheng R, Qin Z. The neuroprotective mechanism of brain ischemic preconditioning. *Acta Pharmacol. Sin.* 2009;30:1071–80. doi: 10.1038/aps.2009.105

244. Abramycheva N, Stepanova M, Kalashnikova L, Zakharova M, Maximova M, Tanashyan M, et al. New mutations in the Notch3 gene in patients with cerebral autosomal dominant arteriopathy with subcortical infarcts and leucoencephalopathy (CADASIL). *J. Neurol. Sci.* 2015;349:196–201. doi: 10.1016/j.jns.2015.01.018

245. Markus HS, Martin RJ, Simpson MA, Dong YB, Ali N, Crosby AH, et al. Diagnostic strategies in CADASIL. *Neurology*. 2002;59:1134–8.

246. MacArthur JAL, Morales J, Tully RE, Astashyn A, Gil L, Bruford EA, et al. Locus Reference Genomic: reference sequences for the reporting of clinically relevant sequence variants. *Nucleic Acids Res.* 2014;42:D873–D878. doi: 10.1093/nar/gkt1198

247. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25:1754–1760. doi: 10.1093/bioinformatics/btp324

248. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26:841–2. doi: 10.1093/bioinformatics/btq033

249. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;20:1297–303. doi: 10.1101/gr.107524.110

250. Plagnol V, Curtis J, Epstein M, Mok KY, Stebbings E, Grigoriadou S, et al. A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. *Bioinformatics*. 2012;28:2747–54. doi:

10.1093/bioinformatics/bts526

251. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly (Austin).* 2012;6:80–92. doi: 10.4161/fly.19695

252. Firth H V, Richards SM, Bevan AP, Clayton S, Corpas M, Rajan D, et al. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *Am. J. Hum. Genet.* 2009;84:524–33. doi: 10.1016/j.ajhg.2009.03.010

253. Costrop LM, Vanakker OO, Van Laer L, Le Saux O, Martin L, Chassaing N, et al. Novel deletions causing pseudoxanthoma elasticum underscore the genomic instability of the ABCC6 region. *J. Hum. Genet.* 2010;55:112–117. doi: 10.1038/jhg.2009.132

254. Le Saux O, Beck K, Sachsinger C, Silvestri C, Treiber C, Göring HH, et al. A spectrum of ABCC6 mutations is responsible for pseudoxanthoma elasticum. *Am. J. Hum. Genet.* 2001;69:749–64.

255. Gale DP, Oygar DD, Lin F, Oygar PD, Khan N, Connor TMF, et al. A novel *COL4A1* frameshift mutation in familial kidney disease: the importance of the C-terminal NC1 domain of type IV collagen. *Nephrol. Dial. Transplant.* 2016;31:1908–1914. doi: 10.1093/ndt/gfw051

256. Moreton FC, Razvi SSM, Davidson R, Muir KW. Changing clinical patterns and increasing prevalence in CADASIL. *Acta Neurol. Scand.* 2014;130:197–203. doi: 10.1111/ane.12266

257. Peters N, Opherk C, Bergmann T, Castro M, Herzog J, Dichgans M. Spectrum of mutations in biopsy-proven CADASIL: implications for diagnostic strategies. *Arch. Neurol.* 2005;62:1091–4. doi: 10.1001/archneur.62.7.1091

258. Opherk C, Peters N, Herzog J, Luedtke R, Dichgans M. Long-term prognosis and causes of death in CADASIL: a retrospective study in 411 patients. *Brain*. 2004;127:2533–9. doi: 10.1093/brain/awh282

259. Liao Y-C, Hsiao C-T, Fuh J-L, Chern C-M, Lee W-J, Guo Y-C, et al. Characterization of CADASIL among the Han Chinese in Taiwan: Distinct Genotypic and Phenotypic Profiles. *PLoS One*. 2015;10:e0136501. doi: 10.1371/journal.pone.0136501

260. Oberstein SAJL. Diagnostic strategies in CADASIL. *Neurology*. 2003;60:2020; author reply 2020.

261. Ishida C, Sakajiri K, Yoshita M, Joutel A, Cave-Riant F, Yamada M. CADASIL with a novel mutation in exon 7 of NOTCH3 (C388Y). *Intern. Med.* 2006;45:981–5.

262. Kim Y, Choi EJ, Choi CG, Kim G, Choi JH, Yoo HW, et al. Characteristics of CADASIL in Korea: a novel cysteine-sparing Notch3 mutation. *Neurology*. 2006;66:1511–6. doi: 10.1212/01.wnl.0000216259.99811.50

263. Ampuero I, Alegre-Abarrategui J, Rodal I, España A, Ros R, Sendón JLL, et al. On the Diagnosis of CADASIL. *J. Alzheimer's Dis.* 2009;17:787–794. doi: 10.3233/JAD-2009-1112

264. Tateoka T, Onda H, Hirota K, Kasuya H, Shinohara T, Kinouchi H, et al. Unusual case of cerebral small vessel disease with a heterozygous nonsense mutation in HTRA1. *J. Neurol. Sci.* 2016;362:144–146. doi: 10.1016/j.jns.2016.01.037

265. Di Donato I, Bianchi S, Gallus GN, Cerase A, Taglia I, Pescini F, et al. Heterozygous mutations of *HTRA1* gene in patients with familial cerebral small vessel disease. *CNS Neurosci. Ther.* 2017;doi: 10.1111/cns.12722

266. Söder S, Pöschl E. The NC1 domain of human collagen IV is necessary to initiate triple helix formation. *Biochem. Biophys. Res. Commun.* 2004;325:276–280. doi: 10.1016/j.bbrc.2004.10.034

267. Decio A, Tonduti D, Pichiecchio A, Vetro A, Ciccone R, Limongelli I, et al. A novel mutation in *COL4A1* gene: A possible cause of early postnatal cerebrovascular events. *Am. J. Med. Genet. Part A*. 2015;167:810–815. doi: 10.1002/ajmg.a.36907

268. Labelle-Dumais C, Dilworth DJ, Harrington EP, de Leau M, Lyons D, Kabaeva Z, et al. COL4A1 mutations cause ocular dysgenesis, neuronal localization defects, and myopathy in mice and Walker-Warburg syndrome in humans. *PLoS Genet.* 2011;7:e1002062. doi: 10.1371/journal.pgen.1002062

269. Meuwissen MEC, Halley DJJ, Smit LS, Lequin MH, Cobben JM, de Coo R, et al. The expanding phenotype of COL4A1 and COL4A2 mutations: clinical data on 13 newly identified families and a review of the literature. *Genet. Med.* 2015;17:843–853. doi: 10.1038/gim.2014.210

270. Kumar M, Chambers C, Dhamija R. Axenfeld-Rieger Syndrome and Leukoencephalopathy Caused by a Mutation in FOXC1. *Pediatr. Neurol.* 2017;66:113–114. doi: 10.1016/j.pediatrneurol.2016.08.020

271. Crow YJ, Hayward BE, Parmar R, Robins P, Leitch A, Ali M, et al. Mutations in the gene encoding the 3'-5' DNA exonuclease TREX1 cause Aicardi-Goutières syndrome at the AGS1 locus. *Nat. Genet.* 2006;38:917–20. doi: 10.1038/ng1845

272. Rice G, Newman WG, Dean J, Patrick T, Parmar R, Flintoff K, et al. Heterozygous mutations in TREX1 cause familial chilblain lupus and dominant Aicardi-Goutieres syndrome. *Am. J. Hum. Genet.* 2007;80:811–5. doi: 10.1086/513443

273. McGlasson S, Rannikmäe K, Bevan S, Logan C, Bicknell LS, Jury A, et al. Rare variants of the 3'-5' DNA exonuclease TREX1 in early onset small vessel stroke. *Wellcome Open Res.* 2017;2:106. doi: 10.12688/wellcomeopenres.12631.1

274. Moore DF, Kaneski CR, Askari H, Schiffmann R. The cerebral vasculopathy of Fabry disease. *J. Neurol. Sci.* 2007;257:258–63. doi: 10.1016/j.jns.2007.01.053

275. Yasuda M, Shabbeer J, Benson SD, Maire I, Burnett RM, Desnick RJ. Fabry disease: Characterization of ?-galactosidase A double mutations and the D313Y plasma enzyme pseudodeficiency allele. *Hum. Mutat.* 2003;22:486–492. doi: 10.1002/humu.10275

276. Hoffmann B, Georg Koch H, Schweitzer-Krantz S, Wendel U, Mayatepek E. Deficient alpha-galactosidase A activity in plasma but no Fabry disease--a pitfall in diagnosis. *Clin. Chem. Lab. Med.* 2005;43:1276–7. doi: 10.1515/CCLM.2005.219

277. Eckman CB, Mehta ND, Crook R, Perez-tur J, Prihar G, Pfeiffer E, et al. A new pathogenic mutation in the APP gene (I716V) increases the relative proportion of A beta 42(43). *Hum. Mol. Genet.* 1997;6:2087–9.

