STRATEGIES TOWARDS UNDERSTANDING THE GENETICS OF ISCHEMIC STROKE

SUNAINA YADAV

IMPERIAL COLLEGE LONDON

DEPARTMENT OF MEDICINE

DOCTOR OF PHILOSOPHY

NOVEMBER 2013

ABSTRACT

Stroke is the second largest cause of death and disability in the world and 80% of all strokes are ischemic in nature. While most risk factors have been investigated in depth, others such as South Asian ethnicity, long term blood pressure variability and symptomatic carotid stenosis remain largely understudied. In order to understand the genetic etiology underlining ischemic stroke, it is important to study the genetic burden shouldered by these phenotypes as well.

This thesis examined the genetics of ischemic stroke in the presence of the above risk factors using three independent lines of investigation: a) literature based meta-analysis, b) candidate gene based approach and c) genome-wide association study.

Using a literature based meta-analysis comprising 2529 ischemic stroke cases and 2881 healthy controls, genetic risk variants associated with ischemic stroke in South Asians were investigated. Genes *PDE4D SNP 83, ACE I/D* and *IL10 G1082A* were associated with South Asian ischemic stroke risk, with no major differences in strength of association for the same susceptibility genes in other ethnic groups.

A candidate gene study was conducted using 8295 ischemic stroke cases and 12722 controls of European ancestry to test the association of GWAS-derived blood pressure variability associated cluster of 17 *NLGN1* intronic SNPs. The study did not confirm the risk association of *NLGN1* with ischemic stroke.

Finally, a genome-wide association study was conducted to identify novel gene variants associated with ≥ 50% carotid stenosis in ischemic stroke, using 1164 cases and 13,703 healthy controls from seven independent cohorts. Three genetic loci at *LRIG1*, *ROBO1* and *CAPN7* were found to be associated at genome-wide significance.

The findings of this thesis provide new insights into the genetic mechanisms underpinning ischemic stroke, by examining genetic variants associated with unusual stroke related phenotypes. Future directions include replication of sentinel SNPs in larger study populations and a GWAS for South Asian ischemic stroke cases.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACRONYMS & ABBREVIATIONS

All ischemic strokes (**IS**)

American Heart Association (**AHA**)

Angiotensin-converting enzyme (**ACE**)

Anglo-Scandinavian Cardiac Outcome Trial (**ASCOT IR-UK cohort**)

Apolipoprotein E (**APOE**)

Arachidonate 5-lipoxygenase-activating protein (**ALOX5AP**)

ASCOT Blood Pressure Lowering Arm (**ASCOT-BPLA**)

Atherosclerosis Risk in Communities (**ARIC**)

Atrial Fibrillation (**AF**)

Australian Stroke Genetics Collaborative (**ASGC**)

Beta regression coefficient (**β**)

Bio-Repository of DNA in Stroke (**BRAINS**)

Body mass index (**BMI**)

Calpain Like Protease 7 gene (**CAPN7**)

Cardioembolism (**CE**)

Causative Classification System (**CCS**)

Cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (**CADASIL**)

Cerebral autosomal recessive arteriopathy with subcortical infarcts and leukoencephalopathy (**CARASIL**)

Cerebral blood flow (**CBF**)

Cerebral Venous Thrombosis (**CVT**)

Cerebrovascular Disease (**CVD**)

Chi Square Statistic (**Q**)

Chromosome (**Chr**)

Comprehensive Meta-Analysis *v*2.0(**CMA**)

Computer tomography (**CT**)

Confidence Interval (**CI**)

Copy number variants (**CNV**)

Coronary artery disease (**CAD**)

Coronary heart disease (**CHD**)

Degrees of freedom (**df**)

Deoxyribonucleic acid (**DNA**)

Disability adjusted years (**DALY's**)

Diastolic Blood Pressure (**DBP**)

Echocardiography (**ECG**)

Epidermal growth factor (**EGF**)

Female (**F**)

Framingham Heart Study (**FHS**)

Genes Associated with Stroke Risk and Outcomes Study (**GAROS**)

Genetics of Early Onset Stroke (**GEOS**)

Genome Wide Association study (**GWAS**)

Genome-wide complex trait analysis (**GCTA**)

Genotyped SNP (**G**)

Globotriaosylceramide (**GL-3**)

Hardy Weinberg Equilibrium (**HWE**)

Histone deacetylase 9 (**HDAC9**)

Homocysteine (**Hcys**)

HtrA serine protease 1 (**HTRA1**)

Hypertension (**HTN**)

Imputation metric (r^2)

Imputed SNP (**I**)

Instrumental variable (**IV**)

International Stroke Genetics Consortium (**ISGC**)

Intimal medial wall thickening (**IMT**)

Ischemic Stroke Genetics Study (**ISGS**)

Large vessel disease (**LVD**)

Large-artery atherosclerosis (**LAA**)

Leucine-rich repeats and immunoglobulin-like domain 1 (**LRIG 1**)

Magnetic resonance angiography (**MRA**)

Magnetic resonance imaging scans (**MRI**)

Male (**M**)

Mean arterial pressure (**MAP**)

Methylenetetrahydrofolate reductase (**MTHFR**)

Minor Allele Frequency (**MAF**)

Mitochondrial myopathy, encephalopathy, lactic acidosis and stroke-like episodes (**MELAS**)

Modified Rankin Scale of Disability (**MRS**)

Myocardial Infarction (**MI**)

National Blood Service (**NBS**)

National Health Service (**NHS**)

National Human Genome Research Institute (**NHGRI**)

National Institute of Neurological Disorders and Stroke (**NINDS**)

National Institutes of Health Stroke Scale (**NIHSS**)

National Stroke Association (**NSA**)

New England Medical Centre Posterior Circulation Registry (**NEMC-PCR**)

Next generation sequencing (**NGS**)

Non-communicable diseases (**NCD's**)

Non-Contrast Computer tomography (**NCCT**)

Non-Risk Allele (**NAR or A2**)

Not Reported (**NR**)

Odds ratios (**OR**)

Oxford Vascular study (**OXVASC**)

Phosphatase and actin regulator 1 (**PHACTR1**)

Phosphodiesterase 4D (**PDE4D**)

Pituitary homeobox 2 (**PITX2**)

Plasminogen activator inhibitor-1 (**PAI**)

Population attributable risk (**PAR**)

Principal component analysis (**PCA**)

Probability value (**p**)

Pulse Pressure (**PP**)

Quantile-quantile (**Q-Q**)

Risk allele (**RA or A1**)

Risk allele frequency (**RAF**)

Risk Ratio (**RR**)

Roundabout, axon guidance receptor, homolog 1 gene (**ROBO 1**)

Sibling TDT (**S-TDT**)

Siblings with Ischemic Stroke Study (**SWISS**)

Single nucleotide polymorphisms (**SNPs**)

Small vessel disease (**SVD**)

Standard Deviation (**SD**)

Standard error (**SE**)

Systolic Blood Pressure (**SBP**)

The Atherosclerosis-Small vessel disease-Cardiac disease-Other (Phenotypic) Stroke Classification (**A-S-C-O)**

The Oxfordshire Community Stroke Project (**OCSP**)

The Trial of Org 10172 in Acute Stroke Treatment (**TOAST**)

Tissue plasminogen activator (**tPA**)

Transforming Growth Factor β (**TGF-β**)

Transient ischemic attacks (**TIA's**)

Transmission disequilibrium testing (**TDT**)

Type 2 diabetes and Diabetes melliatus (**T2D**)

Vitamin Intervention for Stroke Prevention trial (**VISP**)

Waist-hip ratio (**WHR**)

Wellcome Trust Case Control Consortium 2 (**WTCCC 2**)

World Health Organization (**WHO**)

Young all ischemic Stroke (**YS**)

Zinc finger homeobox 3 (**ZFHX3**)

α-galactosidase A (**α-gal**)

DECLARATION OF ORIGINALITY

The work presented in this thesis is original and has not been submitted in support of an application for another degree or qualification to any university.

Images used from published literature are clearly referenced.

The ASCOT Genome wide association study (Chapter 3) was conducted by the ASCOT group and the results are reproduced with permission from Dr. Patricia Munroe.

COPYRIGHT DECLARATION

ACKNOWLEDGEMENTS

CHAPTER 1: INTRODUCTION

1.1. Global Impact of Ischemic Stroke

1.1.1 Prevalence

High quality and detailed collection of data on stroke incidence and prevalence are essential for estimating the burden of stroke on the general population and constructing prevention and treatment strategies. Much remains to be discovered about the global distribution of stroke and its association with the prevalence of vascular disease risk factors, demographic and economic characteristics. The World Health Organization is the leading authority on assessing the global burden of stroke and releases updated annual reports on incidence and mortality data consolidated from 14 different regions around the world. According to the WHO, stroke is the third most common cause of death in the developed world behind coronary heart disease and cancer, and around 25% of all strokes afflict those less than 65 years of age. Stroke is also the second largest cause of death in individuals above 60 years of age and fifth largest cause of death in individuals between 15-59 years of age (WHO, 2002). Nearly 15 million individuals suffer from stroke each year of which 5.5 million (10% of all global deaths) die. Stroke related disability counts as the 6th largest cause for reduced DALY's and is expected to rise in the rankings due to the ever-increasing numbers of the elderly in western populations (Murray and Lopez, 1997). Stroke has already risen from the fifth (in 1990) to the third largest (in 2010) cause of global disability adjusted years, accounting for a 19% increase from 1990 (Murray et al., 2012). The WHO predicts that the burden of stroke is projected to rise from around 38 million DALYs globally in 1990 to 61 million DALYs in 2020 (WHO, 2002). Such a huge epidemiological transition of disease incidence will surely lead to an exponential increase in the number of global deaths attributable to stroke alone.