278. Jurkat-Rott K, Freilinger T, Dreier JP, Herzog J, Göbel H, Petzold GC, et al. Variability of familial hemiplegic migraine with novel A1A2 Na+/K+-ATPase variants. *Neurology*. 2004;62:1857–61.

279. Xu X, Yang X, Wu Q, Liu A, Yang X, Ye AY, et al. Amplicon Resequencing Identified Parental Mosaicism for Approximately 10% of “de novo” SCN1A Mutations in Children with Dravet Syndrome. *Hum. Mutat.* 2015;36:861–872. doi: 10.1002/humu.22819

280. Miksch S, Lumsden A, Guenther UP, Foernzler D, Christen-Zäch S, Daugherty C, et al. Molecular genetics of pseudoxanthoma elasticum: type and frequency of mutations in ABCC6. *Hum. Mutat.* 2005;26:235–48. doi: 10.1002/humu.20206

281. Ringpfeil F, McGuigan K, Fuchsel L, Kozic H, Larralde M, Lebwohl M, et al. Pseudoxanthoma elasticum is a recessive disease characterized by compound heterozygosity. *J. Invest. Dermatol.* 2006;126:782–6. doi: 10.1038/sj.jid.5700115

282. Chassaing N, Martin L, Calvas P, Le Bert M, Hovnanian A. Pseudoxanthoma elasticum: a clinical, pathophysiological and genetic update including 11 novel ABCC6 mutations. *J. Med. Genet.* 2005;42:881–92. doi: 10.1136/jmg.2004.030171

283. Pantoni L, Pescini F, Nannucci S, Sarti C, Bianchi S, Dotti MT, et al. Comparison of clinical, familial, and MRI features of CADASIL and NOTCH3-negative patients. *Neurology*. 2009;74:57–63. doi: 10.1212/WNL.0b013e3181c7da7c

284. Cummings BB, Marshall JL, Tukiainen T, Lek M, Donkervoort S, Foley AR, et al. Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. *Sci. Transl. Med.* 2017;9:eaal5209. doi: 10.1126/scitranslmed.aal5209

285. Raczy C, Petrovski R, Saunders CT, Chorny I, Kruglyak S, Margulies EH, et al. Isaac: ultra-fast whole-genome secondary analysis on Illumina sequencing platforms. *Bioinformatics*. 2013;29:2041–2043. doi: 10.1093/bioinformatics/btt314

286. Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Källberg M, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics*. 2016;32:1220–1222. doi: 10.1093/bioinformatics/btv710

287. Roller E, Ivakhno S, Lee S, Royce T, Tanner S. Canvas: versatile and scalable detection of copy number variants. *Bioinformatics*. 2016;32:2375–2377. doi: 10.1093/bioinformatics/btw163

288. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera?A visualization system for exploratory research and analysis. *J. Comput. Chem.* 2004;25:1605–1612. doi: 10.1002/jcc.20084

289. Capo-Chichi J-M, Tcherkezian J, Hamdan FF, Décarie JC, Dobrzeniecka S, Patry L, et al. Disruption of TBC1D7, a subunit of the TSC1-TSC2 protein complex, in intellectual disability and megalencephaly. *J. Med. Genet.* 2013;50:740–744. doi: 10.1136/jmedgenet-2013-101680

290. Alfaiz AA, Micale L, Mandriani B, Augello B, Pellico MT, Chrast J, et al. *TBC1D7* Mutations are Associated with Intellectual Disability, Macrocrania, Patellar Dislocation, and Celiac Disease. *Hum. Mutat.* 2014;35:447–451. doi: 10.1002/humu.22529

291. Bayrakli F, Balaban H, Gurelik M, Hizmetli S, Topaktas S. Mutation in htra1 gene in a patient with degenerated spine as a component of carasil syndrome. *Turk. Neurosurg.* 2012;24:67–9. doi: 10.5137/1019-5149.JTN.6226-12.1

292. Oka C, Tsujimoto R, Kajikawa M, Koshiba-Takeuchi K, Ina J, Yano M, et al. HtrA1 serine protease inhibits signaling mediated by Tgfbeta family proteins. *Development*. 2004;131:1041–53. doi: 10.1242/dev.00999

293. Chen Y, He Z, Meng S, Li L, Yang H, Zhang X. A novel mutation of the high-temperature requirement A serine peptidase 1 (*HTRA1*) gene in a Chinese family with cerebral autosomal recessive arteriopathy with subcortical infarcts and leukoencephalopathy (CARASIL). *J. Int. Med. Res.* 2013;41:1445–1455. doi: 10.1177/0300060513480926

294. Lemmens R, Maugeri A, Niessen HWM, Goris A, Tousseyn T, Demaerel P, et al. Novel COL4A1 mutations cause cerebral small vessel disease by haploinsufficiency. *Hum. Mol. Genet.* 2013;22:391–7. doi: 10.1093/hmg/dds436

295. Siitonen M, Börjesson-Hanson A, Pöyhönen M, Ora A, Pasanen P, Bras J, et al. Multi-infarct dementia of Swedish type is caused by a 3’UTR mutation of COL4A1. *Brain*. 2017;140:e29–e29. doi: 10.1093/brain/awx062

296. de Vries LS, Koopman C, Groenendaal F, Van Schooneveld M, Verheijen FW, Verbeek E, et al. COL4A1 mutation in two preterm siblings with antenatal onset of parenchymal hemorrhage. *Ann. Neurol.* 2009;65:12–8. doi: 10.1002/ana.21525

297. Dölz R, Engel J, Kühn K. Folding of collagen IV. *Eur. J. Biochem.* 1988;178:357–66.

298. Nordborg C, Kyllerman M, Conradi N, Månsson JE. Early-infantile galactosialidosis with multiple brain infarctions: morphological, neuropathological and neurochemical findings. *Acta Neuropathol.* 1997;93:24–33.

299. Bugiani M, Kevelam SH, Bakels HS, Waisfisz Q, Ceuterick-de Groote C, Niessen HWM, et al. Cathepsin A-related arteriopathy with strokes and leukoencephalopathy (CARASAL). *Neurology*. 2016;doi: 10.1212/WNL.0000000000003251

300. Butteriss DJA, Ramesh V, Birchall D. Serial MRI in a case of familial hemiplegic migraine. *Neuroradiology*. 2003;45:300–303. doi: 10.1007/s00234-003-0979-z

301. Lindahl AJ, Allder S, Jefferson D, Allder S, Moody A, Martel A. Prolonged hemiplegic migraine associated with unilateral hyperperfusion on perfusion weighted magnetic resonance imaging. *J. Neurol. Neurosurg. Psychiatry*. 2002;73:202–3. doi: 10.1136/JNNP.73.2.202

302. Nimmrich V, Gross G. P/Q-type calcium channel modulators. *Br. J. Pharmacol.* 2012;167:741–59. doi: 10.1111/j.1476-5381.2012.02069.x

303. Veneziano L, Guida S, Mantuano E, Bernard P, Tarantino P, Boccone L, et al. Newly

characterised 5’ and 3’ regions of CACNA1A gene harbour mutations associated with Familial Hemiplegic Migraine and Episodic Ataxia. *J. Neurol. Sci.* 2009;276:31–7. doi: 10.1016/j.jns.2008.08.030

304. Nachbauer W, Nocker M, Karner E, Stankovic I, Unterberger I, Eigentler A, et al. Episodic ataxia type 2: phenotype characteristics of a novel CACNA1A mutation and review of the literature. *J. Neurol.* 2014;261:983–91. doi: 10.1007/s00415-014-7310-2

305. Camia F, Pisciotta L, Morana G, Schiaffino MC, Renna S, Carrera P, et al. Combined early treatment in hemiplegic attacks related to CACNA1A encephalopathy with brain oedema: Blocking the cascade? *Cephalalgia*. 2016;doi: 10.1177/0333102416668655

306. Kume T, Jiang H, Topczewska JM, Hogan BL. The murine winged helix transcription factors, Foxc1 and Foxc2, are both required for cardiovascular development and somitogenesis. *Genes Dev.* 2001;15:2470–82. doi: 10.1101/gad.907301

307. Seo S, Fujita H, Nakano A, Kang M, Duarte A, Kume T. The forkhead transcription factors, Foxc1 and Foxc2, are required for arterial specification and lymphatic sprouting during vascular development. *Dev. Biol.* 2006;294:458–70. doi: 10.1016/j.ydbio.2006.03.035

308. Mutations in FOXC2 (MFH-1), a Forkhead Family Transcription Factor, Are Responsible for the Hereditary Lymphedema-Distichiasis Syndrome. *Am. J. Hum. Genet.* 2000;67:1382–1388. doi: 10.1086/316915

309. Cederberg A, Grønning LM, Ahrén B, Taskén K, Carlsson P, Enerbäck S. FOXC2 is a winged helix gene that counteracts obesity, hypertriglyceridemia, and diet-induced insulin resistance. *Cell*. 2001;106:563–73.