Global disease burden is continuing to shift away from communicable to non-communicable diseases and from premature death to years lived with disability. Stroke falls under the NCD's along with heart disease, chronic lung diseases, cancer and diabetes, which are the major causes of death and disability in the world. NCD's are responsible for over two-thirds or 63% of all global deaths (36 million of 57 million global deaths), 80% (29 million) of which occur in the lower and middle-income countries (WHO, 2010). 29% of all NCD deaths in the developing world occur before the

age of 60 years. Vascular disease is the leading cause of NCD deaths and accounts for nearly 17 million (48%) of mortalities. Stroke contributes to over one-third (5 million) of the vascular NCD deaths and leaves another 5 million people permanently disabled (WHO, 2010)

A 2012 study by the American Heart Association on heart disease and stroke statistics reported an over-all decline in stroke mortality (34.8%) while nearly 795,000 Americans experienced a new or recurrent stroke. Of these stroke cases 77% (610,000) were first attacks, and 23% (185,000) were recurrent attacks. Stroke accounted for 1 in 18 deaths in the United States which amounted to 1 stroke occurrence every 40 seconds (Roger et al., 2012). The incidence of stroke in the United States is reported as ~269 per 100,000 (Williams, 2001) and is higher than their European counterparts where the incidence of stroke ranges between 95 per 100,000 for women and 141 per 100,000 for men (Heuschmann et al., 2009).

The average age adjusted stroke mortality in developed countries is 50-100 per 100,000 (Donnan et al., 2008), although there are vast geographical differences in the incidence of stroke. Estimated projections in 2005 forage-standardized death rates from stroke per 100,000 individuals (aged 30–69 years) for nine selected countries representative of all regions and income groups across the globe were determined (Strong et al., 2007). Geographically Russia, India and China accounted for some of the highest number of deaths from stroke by region with an estimated mortality projection of ~180-100 per 100,000 individuals (**Figure 1.1a**). Middle to low-income countries had a higher mortality rate than developed countries such as the United Kingdom (~20 per 100,000 individuals) and Canada (~15 per 100,000 individuals) and this held true for premature death and years of life lost as well (Strong et al., 2007).

A 100% increase in stroke incidence in low to middle income countries in the past 4 decades (Feigin et al., 2009) is accompanied by reports of a 33% increase in stroke mortality (4 million deaths) (Strong et al., 2007), pointing towards an alarming occurrence of a stroke epidemic (**Figure 1.1b**). This is in contrast to the reduction in stroke incidence in developed and high-income countries where incidence rates have seen a sharp decline of over 42% in the last four decades (Feigin et al., 2009). These findings were later supported by other studies that showed that the rates of stroke

mortality and DALY loss were highest in low income countries such as eastern Europe, north Asia, central Africa, and the south Pacific, and this accounted for a 10-fold difference in mortality rates and DALY loss between the affected and non-affected regions (Johnston et al., 2009).

Prevalence studies from within South Asia (India, Pakistan, Sri Lanka and Bangladesh) are few and far between with widely variable results over 40 years of analysis (44-1000 per 100,000 individuals) (Prasad et al., 2012). Disparities in study population demographics and statistical methodology are probably a source of heterogeneity in these small community based studies leading to variable results. However, it is clear that South Asia shoulders a huge burden for stroke, which is higher than in Europe and United States. With an estimated 62 million diabetics (Shetty., 2012), the largest concentration in the world, 1 million Indians are estimated to suffer from stroke each year (Mehndiratta et al., 2013). The newly established Indian Stroke Association (http://www.indianstrokeassociation.com) is the first national effort to tackle the problem of stroke in India and aims to conduct research at a national level. Large well-powered epidemiological studies are of utmost importance to estimate the true prevalence rates in this region and efforts are underway to improve stroke research in the region (Biswas, 2013, Bhaumik, 2013).

Disparity in the prevalence of stroke across various geographical and ethnic lines could be partly due to the identification and improved control of modifiable risk factors such as HTN and diabetes. However, it is likely that genetic differences between populations may also account for this.
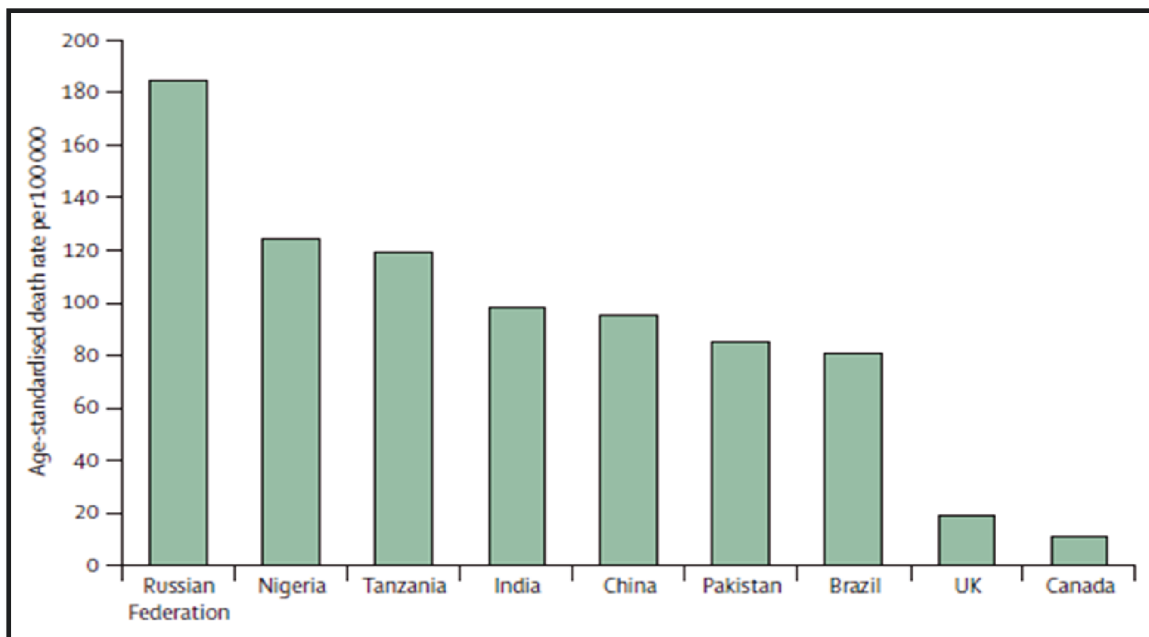
Figure 1.1a: Age-standardized death rates from stroke per 100,000 for ages 30-69 years, selected countries and projections for the year 2005 (Source: Strong et al., 2007).
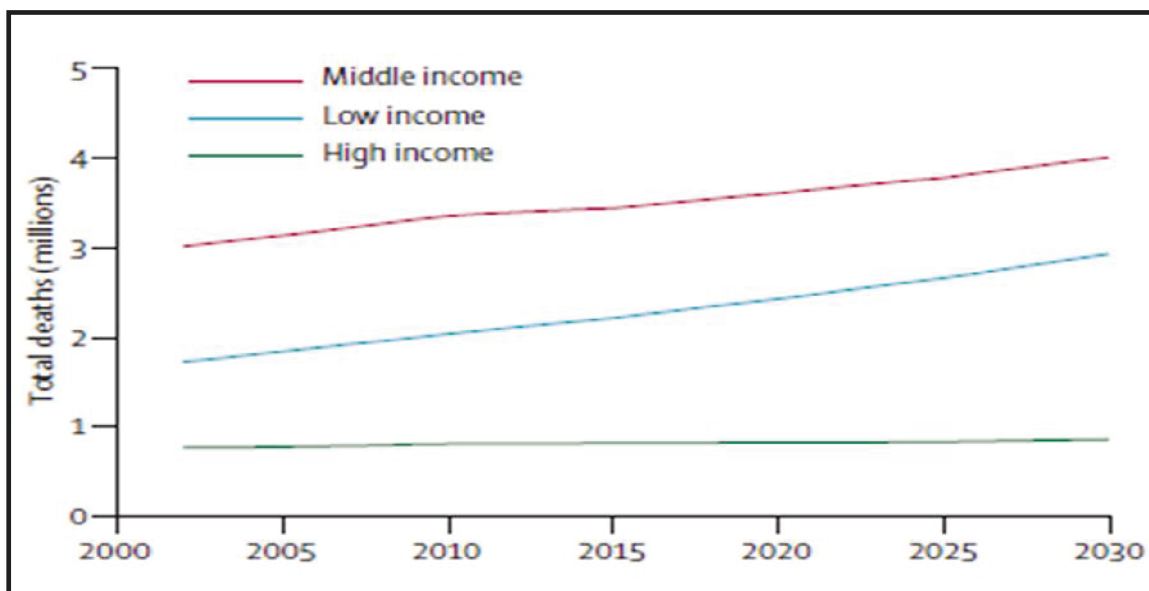
Figure 1.1b: Projected trends for stroke deaths by World Bank income group 2002-30 (Source: Strong et al., 2007).

1.1.2 Socio-economic Burden of Stroke

Preventive stroke strategies and post-stroke care costs global national health institutes millions of dollars each year. Stroke consumes 2-4% of global health care costs and about 4% of direct health care costs in industrialized countries (Donnan et al., 2008).

In 2004 the AHA estimated US stroke costs of $53.6 billion, a rise of $13.6 billion from 1997. This included direct costs for medical care and therapy averaged at $33 billion and indirect costs from lost productivity at $20.6 billion. The updated 2012 AHA report showed a staggering increase in total direct costs (physicians cost, hospital services, prescribed medications, home health care, and other medical durables) and indirect costs (lost productivity resulting from mortality) of CVD and stroke to $297.7 billion. This was higher than the cost for any other disease (Roger et al., 2012). A breakdown of average cost per patient for the first 90 days after a stroke was estimated at $15,000, although 10% of cases cost more than $35,000 (WHO, 2002).

On the other side of the pond, the British NHS spent an estimated £7.6 billion in 1997 on stroke health care costs, which rose considerably by the year 2000 and cost the NHS over 4% of the total health care budget (WHO, 2002). Stroke was estimated to be responsible for 3% of total health care costs in the Netherlands in 1994 with updated average total costs of care per patient for six months following a stroke estimated at €16,000 in 2003 (WHO, 2002). New studies are underway to estimate the socioeconomic impact of stroke in Netherlands. (van Eeden et al., 2012). Additionally an Italian study reflected the average stroke costs across Europe with one-year healthcare and societal costs estimated at €11,747 and €19,953 per stroke survivor respectively (Fattore et al., 2012).

Asian countries are also under severe economic stress related to post stroke health care costs. In Singapore, average hospital costs for stroke were reported in 2000 as $5000 per patient. Ward charges accounted for 38%, radiology 15%, doctors' fees 10%, medications 8%, therapy 7% (WHO, 2002). Annual direct medical costs in Korea were estimated at $8732 per patient with female gender, young age and first stroke associated with increasing costs (Rha et al., 2013). The economic burden of stroke is largely unknown in South Asian countries and no studies on the costs of stroke have emerged from the region. However, the WHO predicts that over the period of 2011-

2025, low and middle-income countries (such as in South Asia) will shoulder 51% of the global costs of NCDs (vascular disease), which amounts to $3.76 trillion (WHO, 2013).

Besides the economic effects of stroke on the world's economies and individual patients, there are several social implications of the disease. Stroke leaves over 60% of the survivors with moderate to severe disability, limiting their ability to gain employment and resulting in a decline in their social functioning (Sreedharan et al., 2013). A 60% loss of personal income (McAllister et al., 2013) and mounting health care costs can lead to a decline in the quality of life for stroke survivors. Stroke survivors are also known to suffer from psychological disorders such as post stroke depression (Chatterjee et al., 2010), which results in greater mortality rates than non-depressive survivors (Paul et al., 2013).

Socio-economic status of stroke patients also affects their propensity to risk factors and mortality. Low economic status affects an individual's ability to access healthcare, take medication and maintain a healthy lifestyle. According to the WHO, individuals from low socio-economic strata of society with a poor educational background are more likely to suffer from vascular diseases and mortality due to heart attacks (WHO, 2002). Due to high poverty and illiteracy rates, this places South Asian countries at a particularly risky juncture.

1.2. Ischemic Stroke

1.2.1 General Definition

The WHO introduced a formal definition for stroke in the 1970's, which is still being used today. WHO describes stroke as "rapidly developing clinical signs of focal (or global) disturbance of cerebral function, lasting more than 24 hours or leading to death, with no apparent cause other than that of vascular origin". This definition was clinical and did not take into account the latest developments in stroke research and imaging technology. Experts across the globe felt the need to update this definition and resultantly the Stroke Council of the AHA/ASA organized a writing group composed of neurologists and other experts to re-examine the modern day definition of stroke (Sacco et al., 2013). They described ischemic stroke as "an episode of neurological dysfunction caused by focal cerebral, spinal, or retinal infarction". Central nervous system infarction was defined as "infarction in the brain, spinal cord, or retinal cell death attributable to ischemia, based on, (a) pathological, imaging, or other objective evidence of cerebral, spinal cord, or retinal focal ischemic injury in a defined vascular distribution; or (b) clinical evidence of cerebral, spinal cord, or retinal focal ischemic injury based on symptoms persisting ≥24 hours or until death, and other aetiologies excluded"(Sacco et al., 2013).

Symptoms of stroke include dysesthesia, dysphasia, paresis or paralysis, dysarthria, ataxia and hemianopsia. Vertigo and headache maybe present but are not typical symptoms of ischemic stroke. Several disability scales such as the Modified Rankin Scale of Disability (van Swieten et al., 1988), Barthel Index (Mahoney and Barthel, 1965) and the National Institutes of Health Stroke Scale (NIHSS) can measure the severity of disability caused by stroke.

Ischemic stroke is a heterogeneous disorder of a complex etiology and 7-40% of all ischemic stroke patients have a TIA. TIA's are a transient episode of neurological dysfunction caused by focal brain, spinal cord, or retinal ischemia, without acute infarction (Easton et al., 2009). They act as a precursor to a full-blown stroke with affects that last less than 24 hours. They act as warning events to an impending stroke and usually cannot be detected by imaging as the symptoms are passing with no visible

tissue damage. TIA's are difficult to diagnose and often confused with other stroke mimics such at vertigo and migraines with aura.

1.2.2 Ischemic Stroke Subtypes

The sub type of ischemic stroke affects its prognosis, outcome, and management. Recent developments in diagnostic and imaging technologies have allowed clinicians to identify differential etiologies underpinning ischemic stroke sub types and molecular mechanisms of ischemic injury. The correct diagnosis of a stroke subtype requires an in depth analysis of clinical phenotypes, imaging and blood biochemistry results, which is imperative to good clinical and research practice. The etiological classification of ischemic stroke can be broadly divided into two categories; causative and phenotypic. Causative sub typing classifies stroke patients into one etiological subtype through a decision-making process that involves the combination of clinical phenotypes, modifiable risk factors, and imaging (Chen et al., 2012). Phenotypic sub typing patient can categorize a patient into more than one etiologic subtype and uses clinical phenotypes and diagnostic test findings (Chen et al., 2012).

1.2.2.1 Phenotypic Sub-Typing of Ischemic Stroke

1.2.2.1.1 Harvard Cooperative Stroke Registry: Before the advent of computed tomography, diagnosis and classification of stroke was primarily based on clinical symptoms of deceased patients. The Harvard Cooperative Stroke Registry was the first computer-based diagnostic program using a prospective published database for stroke sub classification using 694 stroke patients (Mohr et al., 1978). The basis of diagnosis was primarily angiography with only 3% patients with CT scans and 4% with necropsies. The computer-based diagnostic program used clinical data to differentiate between subarachnoid hemorrhage, intracerebral hemorrhage, brain embolism, large artery-related brain infarction, and lacunar infarction (Goldstein RJ, 1976).

1.2.2.1.2 The Stroke Data Bank: Major limitation of the Harvard Cooperative Stroke Registry was that patients were predominantly white. In order to develop a reliable and reproducible diagnostic program, a racially diverse study population was needed. The Stroke Data Bank was an offshoot of the Harvard Cooperative Stroke Registry and was

funded by the National Institute of Neurological Disease and Stroke (Foulkes et al., 1988). Most patients received CT scans.

1.2.2.1.3 The New England Medical Centre Posterior Circulation Registry (NEMC-PCR) was a prospective computerized registry between 1988 and 1996 and was entirely focused on the posterior circulation ischemic stroke (Caplan et al., 2004). The major aims of the registry were to clarify outcomes and the frequency of various stroke mechanisms and vascular lesions, and to understand the relationship of these mechanisms and lesions to the topography of brain infarcts (Caplan, 2011). Brain imaging was performed on all patients and >80% patients had MRI.

1.2.2.1.4 The Atherosclerosis-Small Vessel Disease-Cardiac Disease-Other (Phenotypic) Stroke Classification (A-S-C-O) registry was set up by leaders in the field of stroke from across Europe (Amarenco et al., 2009). Like the NEMC-PCR, A-S-C-O was limited to ischemic stroke. Clinical phenotypes were used for a variety of purposes, such as describing patient characteristics in therapeutic trials, grouping patients in epidemiological studies, phenotyping in genetic studies, and classifying patients for therapeutic decision-making in daily practice (Caplan, 2011).