310. Cerritelli SM, Crouch RJ. Ribonuclease H: the enzymes in eukaryotes. *FEBS J.* 2009;276:1494–1505. doi: 10.1111/j.1742-4658.2009.06908.x

311. Crow YJ, Chase DS, Lowenstein Schmidt J, Szynkiewicz M, Forte GMA, Gornall HL, et al. Characterization of human disease phenotypes associated with mutations in TREX1, RNASEH2A, RNASEH2B, RNASEH2C, SAMHD1, ADAR, and IFIH1. *Am. J. Med. Genet. A*. 2015;167:296–312. doi: 10.1002/ajmg.a.36887

312. McQuillin A, Kalsi G, Moorey H, Lamb G, Mayet S, Quested D, et al. A novel polymorphism in exon 11 of the WKL1 gene, shows no association with schizophrenia. *Eur. J. Hum. Genet.* 2002;10:491–494. doi: 10.1038/sj.ejhg.5200837

313. Santorico SA, Edwards KL. Challenges of Linkage Analysis in the Era of Whole-Genome Sequencing. *Genet. Epidemiol.* 2014;38:S92–S96. doi: 10.1002/gepi.21832

314. Bennett CA, Petrovski S, Oliver KL, Berkovic SF. ExACtly zero or once. *Neurol. Genet.* 2017;3:e163. doi: 10.1212/NXG.0000000000000163

315. Wollenweber FA, Hanecker P, Bayer-Karpinska A, Malik R, Bäzner H, Moreton F, et al. Cysteine-Sparing CADASIL Mutations in NOTCH3 Show Proaggregatory Properties In Vitro. *Stroke.* 2015;46:786–92. doi: 10.1161/STROKEAHA.114.007472

316. Scheid R, Heinritz W, Leyhe T, Thal DR, Schober R, Strenge S, et al. Cysteine-sparing notch3 mutations: cadasil or cadasil variants? *Neurology*. 2008;71:774–6. doi:

10.1212/01.wnl.0000324928.44694.f7

317. Rutten JW, Haan J, Terwindt GM, van Duinen SG, Boon EMJ, Lesnik Oberstein SAJ. Interpretation of NOTCH3 mutations in the diagnosis of CADASIL. *Expert Rev. Mol. Diagn.* 2014;14:593–603. doi: 10.1586/14737159.2014.922880

318. Bersano A, Ranieri M, Ciammola A, Cinnante C, Lanfranconi S, Dotti MT, et al. Considerations on a mutation in the NOTCH3 gene sparing a cysteine residue: A rare polymorphism rather than a CADASIL variant. *Funct. Neurol.* 2012;27:247–252. doi: 5744 [pii]

319. Wollenweber FA, Haffner C, Duering M. Response to Letter Regarding Article, "Cysteine-Sparing CADASIL Mutations in NOTCH3 Show Proaggregatory Properties In Vitro". *Stroke*. 2015;46:e154. doi: 10.1161/STROKEAHA.115.009473

320. Rutten J, van Duinen S, Lesnik Oberstein S. Letter by Rutten et al Regarding Article, "Cysteine-Sparing CADASIL Mutations in NOTCH3 Show Proaggregatory Properties In Vitro". *Stroke*. 2015;46:e153. doi: 10.1161/STROKEAHA.115.009467

321. Muiño E, Gallego-Fabrega C, Cullell N, Carrera C, Torres N, Krupinski J, et al. Systematic Review of Cysteine-Sparing NOTCH3 Missense Mutations in Patients with Clinical Suspicion of CADASIL. *Int. J. Mol. Sci.* 2017;18. doi: 10.3390/ijms18091964

322. Chabriat H, Levy C, Taillia H, Iba-Zizen MT, Vahedi K, Joutel A, et al. Patterns of MRI lesions in CADASIL. *Neurology*. 1998;51:452–7.

323. Brulin P, Godfraind C, Leteurtre E, Ruchoux M-M. Morphometric analysis of ultrastructural vascular changes in CADASIL: analysis of 50 skin biopsy specimens and pathogenic implications. *Acta Neuropathol.* 104:241–248. doi: 10.1007/s00401-002-0530-z

324. Hou X, He C, Jin Q, Niu Q, Ren G, Cheng H. Novel Mutation of the NOTCH3 Gene in a Chinese Pedigree with CADASIL. *CNS Neurol. Disord. Drug Targets*. 2017;16:30–35. doi: 10.2174/1871527315666161024125952

325. Trikamji B, Thomas M, Hathout G, Mishra S. An unusual case of cerebral autosomal-dominant arteriopathy with subcortical infarcts and leukoencephalopathy with occipital lobe involvement. *Ann. Indian Acad. Neurol.* 2016;19:272–4. doi: 10.4103/0972-2327.173403

326. Ungaro C, Mazzei R, Conforti FL, Sprovieri T, Servillo P, Liguori M, et al. CADASIL: Extended polymorphisms and mutational analysis of the NOTCH3 gene. *J. Neurosci. Res.* 2009;87:1162–1167. doi: 10.1002/jnr.21935

327. Brass SD, Smith EE, Arboleda-Velasquez JF, Copen WA, Frosch MP. Case 12-2009. *N. Engl. J. Med.* 2009;360:1656–1665. doi: 10.1056/NEJMcpc0810839

328. Ferreira S, Costa C, Oliveira JP. Novel human pathological mutations. Gene symbol: NOTCH3. Disease: cerebral autosomal dominant arteriopathy with subcortical infarcts and leucoencephalopathy (CADASIL). *Hum. Genet.* 2007;121:649.

329. Nakamura K, Ago T, Tsuchimoto A, Noda N, Nakamura A, Ninomiya T, et al. A CADASIL-Like Case with a Novel Noncysteine Mutation of the *NOTCH3* Gene and Granular Deposits in the Renal Arterioles. *Case Rep. Neurol. Med.* 2015;2015:1–6. doi: 10.1155/2015/431461

330. Di Donato I, Bianchi S, Gallus GN, Cerase A, Taglia I, Pescini F, et al. Heterozygous mutations of *HTRA1* gene in patients with familial cerebral small vessel disease. *CNS Neurosci. Ther.* 2017;23:759–765. doi: 10.1111/cns.12722

331. Gould DB, Phalan FC, Breedveld GJ, van Mil SE, Smith RS, Schimenti JC, et al. Mutations in Col4a1 cause perinatal cerebral hemorrhage and porencephaly. *Science*. 2005;308:1167–71. doi: 10.1126/science.1109418

332. Sibon I, Coupry I, Menegon P, Bouchet J-P, Gorry P, Burgelin I, et al. COL4A1 mutation in Axenfeld-Rieger anomaly with leukoencephalopathy and stroke. *Ann. Neurol.* 2007;62:177–184. doi: 10.1002/ana.21191

333. Meuwissen MEC, Halley DJJ, Smit LS, Lequin MH, Cobben JM, de Coo R, et al. The expanding phenotype of COL4A1 and COL4A2 mutations: clinical data on 13 newly identified families and a review of the literature. *Genet. Med.* 2015;17:843–853. doi: 10.1038/gim.2014.210

334. Dong Y, Hassan A, Zhang Z, Huber D, Dalageorgou C, Markus HS. Yield of Screening for CADASIL Mutations in Lacunar Stroke and Leukoaraiosis. *Stroke*. 2003;34.

335. Joutel A. Pathogenesis of CADASIL: transgenic and knock-out mice to probe function and dysfunction of the mutated gene, Notch3, in the cerebrovasculature. *Bioessays*. 2011;33:73–80. doi: 10.1002/bies.201000093

336. Lesnik Oberstein SAJ, van Duinen SG, van den Boom R, Maat-Schieman MLC, van Buchem MA, van Houwelingen HC, et al. Evaluation of diagnostic NOTCH3 immunostaining in CADASIL. *Acta Neuropathol.* 2003;106:107–11. doi: 10.1007/s00401-003-0701-6

337. Li S, Chen Y, Shan H, Ma F, Shi M, Xue J. Novel heterozygous NOTCH3 pathogenic variant found in two Chinese patients with CADASIL. *J. Clin. Neurosci.* 2017;doi: 10.1016/j.jocn.2017.08.029

338. Oki K, Nagata E, Ishiko A, Shimizu A, Tanaka K, Takahashi K, et al. Novel mutation of the Notch3 gene in a Japanese patient with CADASIL. *Eur. J. Neurol.* 2007;14:464–466. doi: 10.1111/j.1468-1331.2007.01641.x

339. Joutel A, Favrole P, Labauge P, Chabriat H, Lescoat C, Andreux F, et al. Skin biopsy immunostaining with a Notch3 monoclonal antibody for CADASIL diagnosis. *Lancet*. 2001;358:2049–51. doi: 10.1016/S0140-6736(01)07142-2

340. Escary J-L, Cecillon M, Maciazek J, Lathrop M, Tournier-Lasserve E, Joutel A. Evaluation of DHPLC analysis in mutational scanning ofNotch3, a gene with a high G-C content. *Hum. Mutat.* 2000;16:518–526. doi: 10.1002/1098-1004(200012)16:6<518::AID-HUMU9>3.0.CO;2-Q

341. Lim K-S, Tan A-H, Lim C-S, Chua K-H, Lee P-C, Ramli N, et al. R54C Mutation of NOTCH3 Gene in the First Rungus Family with CADASIL. *PLoS One*. 2015;10:e0135470.

doi: 10.1371/journal.pone.0135470

342. Cleves C, Friedman NR, Rothner AD, Hussain MS. Genetically Confirmed CADASIL in a Pediatric Patient. *Pediatrics*. 2010;126:e1603–e1607. doi: 10.1542/peds.2010-0714

343. Bianchi S, Zicari E, Carluccio A, Di Donato I, Pescini F, Nannucci S, et al. CADASIL in central Italy: a retrospective clinical and genetic study in 229 patients. *J. Neurol.* 2015;262:134–141. doi: 10.1007/s00415-014-7533-2

344. Tikka S, Mykkänen K, Ruchoux M-M, Bergholm R, Junna M, Pöyhönen M, et al. Congruence between NOTCH3 mutations and GOM in 131 CADASIL patients. *Brain*. 2009;132:933–939. doi: 10.1093/brain/awn364

345. Na DL, Moon S-Y, Kim H-Y, Seok J-I, Kwon J-C, Ki C-S, et al. A Novel Mutation (C67Y) in the NOTCH3 Gene in a Korean CADASIL Patient. *J. Korean Med. Sci.* 2003;18:141. doi: 10.3346/jkms.2003.18.1.141

346. Wang Z, Yuan Y, Zhang W, Lv H, Hong D, Chen B, et al. NOTCH3 mutations and clinical features in 33 mainland Chinese families with CADASIL. *J. Neurol. Neurosurg. Psychiatry*. 2011;82:534–539. doi: 10.1136/jnnp.2010.209247

347. Dichgans M, Ludwig H, Müller-Höcker J, Messerschmidt A, Gasser T. Small in-frame deletions and missense mutations in CADASIL: 3D models predict misfolding of Notch3v EGF-like repeat domains. *Eur. J. Hum. Genet.* 2000;8:280–285. doi: 10.1038/sj.ejhg.5200460

348. Lackovic V, Bajcetic M, Lackovic M, Novakovic I, Labudović Borović M, Pavlovic A, et al. Skin and Sural Nerve Biopsies: Ultrastructural Findings in the First Genetically Confirmed Cases of CADASIL in Serbia. *Ultrastruct. Pathol.* 2012;36:325–335. doi: 10.3109/01913123.2012.679352

349. Pavlovic AM, Dobricic V, Semnic R, Lackovic V, Novakovic I, Bajcetic M, et al. A novel Notch3 Gly89Cys mutation in a Serbian CADASIL family. *Acta Neurol. Belg.* 2013;113:299–302. doi: 10.1007/s13760-012-0174-2

350. Mazzei R, Guidetti D, Ungaro C, Conforti FL, Muglia M, Cenacchi G, et al. First evidence of a pathogenic insertion in the NOTCH3 gene causing CADASIL. *J. Neurol. Neurosurg. Psychiatry*. 2008;79:108–110. doi: 10.1136/jnnp.2007.128009

351. Kalimo H, Ruchoux M-M, Viitanen M, Kalaria RN. CADASIL: a common form of hereditary arteriopathy causing brain infarcts and dementia. *Brain Pathol.* 2002;12:371–84.

352. Yamada K, Sakai K, Akazawa K, Sugimoto N, Nakagawa M, Mizuno T. Detection of early neuronal damage in CADASIL patients by q-space MR imaging. *Neuroradiology*. 2013;55:283–290. doi: 10.1007/s00234-012-1105-x

353. Testi S, Malerba G, Ferrarini M, Ragno M, Pradotto L, Mauro A, et al. Mutational and haplotype map of NOTCH3 in a cohort of Italian patients with cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL). *J. Neurol. Sci.* 2012;319:37–41. doi: 10.1016/j.jns.2012.05.025

354. Rojas-Marcos I, Encarnacion M, Martinez-Yelamos S, Ferrer I, Arbizu T, Gil-Peralta A,

et al. Gene symbol: NOTCH3. Disease: CADASIL. *Hum. Genet.* 2004;115:175.

355. Joutel A, Chabriat H, Vahedi K, Domenga V, Vayssière C, Ruchoux MM, et al. Splice site mutation causing a seven amino acid Notch3 in-frame deletion in CADASIL. *Neurology*. 2000;54:1874–5. doi: 10.1212/WNL.54.9.1874

356. Matsushima T, Conedera S, Tanaka R, Li Y, Yoshino H, Funayama M, et al. Genotype-phenotype correlations of cysteine replacement in CADASIL. *Neurobiol. Aging*. 2017;50:169.e7-169.e14. doi: 10.1016/j.neurobiolaging.2016.10.026

357. Spinicci G, Conti M, Cherchi MV, Mancosu C, Murru R, Carboni N. Unusual clinical presentations in subjects carrying novel NOTCH3 gene mutations. *J. Stroke Cerebrovasc. Dis.* 2013;22:539–44. doi: 10.1016/j.jstrokecerebrovasdis.2013.02.002

358. Lee YC, Yang AH, Liu HC, Wong WJ, Lu YC, Chang MH, et al. Cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy: Two novel mutations in the NOTCH3 gene in Chinese. *J. Neurol. Sci.* 2006;246:111–115. doi: 10.1016/j.jns.2006.02.011

359. Coto E, Menendez M, Navarro R, Garcia-Castro M, Alvarez V. A new de novo Notch3 mutation causing CADASIL. *Eur. J. Neurol.* 2006;13:628–631. doi: 10.1111/j.1468-1331.2006.01337.x

360. Tuominen S, Juvonen V, Amberla K, Jolma T, Rinne JO, Tuisku S, et al. Phenotype of a homozygous CADASIL patient in comparison to 9 age-matched heterozygous patients with the same R133C Notch3 mutation. *Stroke*. 2001;32:1767–74.