1.2.2.1.5 The Oxfordshire Community Stroke Project (OCSP) was a prospective study of all new cases of stroke and transient ischemic attack in a defined population of about 103,000 patients registered with 49 general practitioners across rural and urban Oxfordshire (OCSP, 1983). OCSP was based solely on the clinical symptoms of the patients, or in combination with brain imaging. OCSP addressed the severity and outcome of the stroke but not the causes (Paci et al., 2011) and hence was plagued with several limitations such as the inability to establish a specific brain infarction site, failure to investigate confounding stroke risk factors and inaccurate discrimination between lacunar and small-volume cortical infarcts (Asdaghi et al., 2011).

1.2.2.2 Causative Sub-Typing of Ischemic Stroke

1.2.2.2.1 The Trial of Org 10172 in Acute Stroke Treatment is the most commonly used stroke classification system. TOAST was a placebo-controlled, randomized, blinded study of the low-molecular-weight heparinoid given to patients within 24 hours after stroke (Adams et al., 1993). A classification system was developed for the diagnosis of subtype of ischemic stroke that uses components of existing diagnostic schemes. The

classification system is based on clinical phenotypes and findings from imaging. The TOAST classification has five subtypes of ischemic stroke: large-artery atherosclerosis, cardio embolism, small-vessel occlusion, stroke of other determined etiology, and stroke of undetermined etiology (Adams et al., 1993). Stroke of undetermined etiology is further divided into three categories: two or more causes identified, negative evaluation and incomplete evaluation (Adams et al., 1993).

Large-Artery Atherosclerosis

Large-artery atherosclerosis is usually accompanied by stenosis > 50% or occlusion of an intracranial or extra cranial artery. It is the most common cause of ischemic stroke (including TIA's) and atherosclerotic lesions are commonly found in the carotid artery bifurcation where the artery divides into the external and internal carotid arteries. Clinical manifestations of LAA include large cerebellar lesions and brain stem or sub cortical hemispheric infarcts greater than 1.5 cm in diameter on CT or MRI, cerebral cortical impairment and brain stem dysfunction. A history of TIA's in the same vascular region lends support to the clinical diagnosis (Adams et al., 1993).

Cardio Embolism

Cardio embolic strokes account for nearly 30% of all ischemic strokes. Patients categorized, as having a cardio embolic stroke must have arterial occlusions presumably due to an embolus arising in the heart, with at least one validated cardiac source for an embolus for a possible or probable diagnosis. Clinical and imaging results are similar to LAA with evidence of recurrent TIA supporting the diagnosis (Adams et al., 1993).

Small-Artery Occlusion

Patients with small artery occlusion have lacunar infarcts of <1.5 cm in diameter and a history of HTN and diabetes. Potential sources of cardiac emboli must not be present and stenosis of >50% in an ipsilateral artery should be absent (Adams et al., 1993).

Stroke of Other Determined Etiology

Patients under this category have non-atherosclerotic, non-cardio embolic and rare causes of stroke, such as vasculopathies, hypercoagulable states, hematologic disorders or monogenic disorders. Clinical manifestation and brain imaging results are similar to

an acute ischemic stroke regardless of the size or location of the lesion. Rare causal factors are confirmed via blood tests or arteriography (Adams et al., 1993).

Stroke of Undetermined Etiology

Undetermined etiology refers to a cause of a stroke that cannot be determined with any degree of certainty despite extensive evaluation. This category also includes patients with two or more potential causes of stroke, however the clinician is unable to pin either as the casual sub type (Adams et al., 1993).

1.2.2.2.2 The Causative Classification of Stroke System is a web-based diagnostic program with a questionnaire-style classification scheme for ischemic stroke (http://ccs.martinos.org) (Ay et al., 2007). CCS was developed by Harvard clinicians who wished to improve on the TOAST classification system by incorporating technological advances in stroke imaging, in particular diffusion-weighted MRI and MRA (Lee et al., 2000). The CCS is primarily a redefined TOAST classification system, which is based on the five categories of ischemic stroke (Adams et al., 1993). Improvements to this classification system include revised definitions of lacunar lesions and more sub categories under the 'stroke of undetermined etiology' and 'other causes' category (Ay et al., 2007). The CCS has high inter and intra rater reliability and offers clinicians and researchers the advantage of usability from any web browser. It also allows researchers to have individual components of the stroke work-up so that they can reorganize the data according to the needs of their research (Ay et al., 2007).

1.2.3 Stroke Diagnosis and Imaging Modalities

Stroke is diagnosed by visible cerebral tissue damage and its symptoms are not specific and can be caused by other neurological diseases as well. The small window of treatment warrants accurate assessment and rapid treatment within hours of symptom onset (Wardlaw et al., 1997). The diagnostic unpredictability of stroke results in a lot of stroke mimics to be wrongly administered reperfusion therapy such thrombolysis, which is associated with hemorrhagic complications (Khaja and Grotta, 2007). Therefore clinical diagnosis of stroke is of utmost importance and imaging is the tool of choice to exclude mimics of ischemic stroke.

Non-Contrast Computer tomography scanning is the most popular imaging modality for stroke as it is widely available, cheaper than MRI and is used to exclude acute intracranial hemorrhage (Davis et al., 2006, Radhiana et al., 2013). Although NCCT has been useful in the early detection of ischemic stroke, it lacks sensitivity and accuracy even in the most robust of studies, which report a maximum sensitivity of 50% (Mohr et al., 1995).

Diffusion-weighted MRI uses the principle of Brownian motion and detects the movement of water protons in cerebral tissue, which occurs within few minutes of the ischemic insult (Davis et al., 2006). DWI has proved to be a more sensitive method of detecting early ischemic infarction with detection rates of 46% as compared to only 10% by CT (Chalela et al., 2007).

Magnetic resonance angiography is another imaging technique used to study cervical and cephalic large arteries. The diagnostic usefulness of combined DWI and MRA in characterizing early ischemic stroke sub types has been suggested (Lee et al., 2000).

1.2.4 Stroke Pathophysiology

The brain does not have any stored energy reserves and relies completely on the continuous supply of nutrients and oxygen from the cerebral blood flow (Jones et al., 1981). A 50% decrease in normal CBF (50–60 mL/100g/min) leads to interruption of synaptic transmissions, failure of the membrane and ionic gradients and eventually the advent of an apoptotic cascade that ends in neuronal death (Lewandowski and Barsan, 2001). The CBF in the ischemic penumbra ranges from 25% to 50% of the normal blood flow and sufficient to maintain tissue viability for a few hours. The infarct core has less than 25% blood flow and these results in anaerobic metabolism, acidosis, and pan-necrosis, requiring almost immediate reperfusion to recover (Astrup et al., 1981). The size of the infarct is dependent on the extent of collateral blood supply.

A review by Zivin et al describes three processes involved in the development of pathological symptoms of stroke (Zivin and Choi, 1991). First, ischemic neurons release high toxic concentrations of glutamate by its action on post-synaptic receptors, which causes the intra-cellular build-up of calcium and sodium ions. Accumulation of calcium ions causes the initiation of a toxic cascade and reverses osmosis of water from the extracellular to intracellular compartment forming cytotoxic edema. The intracellular calcium activates enzymes that degrade many intracellular components and the release of free radicals, which damage cell membranes (Zivin and Choi, 1991). Several other cellular and molecular pathways also contribute to neuronal death through ischemia (Won et al., 2002).

1.3. Heritability of Stroke

90% of the population attributable risk for stroke rests with ten conventional stroke risk factors including hypertension, atrial fibrillation, cigarette smoking, diabetes mellitus and obesity (O'Donnell et al., 2010). Management of these risk factors offers the exciting possibility of near complete elimination of stroke. However, stroke risk extends well beyond the boundaries of these risk factors and the disparity in stroke prevalence within a population that is uniformly exposed to environmental risk factors suggests that some other unknown mechanisms are at play. Some of this phenotypic variability has been attributed to genetic differences, with familial patterns of inheritance also lending support.

In the following sections we provide evidence from observational and experimental studies supporting the heritability of stroke.

1.3.1 Observational Studies

1.3.1.1 Family Studies

Positive parental history of stroke can predict the risk of stroke independent of confounding risk factors such as HTN, diabetes and smoking. Although the transmission of stroke in families does not always follow a classic mode of genetic inheritance, it is clear that stroke tends to run in families. An inherited component is likely, although it would not be surprising that a common disease like stroke is often found within families as family members age. Family studies contribute to understanding the heritability of stroke and also serve as a clinically useful risk marker of an individual's susceptibility to stroke, which may serve as a useful tool in stroke diagnosis.

In 1987 a study conducted by Welin et al analyzed parental death from stroke and other potential risk factors in relation to the incidence of stroke among 789 men born in 1913 (54 years old at the base-line examination). They demonstrated that during 18.5 years of follow-up, 7.2% of the subjects suffered from strokes. Subjects with maternal stroke had a 3-fold increase in the risk of stroke as compared with men without such a maternal history (Welin et al., 1987).