361. Siitonen M, Mykkänen K, Pescini F, Rovio S, Kääriäinen H, Baumann M, et al. APOE and AGT in the Finnish p.Arg133Cys CADASIL population. *Acta Neurol. Scand.* 2015;132:430–4. doi: 10.1111/ane.12400

362. Grigg R, Lea R, Sullivan AA, Curtain R, MacMillian J, Griffiths L. Identification of a novel mutation C144F in the notch3 gene in an Australian CADASIL pedigree. *Hum. Mutat.* 2000;16:449–450. doi: 10.1002/1098-1004(200011)16:5<449::AID-HUMU26>3.0.CO;2-I

363. Malandrini A, Gaudiano C, Gambelli S, Berti G, Serni G, Bianchi S, et al. Diagnostic value of ultrastructural skin biopsy studies in CADASIL. *Neurology*. 2007;68:1430–2. doi: 10.1212/01.wnl.0000264018.46335.c8

364. Peters N, Opherk C, Zacherle S, Capell A, Gempel P, Dichgans M. CADASIL-associated Notch3 mutations have differential effects both on ligand binding and ligand-induced Notch3 receptor signaling through RBP-Jk. *Exp. Cell Res.* 2004;299:454–64. doi: 10.1016/j.yexcr.2004.06.004

365. Sangle N, Baringer JR, Majersik J, DeWitt LD. CADASIL with Multiorgan Involvement: a Complete Autopsy Examination Report. *Can. J. Neurol. Sci. / J. Can. des Sci. Neurol.* 2016;43:202–205. doi: 10.1017/cjn.2015.240

366. Andreadou E, Papadimas G, Sfagos C. A novel heterozygous mutation in the NOTCH3 gene causing CADASIL. *Swiss Med. Wkly.* 2008;138:614–7. doi: 2008/41/smw-12394

367. Santa Y, Uyama E, Chui DH, Arima M, Kotorii S, Takahashi K, et al. Genetic, clinical and

pathological studies of CADASIL in Japan: a partial contribution of Notch3 mutations and implications of smooth muscle cell degeneration for the pathogenesis. *J. Neurol. Sci.* 2003;212:79–84.

368. Suda S, Okubo S, Ueda M, Sowa K, Abe A, Aoki J, et al. A Japanese CADASIL kindred with a novel two-base *NOTCH3* mutation. *Eur. J. Neurol.* 2016;23:e32–e34. doi: 10.1111/ene.12977

369. Kotorii S, Takahashi K, Kamimura K, Nishio T, Arima K, Yamada H, et al. Mutations of the notch3 gene in non-caucasian patients with suspected CADASIL syndrome. *Dement. Geriatr. Cogn. Disord.* 12:185–93. doi: 51256

370. GRANILD-JENSEN J, JENSEN UB, SCHWARTZ M, HANSEN US. Cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy resulting in stroke in an 11-year-old male. *Dev. Med. Child Neurol.* 2009;51:754–757. doi: 10.1111/j.1469-8749.2008.03241.x

371. Vinciguerra C, Rufa A, Bianchi S, Sperduto A, De Santis M, Malandrini A, et al. Homozygosity and severity of phenotypic presentation in a CADASIL family. *Neurol. Sci.* 2014;35:91–93. doi: 10.1007/s10072-013-1580-9

372. Maksemous N, Smith RA, Haupt LM, Griffiths LR. Targeted next generation sequencing identifies novel NOTCH3 gene mutations in CADASIL diagnostics patients. *Hum. Genomics*. 2016;10:38. doi: 10.1186/s40246-016-0093-z

373. Uyguner ZO, Siva A, Kayserili H, Saip S, Altıntaş A, Apak MY, et al. The R110C mutation in Notch3 causes variable clinical features in two Turkish families with CADASIL syndrome. *J. Neurol. Sci.* 2006;246:123–130. doi: 10.1016/j.jns.2006.02.021

374. Matsumoto H, Tsumoto M, Yamamoto T, Takahashi K, Tahira T, Ugawa Y, et al. [A case of early stage CADASIL showing only dizziness and vertigo with a novel mutation of Notch 3 gene]. *Rinsho Shinkeigaku*. 2005;45:27–31.

375. Oberstein SA, Ferrari MD, Bakker E, van Gestel J, Kneppers AL, Frants RR, et al. Diagnostic Notch3 sequence analysis in CADASIL: three new mutations in Dutch patients. Dutch CADASIL Research Group. *Neurology*. 1999;52:1913–5. doi: 10.1212/WNL.52.9.1913

376. Bentley P, Wang T, Malik O, Nicholas R, Ban M, Sawcer S, et al. CADASIL with cord involvement associated with a novel and atypical NOTCH3 mutation. *J. Neurol. Neurosurg. Psychiatry*. 2011;82:855–860. doi: 10.1136/jnnp.2010.223297

377. Dichgans M, Herzog J, Gasser T. NOTCH3 mutation involving three cysteine residues in a family with typical CADASIL. *Neurology*. 2001;57:1714–7.

378. Vikelis M, Papatriantafyllou J, Karageorgiou CE. A novel CADASIL-causing mutation in a stroke patient. *Swiss Med. Wkly.* 2007;137:323–5. doi: 2007/21/smw-11816

379. De Silva KRD, Gamage R, Dunuwille J, Gunarathna D, Sirisena D, Weerasinghe A, et al. Cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL): A patient from Sri Lanka. *J. Clin. Neurosci.* 2009;16:1492–1493. doi: 10.1016/j.jocn.2009.01.019

380. Dotti MT, Federico A, Mazzei R, Bianchi S, Scali O, Conforti FL, et al. The spectrum of Notch3 mutations in 28 Italian CADASIL families. *J. Neurol. Neurosurg. Psychiatry*. 2005;76:736–8. doi: 10.1136/jnnp.2004.048207

381. Au K-M, Li H-L, Sheng B, Chow T-C, Chen M-L, Lee K-C, et al. A novel mutation (C271F) in the Notch3 gene in a Chinese man with cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy. *Clin. Chim. Acta*. 2007;376:229–232. doi: 10.1016/j.cca.2006.07.022

382. García-Estévez DA, Barros-Angueira F, Navarro C. [CADASIL: brief report on a family with a new p.G296C mutation in exon 6 of the Notch-3 gene]. *Rev. Neurol.* 2010;51:729–32.

383. Golomb MR, Sokol DK, Walsh LE, Christensen CK, Garg BP. Recurrent hemiplegia, normal MRI, and NOTCH3 mutation in a 14-year-old: is this early CADASIL? *Neurology*. 2004;62:2331–2.

384. Oliveri RL, Muglia M, De Stefano N, Mazzei R, Labate A, Conforti FL, et al. A novel mutation in the Notch3 gene in an Italian family with cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy: genetic and magnetic resonance spectroscopic findings. *Arch. Neurol.* 2001;58:1418–22.

385. Nishida Y, Ueda A, Ando Y, Ichikawa T. Skin Biopsy-based Diagnosis of CADASIL with Atypical MRI Findings. *Intern. Med.* 2015;54:537–538. doi: 10.2169/internalmedicine.54.3350

386. Pettersen JA, Keith J, Gao F, Spence JD, Black SE. CADASIL accelerated by acute hypotension. *Neurology*. 2017;88:1077–1080. doi: 10.1212/WNL.0000000000003717

387. Lee SJ, Meng H, Elmadhoun O, Blaivas M, Wang MM-H. Cerebral Autosomal Dominant Arteriopathy With Subcortical Infarcts and Leukoencephalopathy Affecting an African American Man. *Arch. Neurol.* 2011;68:1584. doi: 10.1001/archneurol.2011.781

388. Kim Y-E, Yoon CW, Seo SW, Ki C-S, Kim YB, Kim J-W, et al. Spectrum of NOTCH3 mutations in Korean patients with clinically suspicious cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy. *Neurobiol. Aging*. 2014;35:726.e1-726.e6. doi: 10.1016/j.neurobiolaging.2013.09.004

389. Pradotto L, Azan G, Doriguzzi C, Valentini C, Mauro A. Sporadic vascular dementia as clinical presentation of a new missense mutation within exon 7 of NOTCH3 gene. *J. Neurol. Sci.* 2008;271:207–10. doi: 10.1016/j.jns.2008.04.015

390. del Río-Espínola A, Fernández-Cadenas I, Mendióroz M, Gutiérrez-Agulló M, Fernández MT, Fernández-Morales J, et al. Novel human pathological mutations. Gene symbol: NOTCH3. Disease: CADASIL. *Hum. Genet.* 2010;127:474.

391. Federico A, Bianchi S, Dotti MT. The spectrum of mutations for CADASIL diagnosis. *Neurol. Sci.* 2005;26:117–124. doi: 10.1007/s10072-005-0444-3

392. Arboleda-Velasquez JF, Lopera F, Lopez E, Frosch MP, Sepulveda-Falla D, Gutierrez JE, et al. C455R notch3 mutation in a Colombian CADASIL kindred with early onset of stroke. *Neurology*. 2002;59:277–9.

393. Rice CM, McGuone D, Kurian KM, Love S, Renowden SA, Giffin NJ. Autopsy-confirmed, co-existent CADASIL and multiple system atrophy. *Parkinsonism Relat. Disord.* 2011;17:390–2. doi: 10.1016/j.parkreldis.2011.01.004

394. Ozaki K, Irioka T, Ishikawa K, Mizusawa H. CADASIL with a Novel NOTCH3 Mutation (Cys478Tyr). *J. Stroke Cerebrovasc. Dis.* 2015;24:e61–e62. doi: 10.1016/j.jstrokecerebrovasdis.2014.11.022

395. Lee Y-C, Liu C-S, Chang M-H, Lin K-P, Fuh J-L, Lu Y-C, et al. Population-specific spectrum of NOTCH3 mutations, MRI features and founder effect of CADASIL in Chinese. *J. Neurol.* 2009;256:249–255. doi: 10.1007/s00415-009-0091-3

396. Bianchi S, Scali O, Dotti MT, Pantoni L, Parnetti L, Inzitari D, et al. Gene symbol: Notch3. Disease: CADASIL. *Hum. Genet.* 2005;118:546.

397. Mosca L, Marazzi R, Ciccone A, Santilli I, Bersano A, Sansone V, et al. NOTCH3 gene mutations in subjects clinically suspected of CADASIL. *J. Neurol. Sci.* 2011;307:144–148. doi: 10.1016/j.jns.2011.04.019

398. Ragno M, Pianese L, Morroni M, Cacchiò G, Manca A, Di Marzio F, et al. “CADASIL coma” in an Italian homozygous CADASIL patient: comparison with clinical and MRI findings in age-matched heterozygous patients with the same G528C NOTCH3 mutation. *Neurol. Sci.* 2013;34:1947–1953. doi: 10.1007/s10072-013-1418-5

399. Mazzei R, Conforti FL, Ungaro C, Liguori M, Sprovieri T, Patitucci A, et al. Gene symbol: NOTCH3. *Hum. Genet.* 2007;121:296.

400. Bianchi S, Dotti MT, De Stefano N, Stromillo ML, Federico A. Novel human pathological mutations. Gene symbol: NOTCH3. Disease: CADASIL, exon 2 mutation. *Hum. Genet.* 2007;122:558–9.

401. Oh J-H, Kang BS, Choi JC. CADASIL Initially Presented with a Seizure. *J. epilepsy Res.* 2016;6:104–106. doi: 10.14581/jer.16020

402. Lee JS, Kang C, Park SQ, Choi HA, Sim K-B. Clinical Significance of Cerebral Microbleeds Locations in CADASIL with R544C NOTCH3 Mutation. *PLoS One*. 2015;10:e0118163. doi: 10.1371/journal.pone.0118163

403. Liem MK, Lesnik Oberstein SAJ, Vollebregt MJ, Middelkoop HAM, Grond J, Helderman-van den Enden ATJM. Homozygosity for a NOTCH3 mutation in a 65-year-old CADASIL patient with mild symptoms. *J. Neurol.* 2008;255:1978–1980. doi: 10.1007/s00415-009-0036-x

404. Roy B, Maksemous N, Smith RA, Menon S, Davies G, Griffiths LR. Two novel mutations and a previously unreported intronic polymorphism in the NOTCH3 gene. *Mutat. Res. Mol. Mech. Mutagen.* 2012;732:3–8. doi: 10.1016/j.mrfmmm.2012.02.004

405. Praline J, Limousin N, Vourc’h P, Pallix M, Debiais S, Guennoc A-M, et al. CADASIL and ALS: A link? *Amyotroph. Lateral Scler.* 2009;11:1–3. doi: 10.1080/17482960903033153

406. Bohlega S. Novel mutation of the notch3 gene in arabic family with CADASIL. *Neurol. Int.* 2011;3:e6. doi: 10.4081/ni.2011.e6

407. Pradotto L, Orsi L, Daniele D, Caroppo P, Lauro D, Milesi A, et al. A new NOTCH3 mutation presenting as primary intracerebral haemorrhage. *J. Neurol. Sci.* 2012;315:143–145. doi: 10.1016/j.jns.2011.12.003

408. Rutten JW, Boon EMJ, Liem MK, Dauwerse JG, Pont MJ, Vollebregt E, et al. Hypomorphic NOTCH3 alleles do not cause CADASIL in humans. *Hum. Mutat.* 2013;34:1486–9. doi: 10.1002/humu.22432

409. Saiki S, Sakai K, Saiki M, Kitagawa Y, Umemori T, Murata K, et al. Varicose veins associated with CADASIL result from a novel mutation in the Notch3 gene. *Neurology*. 2006;67:337–339. doi: 10.1212/01.wnl.0000224758.52970.19

410. Pippucci T, Maresca A, Magini P, Cenacchi G, Donadio V, Palombo F, et al. Homozygous NOTCH3 null mutation and impaired NOTCH3 signaling in recessive early-onset arteriopathy and cavitating leukoencephalopathy. *EMBO Mol. Med.* 2015;7:848–58. doi: 10.15252/emmm.201404399

411. Ragno M, Sanguigni S, Manca A, Pianese L, Paci C, Berbellini A, et al. Parkinsonism in a pair of monozygotic CADASIL twins sharing the R1006C mutation: a transcranial sonography study. *Neurol. Sci.* 2016;37:875–881. doi: 10.1007/s10072-016-2497-x

412. Vedeler C, Bindoff L. A family with atypical CADASIL. *J. Neurol.* 2011;258:1888–9. doi: 10.1007/s00415-011-6023-z

413. Viana-Baptista M, Ferreira S, Costa P, Venancio M, Fernandes S, Carvalho F, et al. The spectrum of NOTCH3 mutations in Portuguese patients with CADASIL: implications for diagnostic strategies. In: Cerebrovascular Diseases. Karger Publishers; 2007. p. 75.

414. Pescini F, Bianchi S, Salvadori E, Poggesi A, Dotti MT, Federico A, et al. A pathogenic mutation on exon 21 of the NOTCH3 gene causing CADASIL in an octogenarian paucisymptomatic patient. *J. Neurol. Sci.* 2008;267:170–173. doi: 10.1016/j.jns.2007.10.017

415. Rinnoci V, Nannucci S, Valenti R, Donnini I, Bianchi S, Pescini F, et al. Cerebral hemorrhages in CADASIL: Report of four cases and a brief review. *J. Neurol. Sci.* 2013;330:45–51. doi: 10.1016/j.jns.2013.04.002

416. Valenti R, Bianchi S, Pescini F, D'Eramo C, Inzitari D, Dotti MT, et al. First report of a pathogenic mutation on exon 24 of the NOTCH3 gene in a CADASIL family. *J. Neurol.* 2011;258:1632–1636. doi: 10.1007/s00415-011-5983-3

417. Mizuno T, Muranishi M, Torugun T, Tango H, Nagakane Y, Kudeken T, et al. Two Japanese CADASIL families exhibiting Notch3 mutation R75P not involving cysteine residue. *Intern. Med.* 2008;47:2067–72.

418. Roy B, Maksemous N, Smith RA, Menon S, Davies G, Griffiths LR. Two novel mutations and a previously unreported intronic polymorphism in the NOTCH3 gene. *Mutat. Res. Mol. Mech. Mutagen.* 2012;732:3–8. doi: 10.1016/j.mrfmmm.2012.02.004

419. Uchino M, Hirano T, Uyama E, Hashimoto Y. Cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL) and CADASIL-like disorders in Japan. *Ann. N. Y. Acad. Sci.* 2002;977:273–8.