Several familial studies on stroke were published in the 1990's. In 1997 Liao et al examined the prevalence of stroke in 3168 individuals of African American and European American ancestry from the Family Heart Study, a National Heart, Lung, and Blood Institute (Liao et al., 1997). The study showed that the age, ethnicity, and sex-adjusted stroke prevalence rates were higher in individuals with a positive family history as compared to individuals without any positive familial history. The age, ethnicity, and sex-adjusted ORs for stroke were larger for individuals with a positive paternal history (OR 2.00; 95% CI 1.13-3.54) compared to a positive maternal stroke history (OR 1.41; 95% CI 0.80-2.50) (Liao et al., 1997). The results remain unchanged when adjusted for additional confounders such as cholesterol, cigarette, coronary heart disease, HTN, and diabetes.

Another prospective follow-up study examined 14,371 men and women with a positive parental history of stroke before the age of 60 years (Jousilahti et al., 1997). The risk ratio of stroke after adjustment for age, smoking, blood pressure, cholesterol, diabetes, and education associated with a positive parental history of stroke was similar in men and women (RR 1.89 and 1.80) and was stronger among subjects aged 25 to 49 (Jousilahti et al., 1997).

A case-control study in 470 cases and 477 controls (40-85 years) further examined the family history of stroke in first-degree relatives as an independent risk factor for all stroke types. The study established no difference in paternal or maternal stroke, although the results were minimally significant for maternal inheritance. Authors speculated that genetic factors, in combination with environmental risk factors might play role in heritability and eventual manifestation of stroke in families (Caicoya et al., 1999).

In 2000 a study published by the ARIC cohort investigated the association of parental history of stroke with subclinical or clinical stroke and also examined the role of confounding risk factors (Morrison et al., 2000). Subclinical stroke refers to asymptomatic stroke that is detected by chance by neuroimaging and exists in individuals with AF and carotid disease (Yatsu and Shaltoni, 2004). Parental history of stroke was significantly associated with silent or subclinical stroke after adjusting for

age, gender, race, BMI, WHR and blood pressure. (OR1.64, 95% CI1.20-2.24) but the same was not true for clinical stroke (Morrison et al., 2000).

The OXVASC study of patients with ischemic stroke or transient ischemic attack, examined the relation between the sex and phenotype of 865 affected probands and history of stroke in mothers, fathers, and siblings. The study revealed gender differences in stroke subtypes, with women less likely than men to have small-vessel disease or large artery atherosclerosis and more likely to have a stroke of unexplained etiology. Maternal stroke was more common than paternal stroke and the maternal excess of stroke was concentrated only in female stroke patients (Touze and Rothwell, 2007). The initial analysis was followed up by a meta-analysis of 18 studies including 7941 patients and supported the findings of the original study. The meta-analysis revealed that women with stroke were more likely than men to have a parental history of stroke. The authors argued for both genetic and non-genetic processes to account for this phenomenon (Touze and Rothwell, 2008).

A Swedish study conducted by the Lund Stroke Register study demonstrated that the prevalence of stroke or TIA was higher among first-degree relatives of stroke patients (12.3%) than among first-degree relatives of control subjects (7.5%) (OR 1.74, 95% CI 1.36-2.22). This could be partly explained by the high prevalence of HTN amongst first-degree relatives of stroke patients as compared to relatives of controls; however authors suggested that it might also indicate a heritable component of HTN (Lindgren et al., 2005).

The Framingham Heart Study provided important evidence on the role of genetic risk factors in stroke risk in the offspring of stroke parents. In the study, 3443 stroke-free FHS offspring participants with confirmed stroke parents (by 65 years of age) were followed up for 8 years after baseline examination and the incident of stroke ascertained (Seshadri et al., 2010). Parental stroke was associated with an increased risk of incident stroke of the same type in the offspring with a 3-fold increase in the risk of stroke.

Most family studies suggest that the genetic liability is greater in individuals aged younger than 70 years (Flossmann et al., 2004, Jood et al., 2005) and varies with stroke subtype (Schulz et al., 2004, Jerrard-Dunne et al., 2003). Family history of stroke has

been shown to be an independent risk factor for both LVA (OR 2.93; 95% CI 1.68-5.13) and SVD (OR 3.15; 95% CI 1.81-5.50) for stroke survivors < 65 years of age but was not associated with CE stroke or stroke of undetermined etiology (Jerrard-Dunne et al., 2003). Case-control studies suggested a 76% increase in the risk of ischemic stroke in the presence of a family history of stroke (Flossmann et al., 2004), although not all reports have demonstrated a positive relationship with family history (Flossmann and Rothwell, 2006) possibly due to confounding factors such as blood pressure (Flossmann and Rothwell, 2005).

1.3.1.2 Twin Studies

Twins provide a unique model to study genetic effects in a shared environment in shaping individual behavior. Monozygotic twins are 100% genetically alike whereas dizygotic twins share 50% of their genetic material. In the classic twin model, the difference between the within-pair variances of dizygotic and monozygotic twins is used to calculate a maximum likelihood estimate of the genetic variation within dizygotic twins. If genes influence the prevalence of stroke, there should be a greater concordance rate for stroke among monozygotic twin pairs than dizygotic pairs. Over the years, twin studies looking at concordance rates for a disease or phenotypic trait between monozygotic and dizygotic twins have been widely used to establish the heritability for complex disorders such HTN, cardiovascular disease and diabetes.

Although twin studies provide the most reliable method to ascertain a genetic component of stroke, they have been particularly challenging to conduct due to the fact that stroke is an age related disorder and recruiting elderly stroke twin pairs is a difficult task. It is also impossible to ascertain the cause of death as being solely due to stroke or compounded by other age-related disorders. Additionally, sample size in twin studies is small and provides no information on stroke sub-types. To date, three twin's studies in stroke have been published with disparity in results. These studies failed to distinguish between stroke subtypes or assess the role of confounding risk factors.

In the first of the twin studies, a Swedish group analyzed a cohort of 10,900 twin pairs but failed to establish a significant difference in stroke mortality concordance rates between monozygotic and dizygotic twins(de Faire U, Friberg L 1975).

A study of 9,475 male twins born between 1917 and 1927 from the Twin Registry maintained by the National Academy of Sciences-National Research Council showed a five-fold greater prevalence of stroke in monozygotic twins and a four-fold increase in stroke risk in monozygotic twins as compared to dizygotic twins (RR 4.3, p<0.05) (Brass et al., 1992).

Similar results were observed by another study on 869 same-sex twins from the Danish Twin Register, which demonstrated that the observed risk of stroke death was 2.3 times greater in monozygotic twins as compared with dizygotic twins after adjustment with age and gender (Bak et al., 2002). Male gender and increasing age were also associated with stroke death. Monozygotic twins had a 1.5 times higher risk of stroke hospitalization compared with dizygotic twins after adjustment for sex and age although the difference was not statistically significant (p=0.11). Heritability for stroke death and stroke hospitalization in twins was estimated at 32% (95% CI, 0.04-0.47) and 17% (95% CI, 0.00 to 0.44) respectively.

Twin studies have shown that a positive family history is a risk factor for stroke and monozygotic twins are more likely to be concordant than dizygotic twins (Flossmann et al., 2004).

1.3.1.3 Sibling Studies

Familial and twin studies have effectively established the role of genetics in the increased risk of ischemic stroke. Unlike twins who are genetically identical, siblings have a common genetic element, which biologically predisposes them to death and disease. Susceptibility due to similar environmental exposures can lead to increased risk of mortality from stroke, especially if one sibling has already died from stroke.

A recent Swedish study examined this hypothesis in 30,735 exposed study participants whose sibling had suffered or died due to stroke. Individuals, who had a sibling with prior stroke, possessed a 60% increased risk for ischemic stroke (RR 1.61, 95% CI 1.48–1.75, p=0.001). Familial risk was higher in full siblings (RR 1.64, 95% CI 1.50–1.81, p=0.001) as compared to half siblings (RR 1.41, 95% CI 1.10–1.82, p=0.007) suggesting that greater genetic similarity leads to a higher disease burden. The age of stroke in siblings was also important with familial risk of early ischemic stroke almost doubling

when exposed to early ischemic stroke (RR 1.94, 95% CI 1.41–2.67, p=0.001) (Kasiman et al., 2012).

In another interesting study, researchers found that individuals who had experienced the loss of a sibling had an increased risk of fatal stroke in the second or third year since death (Rostila et al., 2013). Bereavement was suggested as a cause of stroke inducing stress but doesn't rule out the role of a shared genetic predisposition or environmental determinants. In order to dissect causation from confounding it is essential to examine pairs of siblings that died of discordant causes as compared to the same specific cause (e.g. stroke)(Rostila et al., 2013).

Another study on Mexican Americans found a doubling in stroke risk as compared to the national stroke prevalence estimates for siblings of ischemic stroke/TIA cases. A positive family history of stroke could be utilized to identify high-risk individuals within the Mexican American population (Lisabeth et al., 2008).

Other studies have examined sibling pairs for specific stroke sub types via the presence of phenotypic markers such as microangiopathic brain lesions in siblings of probands with lacunar infarct and disproved the hypothesis that the subtype of ischemic stroke in a proband could be a determinant of the stroke subtype in the respective sibling (Leistner et al., 2008). It is likely that genetic risk factors for different stroke sub types differ and may not be specific for one subtype.

Stroke risk factors have also been analyzed in siblings. Elevated plasma Hcys levels in hypertensiveprobands (18.96±8.08 μmol/L) as compared to their normotensive siblings (14.84±5.55 μmol/L) have been observed (Jain et al., 2003). Increased risk of HTN to siblings of stroke patients is also known (Nicolaou et al., 2000). Studies have also identified a familial predisposition to stroke in sickle cell anemia (Driscoll et al., 2003).

1.3.1.4 Migration Studies

Migration studies are useful for identifying changes in environmental causes of common diseases and help reveal disease risk factors, which may not be readily visible in a general population. Besides comparing existing risk factors between migrants and native populations, such studies can also identify novel risk factors. Although non-modifiable risk factors of stroke such as gender and genetics do not change with migration, there can be dramatic changes in environmental parameters that can cause additional stress. Changes in life style such as diet and physical activity can cause an increase in stroke risk factors such as HTN (Poulter et al., 1990), diabetes and high cholesterol, leading to increase in stroke incidence (Kusuma et al., 2009). Alternately, migration can lead to a decline in mortality (Hammar et al., 2002).

In a large study based in India, Ebrahim et al examined the effect of rural-to-urban migration on the increase of stroke risk factors such as blood pressure, BMI, WHR, cholesterol and glucose (Ebrahim et al., 2010). The study showed that the rural male population had lower blood pressure, lipids, and fasting blood glucose than urban and migrant men, whereas no differences were seen in women. Within the migrant population, recent migrant men had slightly lower prevalence of both diabetes and obesity than long stay migrants.

A 30-year follow-up study examined patterns of morbidity from stroke in Finnish migrants in Sweden. Finnish migrants were younger at the time of stroke and carried a greater mortality risk than Swedish natives. The greater length of residence in Sweden lowered risk of stroke mortality. The authors argued that a difference in the Finnish lifestyle and an inherent tendency for high blood pressure might underline the higher risk of mortality in the migrants (Albin et al., 2013).

Although most evidence supports the increase in stroke risk factors in migrants, some studies contradict these findings. A large population-based matched Canadian cohort study showed that new immigrants appeared to be at lower risk of premature acute stroke than long-term residents. The incidence rate of acute stroke was greater (1.69 per 10,000 person-years) in new immigrants as compared to long stay immigrants (2.56 per 10,000 person-years) (Saposnik et al., 2010). Disparity in age-standardized mortality rate of immigrants to Canada when compared to the overall population has

been demonstrated before (DesMeules et al., 2005). Similar results for immigrants to the United States have also been observed (Lanska, 1997).

1.3.1.5 Adoption Studies

Large-scale population based adoption studies in stroke are scarce. However Sorensen et al have conducted several adoption studies for risk factors of stroke such as obesity (Stunkard et al., 1986) and smoking behavior (Osler et al., 2001). As early as 1988 it has been known that premature death in adult adoptees (aged 16 and 58 years) has a strong genetic burden for death due to CVD (Sorensen et al., 1988). An adoptee whose biological parent died due to cardiovascular and cerebrovascular causes before the age of 50 had a RR of death of 4.52 times (95% CI 1.32-15.4) (Sorensen et al., 1988). A similar study conducted on Danish adoptees came to the same conclusions i.e. there is a moderate genetic burden on the risk of dying prematurely in adulthood from vascular causes if the biological parents has died of a similar cause (Petersen et al., 2002).

Although adoption studies are useful in dissecting the genetic burden away from shared familial environment, the circumstances leading to adoption and living as an adoptee may introduce bias (Osler et al., 2001). Studies have shown that adoptees have an increased all-cause mortality compared to the general population and this may bias the studies with regard to the relative influence of genes and environment compared to the general population (Petersen et al., 2010).

1.3.2 Experimental Studies

1.3.2.1 Genome Wide Association Studies

Heritability of stroke has also recently been estimated using data derived from GWAS studies. Genome-wide complex trait analysis, a tool that allows researchers to estimate heritability of a complex phenotypic trait such as stroke using genome wide association data on unrelated individuals and estimates variance explained by all SNPs entered against a phenotypic trait rather than individual SNPs (Bevan et al., 2012, Yang et al., 2010). Using this methodology, heritability estimate for all ischemic stroke was 37.9% but varied considerably by subtype with the greatest associated with large vessel (40.3%) and cardio embolic disease (32.6%) and lowest for small vessel disease (16.1%) (Bevan et al., 2012). The phenotypic heterogeneity between stroke sub-types and differences in etiology could potentially explain the differences in heritability estimates.

These results need validation due to several inherent limitations of the GCTA tool. Firstly GCTA uses only GWAS SNP data to calculate heritability and as a result takes into account the genetic variation of the submitted SNPs (both genotyped and imputed) (Bevan et al., 2012). Such estimates miss heritability due to rare and low frequency variants as well as non-SNP variants such as CNVs. GCTA is also greatly affected by underlying population sub-structure and require stringent control from population stratification using up to 20PCAs(Bevan et al., 2012). Other limitations include the lack of broad-sense heritability which captures gene–environment interactions or epistatic (gene–gene interactions) (Bevan et al., 2012).

1.3.2.2 Linkage Studies (Monogenic Stroke)

Linkage studies are the first line of choice for studying the genetics of stroke and have been very successful in identifying Mendelian diseases using solely the distribution of genotypes and phenotypes within narrowly delimited families to determine the location of disease genes (Devlin and Roeder, 1999). While stroke remains principally a common sporadic disorder, our understanding of monogenic forms of stroke has improved greatly in recent times (Muqtadar and Testai, 2012, Della-Morte et al., 2012). Studies conducted on monogenic stroke provide the most convincing evidence for the genetic etiology of stroke. However these rare forms of stroke account for only a small

percentage of stroke incidence and are not useful for determining the incidence of sporadic or polygenic forms of ischemic stroke. Although the monogenic disorders confer a high individual relative risk they contribute very little to population stroke due to their rarity.

CADASIL

Described by Joutel et al in 1996 (Joutel et al., 1996), CADASIL is a Mendelian form of hereditary small-vessel disease and vascular dementia. Over 100 pathogenic mutations in the *NOTCH3* gene, an evolutionarily highly conserved transmembrane receptor protein regulating cell fate, (Artavanis-Tsakonas et al., 1999) are known to almost always lead to an odd number of cysteine residues in one of the 33 EGF like repeats in the extracellular domain of the Notch3 protein. These mainly missense mutations are thought to result in conformational changes of the Notch3 protein. Mutations have predominately been identified in individuals of European descent, although cases have been found in other populations such as South Asia (Yadav et al., 2013). Recent sequencing studies have shown the association between common variants in the *NOTCH3* gene and increase in the risk of age-related white matter hyperintensities in hypertensives and stroke, suggesting that *NOTCH3* may play an important role in sporadic stroke as well (Ross et al., 2013, Schmidt et al., 2011).

The prevalence of CADASIL is likely underestimated, as clinical suspicion along with laboratory diagnosis is required. There are few prevalence studies, with one registry in Scotland, UK estimating prevalence rate of confirmed CADASIL cases of 1.98/100,000 (Razvi et al., 2005). Genotype-phenotype correlations have been difficult to determine precisely, mainly because of the heterogeneous nature of the mutations although some mutations are associated with a worse prognosis (Monet-Lepretre et al., 2009, Opherk et al., 2004). Adding to this problem, CADASIL-like symptoms have also been observed in patients without *NOTCH3* mutations (Pantoni et al., 2010). Phenotypic differences such as higher volume of white matter hyperintensities have also been observed in patients with mutations in the *NOTCH3* Delta/Serrate/LAG-2 (DSL) ligand-binding domain as compared to patients with mutations outside of the DSL-binding domain (Meschia, 2011).

Studies investigating CADASIL in monozygotic twins with the *NOTCH3* Cys251Tyr mutation demonstrated significant phenotypic differences in the severity of disease. The study hinted at interplay of genes and environment, with the physically inactive-smoking twin suffering a stroke 14 years earlier than the twin who led an active and healthy lifestyle (Mykkanen et al., 2009).

There is no cure for CADASIL targeted specifically at the *NOTCH3* gene and treatment mainly includes management of clinical symptoms.