420. Abramycheva N, Stepanova M, Kalashnikova L, Zakharova M, Maximova M, Tanashyan M, et al. New mutations in the Notch3 gene in patients with cerebral autosomal dominant arteriopathy with subcortical infarcts and leucoencephalopathy (CADASIL). *J. Neurol. Sci.* 2015;349:196–201. doi: 10.1016/j.jns.2015.01.018

421. Pradotto L, Orsi L, Mencarelli M, Caglio M, Lauro D, Milesi A, et al. Recurrent transient global amnesia as presenting symptoms of CADASIL. *Clin. Case Reports*. 2016;4:1045–1048. doi: 10.1002/ccr3.666

422. Bianchi S, Di Palma C, Gallus GN, Taglia I, Poggiani A, Rosini F, et al. Two novel HTRA1 mutations in a European CARASIL patient. *Neurology*. 2014;82:898–900. doi: 10.1212/WNL.0000000000000202

423. Cai B, Zeng J, Lin Y, Lin Y, Lin W, Lin W, et al. A frameshift mutation in HTRA1 expands CARASIL syndrome and peripheral small arterial disease to the Chinese population. *Neurol. Sci.* 2015;36:1387–1391. doi: 10.1007/s10072-015-2121-5

424. Khaleeli Z, Jaunmuktane Z, Beaufort N, Houlden H, Haffner C, Brandner S, et al. A novel HTRA1 exon 2 mutation causes loss of protease activity in a Pakistani CARASIL patient. *J. Neurol.* 2015;262:1369–1372. doi: 10.1007/s00415-015-7769-5

425. Ibrahimi M, Nozaki H, Lee A, Onodera O, Reichwein R, Wicklund M, et al. A CARASIL Patient from Americas with Novel Mutation and Atypical Features: Case Presentation and Literature Review. *Cerebrovasc. Dis.* 2017;44:135–140. doi: 10.1159/000477358

426. Nishimoto Y, Shibata M, Nihonmatsu M, Nozaki H, Shiga A, Shirata A, et al. A novel mutation in the HTRA1 gene causes CARASIL without alopecia. *Neurology*. 2011;76:1353–1355. doi: 10.1212/WNL.0b013e318215281d

427. Wang X-L, Li C-F, Guo H-W, Cao B-Z. A Novel Mutation in the *HTRA1* Gene Identified in Chinese CARASIL Pedigree. *CNS Neurosci. Ther.* 2012;18:867–869. doi: 10.1111/j.1755-5949.2012.00373.x

428. Huang X, Xiao X, Jia X, Li S, Li M, Guo X, et al. Mutation analysis of the genes associated with anterior segment dysgenesis, microcornea and microphthalmia in 257 patients with glaucoma. *Int. J. Mol. Med.* 2015;36:1111–1117. doi: 10.3892/ijmm.2015.2325

429. Garel C, Rosenblatt J, Moutard ML, Heron D, Gelot A, Gonzales M, et al. Fetal intracerebral hemorrhage and *COL4A1* mutation: promise and uncertainty. *Ultrasound Obstet. Gynecol.* 2013;41:228–230. doi: 10.1002/uog.12268

430. Giorgio E, Vaula G, Bosco G, Giacone S, Mancini C, Calcia A, et al. Two families with novel missense mutations in COL4A1: When diagnosis can be missed. *J. Neurol. Sci.* 2015;352:99–104. doi: 10.1016/j.jns.2015.03.042

431. Chen Z, Migeon T, Verpont M-C, Zaidan M, Sado Y, Kerjaschki D, et al. HANAC Syndrome Col4a1 Mutation Causes Neonate Glomerular Hyperpermeability and Adult Glomerulocystic Kidney Disease. *J. Am. Soc. Nephrol.* 2016;27:1042–1054. doi: 10.1681/ASN.2014121217

432. Plaisier E, Gribouval O, Alamowitch S, Mougenot B, Prost C, Verpont MC, et al. COL4A1 mutations and hereditary angiopathy, nephropathy, aneurysms, and muscle

cramps. *N. Engl. J. Med.* 2007;357:2687–95. doi: 10.1056/NEJMoa071906

433. Zenteno JC, Crespí J, Buentello-Volante B, Buil JA, Bassaganyas F, Vela-Segarra JI, et al. Next generation sequencing uncovers a missense mutation in COL4A1 as the cause of familial retinal arteriolar tortuosity. *Graefe's Arch. Clin. Exp. Ophthalmol.* 2014;252:1789–1794. doi: 10.1007/s00417-014-2800-6

434. Patel NH, Bhadarka HK, Patel KB, Vaniawala SN, Acharya A, Mukhopadhyaya PN, et al. Embryo genome profiling by single-cell sequencing for successful preimplantation genetic diagnosis in a family harboring COL4A1 c.1537G>A; p.G513S mutation. *J. Hum. Reprod. Sci.* 2016;9:200–206. doi: 10.4103/0974-1208.192072

435. Plaisier E, Gribouval O, Alamowitch S, Mougenot B, Prost C, Verpont MC, et al. *COL4A1* Mutations and Hereditary Angiopathy, Nephropathy, Aneurysms, and Muscle Cramps. *N. Engl. J. Med.* 2007;357:2687–2695. doi: 10.1056/NEJMoa071906

436. Yoneda Y, Haginoya K, Kato M, Osaka H, Yokochi K, Arai H, et al. Phenotypic Spectrum of *COL4A1* Mutations: Porencephaly to Schizencephaly. *Ann. Neurol.* 2013;73:48–57. doi: 10.1002/ana.23736

437. Livingston J, Doherty D, Orcesi S, Tonduti D, Piechiecchio A, La Piana R, et al. COL4A1 Mutations Associated with a Characteristic Pattern of Intracranial Calcification. *Neuropediatrics*. 2011;42:227–233. doi: 10.1055/s-0031-1295493

438. Corlobe A, Tournier-Lasserve E, Mine M, Menjot de Champfleur N, Carra Dalliere C, Ayrignac X, et al. COL4A1 mutation revealed by an isolated brain hemorrhage. *Cerebrovasc. Dis.* 2013;35:593–4. doi: 10.1159/000351520

439. John S, Jehi L, Manno EM, Conway DS, Uchino K. COL4A1 gene mutation – beyond a vascular syndrome. *Seizure*. 2015;31:19–21. doi: 10.1016/j.seizure.2015.06.014

440. Deml B, Reis LM, Maheshwari M, Griffis C, Bick D, Semina EV. Whole exome analysis identifies dominant *COL4A1* mutations in patients with complex ocular phenotypes involving microphthalmia. *Clin. Genet.* 2014;86:475–481. doi: 10.1111/cge.12379

441. Smigiel R, Cabala M, Jakubiak A, Kodera H, Sasiadek MJ, Matsumoto N, et al. Novel *COL4A1* mutation in an infant with severe dysmorphic syndrome with schizencephaly, periventricular calcifications, and cataract resembling congenital infection. *Birth Defects Res. Part A Clin. Mol. Teratol.* 2016;106:304–307. doi: 10.1002/bdra.23488

442. Sibon I, Coupry I, Menegon P, Bouchet J-P, Gorry P, Burgelin I, et al. COL4A1 mutation in Axenfeld-Rieger anomaly with leukoencephalopathy and stroke. *Ann. Neurol.* 2007;62:177–84. doi: 10.1002/ana.21191

443. Coupry I, Sibon I, Mortemousque B, Rouanet F, Mine M, Goizet C. Ophthalmological Features Associated With COL4A1 Mutations. *Arch. Ophthalmol.* 2010;128:483. doi: 10.1001/archophthalmol.2010.42

444. Tonduti D, Pichiecchio A, La Piana R, Livingston J, Doherty D, Majumdar A, et al. COL4A1-Related Disease: Raised Creatine Kinase and Cerebral Calcification as Useful Pointers. *Neuropediatrics*. 2012;43:283–288. doi: 10.1055/s-0032-1325116

445. Shah S, Ellard S, Kneen R, Lim M, Osborne N, Rankin J, et al. Childhood presentation of

COL4A1 mutations. *Dev. Med. Child Neurol.* 2012;54:569–574. doi: 10.1111/j.1469-8749.2011.04198.x

446. Colin E, Sentilhes L, Sarfati A, Mine M, Guichet A, Ploton C, et al. Fetal intracerebral hemorrhage and cataract: think COL4A1. *J. Perinatol.* 2014;34:75–77. doi: 10.1038/jp.2013.135

447. Shah S, Kumar Y, McLean B, Churchill A, Stoodley N, Rankin J, et al. A dominantly inherited mutation in collagen IV A1 (COL4A1) causing childhood onset stroke without porencephaly. *Eur. J. Paediatr. Neurol.* 2010;14:182–7. doi: 10.1016/j.ejpn.2009.04.010

448. Rouaud T, Labauge P, Lasserve ET, Mine M, Coustans M, Deburghgraeve V, et al. Acute urinary retention due to a novel collagen COL4A1 mutation. *Neurology*. 2010;75:747–749. doi: 10.1212/WNL.0b013e3181eee440

449. Xia X-Y, Li N, Cao X, Wu Q-Y, Li T-F, Zhang C, et al. A novel COL4A1 gene mutation results in autosomal dominant non-syndromic congenital cataract in a Chinese family. *BMC Med. Genet.* 2014;15:97. doi: 10.1186/s12881-014-0097-2

450. Durrani-Kolarik S, Manickam K, Chen B. COL4A1 Mutation in a Neonate With Intrauterine Stroke and Anterior Segment Dysgenesis. *Pediatr. Neurol.* 2017;66:100–103. doi: 10.1016/j.pediatrneurol.2016.04.010