CARASIL

CARASIL or Maeda syndrome (Fukutake and Hirayama, 1995) is caused by mutations in *HTRA1* gene localized on Chr10q encoding HTRA1 that represses signaling mediated by Transforming Growth Factor β (TGF-β) family (Hara et al., 2009). Resultantly, CARASIL patients have unproteolized cellular proteins, which affect the signal transduction process. Brain MRI shows diffuse white matter changes and multiple lacunar infarctions in the basal ganglia and thalamus (Fukutake, 2011). Histopathologically, arteriosclerosis is seen in the penetrating arteries in the absence of granular osmiophilic or amyloid material (Arima et al., 2003). CARASIL patients are also less likely to have migraines and exhibit psychiatric disorders, such as euphoria and emotional liability (Muqtadar and Testai, 2012).

Prevalence rates for CARASIL are lower than CADASIL, although it is probably more frequent than the few dozen currently reported cases, which to-date have only been described from Japan and China (Fukutake, 2011).

Fabry's Disease

Fabry disease is a congenital metabolic disorder caused by deficient activity of α-galactosidase A, resulting in a progressive accumulation of globotriaosylceramide and related glycosphingolipids within vascular endothelial cells, myocardial cells and neurons (Toyooka, 2011). Prevalence rate of Fabry's is unclear with studies reporting different results. A German study by Rolf et al reported the prevalence of Fabry's in young male stroke patients as 4.9% (Rolfs et al., 2005) and suggested that Fabry's could be a common cause of cryptogenic ischemic stroke. Another multi racial study refutes this finding suggesting reporting Fabry's disease in 0.18% of all strokes and 0.65% of

cryptogenic strokes (Wozniak et al., 2010). Although an X-linked lysosomal storage disorder, female carriers can develop symptoms (Wilcox et al., 2008) that appear comparatively later in life as compared to males, at a median age of 45.7 years (Sims et al., 2009).

Treatment for Fabry's includes bi-weekly recombinant α-gal enzyme replacement therapy at a dose of 1mg/kg body weight, however, continued management of conventional stroke risk factors is important as well (Sharma et al., 2013).

MELAS

MELAS is one of the most clinically prevalent and commonly encountered genetic disorder, 80% of which is accounted for by maternally transmitted mitochondrialtRNA (Leu) A3243G mutations (Goto et al., 1992). Another 10% of patients carry the T3271C mutation. The prevalence of MELAS varies from 7.9/100,000 in England to 236/100,000 in Australia (Manwaring et al., 2007) with an age of onset ranging from 2 to 20 years.

Treatments for MELAS are varied and include the use of vitamin supplements (B complex, E and C) and enzyme co-factors (Q10, idebenone) that enhance mitochondrial metabolism and respiratory chain activity (Sharma et al., 2013).

Other monogenic disorders

A number of other monogenic disorders have been associated with stroke; Marfan syndrome (Della-Morte et al., 2012), Sickle cell disease (Muqtadar and Testai, 2012), homocystinuria (Kelly et al., 2003) and systemic lupus erythematosus (Troedson et al., 2013).

1.4. Genetic Epidemiology of Risk Factors for Stroke

The prevalence rates for various risk factors of stroke vary greatly amongst different countries and mirrors the disparity in rates of stroke mortality. Low-income countries remain as the highest affected regions and current measures of the prevalence of cardiovascular risk factors at the population level fail to predict overall stroke mortality and the greater burden in low-income countries (Johnston et al., 2009). A risk factor can exert its effect on the manifestation of stroke based on several factors including its relative risk and prevalence rate.

Stroke has a large number of modifiable risk factors, which contribute in small increments to the overall risk of disease development. Some of these risk factors such as HTN and diabetes are common and occur is large populations while others such as AF and TIA's are less common. Since the burden of stroke is high, reduction of a risk factor such as HTN can lead to considerable reduction in the incidence of disease at a population level. Importantly, the benefits of modulating risk factors to favorable levels (for example SBP 110 mmHg, total cholesterol 3.8 mmol/L and BMI 21 kg/m^2) (Danaei et al., 2009, Law et al., 2009) produce significantly large and quick results. As a result, the majority of preventive strategies available today such as antiplatelet agents, angiotensin converting enzyme inhibitors, beta-blockers and statins, are aimed at reducing the burden of risk factors. There is an urgent need for effective global strategies and policies to reduce the risk of stroke by reducing the effect of major modifiable risk factors (Ezzati and Riboli, 2012).

Although family history of stroke is an important risk factor several intermediate phenotypes are also involved in the etiology of stroke such as age, gender, HTN, AF, cigarette smoking, T2D and obesity. Such risk factors possess a substantial genetic component themselves and usually cluster such that the risk each imposes varies widely amongst population isolates. There is ambiguity about the exact causative molecular mechanisms for risk factors of stroke and genetic studies have provided some insights into the underpinnings of cellular processes. These studies have also highlighted the genetic contribution to the phenotypic variance of risk factors, which in turn affects their burden in stroke.

Early studies assigned 60% of the PAR for stroke to well known risk factors (Whisnant, 1997). In the only large scale study of its kind, the population based case-control INTERSTROKE study conducted in 22 countries (high income countries, South America, South East Asia (including China), India and Africa) researchers provide compelling evidence towards the role of ten well known modifiable risk factors (**Figure 1.2**), contributing to 90% of the stroke risk in that population (Tu, 2010). On the assumption of a causal relationship, removing all of these risks from the population could prevent 90% of all ischemic strokes. Considering the fact that these risk factors are also under genetic control, the answer to the missing heritability for stroke may lie within the complex interplay of genetic risk factors and phenotypic variance.

	INTERSTROKE (all stroke; 3000 cases, 3000 controls)[3]*	INTERHEART (acute myocardial infarction; 15 152 cases, 14 820 controls)[4]†
Hypertension	34·6% (30·4–39·1)	17·9% (15·7–20·4)
Smoking	18·9% (15·3–23·1)	35·7% (32·5–39·1)
Waist-to-hip ratio (abdominal obesity)	26·5% (18·8–36·0)	20·1% (15·3–26·0)
Diet		
Diet risk score	18·8% (11·2–29·7)	..
Fruits and vegetables daily	..	13·7% (9·9–18·6)
Regular physical activity	28·5% (14·5–48·5)	12·2% (5·5–25·1)
Diabetes	5·0% (2·6–9·5)	9·9% (8·5–11·5)
Alcohol intake	3·8% (0·9–14·4)	6·7% (2·0–20·2)
Psychosocial factors		
All psychosocial factors	..	32·5% (25·1–40·8)
Psychosocial stress	4·6% (2·1–9·6)	..
Depression	5·2% (2·7–9·8)	..
Cardiac causes	6·7% (4·8–9·1)	..
Ratio of apolipoproteins B to A1	24·9% (15·7–37·1)	49·2% (43·8–54·5)

Data are population-attributable risk (99% CI). *Adjusted for all stroke risk factors apart from ratio of apolipoproteins B to A1. †Adjusted for all myocardial infarction risk factors. ‡See original article for definition of risk factor and methods used to calculate population-attributable risk.

Table: Comparison of the population-attributable risk (99% CI) for common risk factors‡ in the INTERSTROKE and INTERHEART studies

Figure 1.2: Comparison of PAR's for common risk factors in the INTERSTROKE and INTERHEART studies (Source: Tu, 2010).

1.4.1 Age

Stroke is of increasing socio-economic importance in aging populations. Historically it has been an age related disorder with an exponential increase in incidence with advancing age (Correia et al., 2004), however recent studies have demonstrated an increase in the incidence of stroke in young adults (aged 20-55 years) as well (Kissela et al., 2012). This trend of younger stroke has huge implications on public health care expenses due to a substantial loss of productive life years and a longer need for post stroke care. Age is also strongly associated with TIA, HTN and smoking, emphasizing the fact that preventive strategies may have greater benefit in younger stroke patients because of the higher relative risk for stroke at younger ages for these risk factors (Whisnant, 1997), whereas the same strategies in older patients may not apply. Preventive strategies in older patients have to reach many more people to have a comparable effect because of the lower relative risk with increasing age (Whisnant, 1997).

Stroke is easier to diagnose in older patients since large extracranial and intracranial atheroma, small-vessel disease and AF are frequent in older individuals and so are traditional risk factors such as diabetes and HTN (Ferro et al., 2010). Age is an important determinant of stroke outcome in the elderly (Konig et al., 2008) and observational studies and randomized trials that have shown poorer outcomes and higher risk of bleeding in this population due to complications from thrombolysis (Mishra et al., 2010, Sylaja et al., 2006). In a review of population based epidemiological studies Feigin et al reported that the age-specific incidence of stroke increases progressively with each decade of life (**Figure 1.3**) (Feigin et al., 2003). Individuals between 75–84 years of age had tens of folds greater incidence rates (12.0–20.0 per 1000 person-years) than those aged less than 45 years (0.1 to 0.3 per 1000 person-years) (Feigin et al., 2003).

Environmental insult also increases with age and intern affects the genetic constitution of an individual. Epigenetic modifications and differential gene expression occurring over an individual's lifespan reflect the natural process of aging and may contribute to the development of age-related phenotypes and diseases such as T2D, autoimmune and vascular disease (Johansson et al., 2013). A recent study by Traylor et all examined the

role of age at stroke onset as a source of heterogeneity and demonstrated that genetic heterogeneity exists within ischemic stroke sub-phenotypes by age at onset and urge researchers to examine their existing datasets for younger onset cases (Traylor et al., 2013).