451. Vahedi K, Boukobza M, Massin P, Gould DB, Tournier-Lasserve E, Bousser M-G. Clinical and brain MRI follow-up study of a family with COL4A1 mutation. *Neurology*. 2007;69:1564–8. doi: 10.1212/01.wnl.0000295994.46586.e7

452. Meuwissen MEC, de Vries LS, Verbeek HA, Lequin MH, Govaert PP, Schot R, et al. Sporadic COL4A1 mutations with extensive prenatal porencephaly resembling hydranencephaly. *Neurology*. 2011;76:844–846. doi: 10.1212/WNL.0b013e31820e7751

453. Plancher JM, Hufnagel RB, Vagal A, Peariso K, Saal HM, Broderick JP. Case of Small Vessel Disease Associated with COL4A1 Mutations following Trauma. *Case Rep. Neurol.* 2015;7:142–7. doi: 10.1159/000431309

454. Abe Y, Matsuduka A, Okanari K, Miyahara H, Kato M, Miyatake S, et al. A severe pulmonary complication in a patient with COL4A1 -related disorder: A case report. *Eur. J. Med. Genet.* 2017;60:169–171. doi: 10.1016/j.ejmg.2016.12.008

455. Lichtenbelt KD, Pistorius LR, De Tollenaer SM, Mancini GM, De Vries LS. Prenatal genetic confirmation of a *COL4A1* mutation presenting with sonographic fetal intracranial hemorrhage. *Ultrasound Obstet. Gynecol.* 2012;39:726–727. doi: 10.1002/uog.11070

456. van der Knaap MS, Smit LME, Barkhof F, Pijnenburg YAL, Zweegman S, Niessen HWM, et al. Neonatal porencephaly and adult stroke related to mutations in collagen IV A1. *Ann. Neurol.* 2006;59:504–11. doi: 10.1002/ana.20715

457. Takenouchi T, Ohyagi M, Torii C, Kosaki R, Takahashi T, Kosaki K. Porencephaly in a fetus and HANAC in her father: Variable expression of *COL4A1* mutation. *Am. J. Med. Genet. Part A*. 2015;167:156–158. doi: 10.1002/ajmg.a.36823

458. Leung M, Lewis E, Humphreys P, Miller E, Geraghty M, Lines M, et al. COL4A1 Mutation in a Pediatric Patient Presenting with Post-Ictal Hemiparesis. *Can. J. Neurol. Sci.* 2012;39:654–657. doi: 10.1017/S0317167100015420

459. Bilguvar K, DiLuna ML, Bizzarro MJ, Bayri Y, Schneider KC, Lifton RP, et al. COL4A1 mutation in preterm intraventricular hemorrhage. *J. Pediatr.* 2009;155:743–5. doi: 10.1016/j.jpeds.2009.04.014

460. Hino-Fukuyo N, Kikuchi A, Iwasaki M, Sato Y, Kubota Y, Kobayashi T, et al. Dramatic response after functional hemispherectomy in a patient with epileptic encephalopathy carrying a de novo COL4A1 mutation. *Brain Dev.* 2017;39:337–340. doi: 10.1016/j.braindev.2016.11.006

461. Değerliyurt A, Ceylaner G, Koçak H, Bilginer Gürbüz B, Cihan BS, Rizzu P, et al. A new family with autosomal dominant porencephaly with a novel Col4A1 mutation. Are arachnoid cysts related to Col4A1 mutations? *Genet. Couns.* 2012;23:185–93.

462. Rødahl E, Knappskog PM, Majewski J, Johansson S, Telstad W, Kråkenes J, et al. Variants of Anterior Segment Dysgenesis and Cerebral Involvement in a Large Family With a Novel COL4A1 Mutation. *Am. J. Ophthalmol.* 2013;155:946–953.e2. doi: 10.1016/j.ajo.2012.11.028

463. Ha TT, Sadleir LG, Mandelstam SA, Paterson SJ, Scheffer IE, Gecz J, et al. A mutation in *COL4A2* causes autosomal dominant porencephaly with cataracts. *Am. J. Med. Genet. Part A*. 2016;170:1059–1063. doi: 10.1002/ajmg.a.37527

464. Gunda B, Mine M, Kovács T, Hornyák C, Bereczki D, Várallyay G, et al. COL4A2 mutation causing adult onset recurrent intracerebral hemorrhage and leukoencephalopathy. *J. Neurol.* 2014;261:500–3. doi: 10.1007/s00415-013-7224-4

465. Yoneda Y, Haginoya K, Arai H, Yamaoka S, Tsurusaki Y, Doi H, et al. De novo and inherited mutations in COL4A2, encoding the type IV collagen α2 chain cause porencephaly. *Am. J. Hum. Genet.* 2012;90:86–90. doi: 10.1016/j.ajhg.2011.11.016

466. Philippakis AA, Azzariti DR, Beltran S, Brookes AJ, Brownstein CA, Brudno M, et al. The Matchmaker Exchange: A Platform for Rare Disease Gene Discovery. *Hum. Mutat.* 2015;36:915–921. doi: 10.1002/humu.22858

467. Naba A, Clauser KR, Hoersch S, Liu H, Carr SA, Hynes RO. The matrisome: in silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices. *Mol. Cell. Proteomics*. 2012;11:M111.014647. doi: 10.1074/mcp.M111.014647

468. Chun HB, Scott M, Niessen S, Hoover H, Baird A, Yates J, et al. The Proteome of Mouse Brain Microvessel Membranes and Basal Lamina. *J. Cereb. Blood Flow Metab.* 2011;31:2267–2281. doi: 10.1038/jcbfm.2011.104

469. Bellenguez C, Bevan S, Gschwendtner A, Spencer CCA, Burgess AI, Pirinen M, et al. Genome-wide association study identifies a variant in HDAC9 associated with large vessel ischemic stroke. *Nat. Genet.* 2012;44:328–333. doi: 10.1038/ng.1081

470. Saint Pierre A, Genin E. How important are rare variants in common disease? *Brief. Funct. Genomics*. 2014;13:353–361. doi: 10.1093/bfgp/elu025

471. MacArthur DG, Tyler-Smith C. Loss-of-function variants in the genomes of healthy humans. *Hum. Mol. Genet.* 2010;19:R125-30. doi: 10.1093/hmg/ddq365

Conference oral presentations

2017 **Screening for Monogenic Cerebral Small Vessel Disease using Whole Genome Sequencing**

- 3rd European Stroke Organisation Conference, 2017, Prague, Czech Republic
- Annual Meeting of the Association of British Neurologists, 2017, Liverpool, UK

2016 **CADASIL: Migraine, Encephalopathy, Stroke and their Inter-relationships**

2015

- 2nd European Stroke Organisation Conference, 2016, Barcelona, Spain
- Annual Meeting of the Association of British Neurologists, 2015, Harrogate, UK

Publications

1. **Tan RYY**, Markus HS (2018) Genetics and Genomics of Stroke. in D.Kumar, P.Elliot, eds. Cardiovascular Genetics and Genomics for Clinicians. Springer.
2. **Tan RYY,** Traylor M, Rutten-Jacobs L, Markus HS (2017) New Insights into Mechanisms of Small Vessel Disease Stroke from Genetics. Clinical Science. doi:10.1042/CS20160825
3. Brookes R, Hollocks M, **Tan RYY**, Markus HS (2016) Brief Screening of Vascular Cognitive Impairment in Patients with CADASIL without Dementia. Stroke. doi:10.1161/strokeaha.116013761
4. **Tan RYY,** Markus HS (2016) CADASIL: Migraine, Encephalopathy, Stroke and their Inter-relationships. PLoS One. doi:10.1371/journal.pone.0157613
5. **Tan RYY**, Markus HS (2015) Monogenic Causes of Stroke: Now and the Future. J Neurol. doi:10.1007/s00415-015-7794-4

Awards and Travel Grants

2018 **Young Investigator Award**

European Stroke Organisation Conference 2018

2017 **Association of British Neurologists (ABN) Bursary**

For attendance at the 2017 Annual Meeting of the ABN

Guarantors of Brain Travel Grant

For attendance at the 3rd European Stroke Organisation Conference

2016 **Fitzwilliam Trust Research Fund Graduate Student Travel Award**

For attendance at the 20th International Stroke Genetics Consortium Workshop

2016 **E.G. Fearnside Fund Travel Grant**

For attendance at the 2nd European Stroke Organisation Conference

2015, 2016 **Fitzwilliam College Senior Scholarship**

Awarded for first class work