The study participants included in our analysis varied in age and included both young and old adults. Therefore age was considered a confounding factor and used as a covariate to adjust the regression analysis in Chapters 3 and 4.
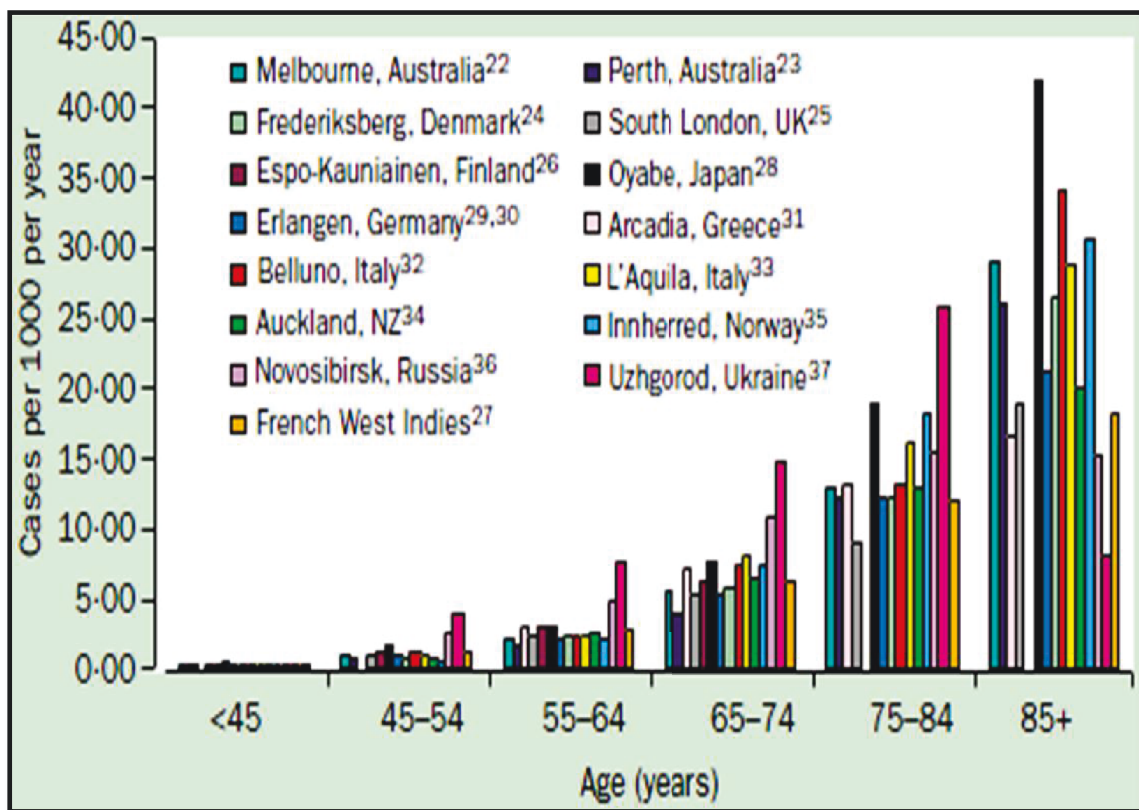
Figure 1.3: Annual incidence by age per 1000 population of all types of stroke combined in selected studies (Source: Feigin et al., 2003).

1.4.2 Gender

Gender based differences in vascular disease are increasingly apparent (Banerjee et al., 2009). While men have a greater incidence of stroke up to 75 years of age (Thom et al., 2006), women shoulder a higher lifetime risk due to their greater life expectancy (Seshadri et al., 2006). Not only do women stroke survivors have worse outcomes post stroke but they also have a completely different risk factor profile from men. Women are older than men at stroke onset, which may help explain the higher presence of age, related risk factors such as cholesterol, HTN and AF (Bushnell, 2008, Haast et al., 2012). Changes in endogenous hormone levels, due to pregnancy or menopause, throughout a woman's life cycle also serve as risk factors (Bushnell, 2008).

The OXVASC was the first large-scale study to investigate gender-specific transmission of stroke. The study concluded that the heritability of ischemic stroke was greater in women than in men, with an excess of affected mothers and affected sisters in female probands independent of traditional vascular risk factors and intermediate phenotypes (Touze and Rothwell, 2007).

The results of the study were validated by performing a large literature based meta-analysis, which consolidated data from the original OXVASC study and 17 other studies (Touze and Rothwell, 2008). Although non-genetic risk factors were thought likely to explain the results of this study to a large extent, the presence of genetic factors was assessed for a possible explanation for the maternal excess of stroke. The authors speculated that mother-to-daughter transmission was probably mitochondrial in nature, which has a higher penetrance in women than in men. Alternately differential gene expression or epigenetics, which are changes in gene expression that do not entail a change in DNA sequence, provide an alternative explanation (Jiang et al., 2004). Further, the authors delve into the realm of genomic imprinting, which is a subtype of epigenetic regulation in which the activity of a gene is reversibly modified depending on the sex of the parent that transmits it (Jiang et al., 2004). Studies have shown that the epigenomic state of a gene can be established through behavioral programming and that the epigenotype, in particular imprinting genes, could be more susceptible than genotype to environmental factors (including maternal behavior in animals) (Weaver et al., 2004, Touze and Rothwell, 2007).

The gender differences in stroke are complex and may reflect the genetic differences that control hormonal state or longer life spans for women.

Since gender is an important marker for clinical and genetic factors, it was considered a confounding factor and used as a covariate to adjust the regression analysis in Chapters 3 and 4.

1.4.3 Ethnicity

Studies have shown race-ethnic disparities in the distribution and prevalence of stroke, its risk factors and mortality among various populations (Cruz-Flores et al., 2011, Sen et al., 2013, O'Donnell et al., 2010b, Khan et al., 2013). The WHO supports these findings and has identified significant differences in the prevalence of stroke risk factors such as HTN, cholesterol and smoking amongst populations from the developed and developing regions of the world.

Ethnic disparities in clinical phenotypes and risk factors also exist. Using a Caucasian reference group, Sen et al (Sen et al., 2013) examined race-ethnic disparities in minority groups (Asian-Indian, African-American and Hispanic) residing in the US. The study showed that racial differences in the impact of age and gender contributed to the risk of stroke i.e. Asian-Indians and the Hispanics were largely younger males as compared with the white stroke patients who were older females. Risk factors such as diabetes (55%) and elevated plasma Hcys levels (12.1 µmol/L, p=0.002) were most prevalent in Asian-Indians, who also had a higher number of strokes related to small vessel occlusive disease (25%).

Similar results were observed by the INTERSTROKE study, which by far remains the largest epidemiological study covering 22 countries 3000 stroke patients (O'Donnell et al., 2010). Whether phenotypic differences in stroke can be attributed to genetic differences between races remains to be validated. Although studies have suggested that genetic risk associations are broadly similar for different ethnicities (Ariyaratnam et al., 2007), this does not rule out differential expression of intermediate biomarkers between ethnic groups (as shown for Hcys) (Bentley et al., 2010). Also most genetic association studies have examined common genetic variants and it is likely that the risk may lie with rare or non-SNP variations in the genome.

Genetic risk variants of stroke are well studied in some ethnic groups such as Caucasians and Chinese, whereas they remain understudied or completely unknown in South Asians from India, Pakistan, Sri Lanka and Bangladesh. Reports on monogenic stroke in South Asians are limited. Three Indian studies on CADASIL have been reported (Yadav et al., 2013, Panagariya et al., 2004, Gurumukhani et al., 2004), of which only 2 obtained a genetically confirmed diagnosis. Only one study on pediatric Fabry's disease

has ever been reported with pedigree analysis for the *α-galactosidaseA (GLA)* gene mutation confirmed based on clinical suspicion only (Phadke et al., 2009). Studies on other forms of monogenic stroke such as CARASIL, MELAS, sickle cell anemia and Marfan syndrome have not been reported from India. It is surprising that despite the common frequency of sickle cell anemia in India, no studies examining the incidence of stroke in sickle cell anemia have been conducted.

South Asian studies on sporadic stroke are also limited consisting mainly of a collection of small underpowered candidate gene studies. A selection criterion for candidate genes either depends on the researcher interest and includes cellular processes of choice (for example signal transduction, metabolism) or on GWAS-derived risk variants associated with other vascular disorders such as MI and CAD (**Figure 1.4**). Associations were confirmed for some risk variants that were previously identified for European populations, *PDE4D, MTHFR, ACE, APOE, LPL* and *eNOS* (Kaul and Munshi, 2012). Although these studies do not carry the same statistical weight as the larger European studies, they provide evidence for a genetic burden of stroke in South Asians and estimate the possible effect size exerted by these risk variants.

In light of the fact that no large genetic association studies have been conducted on South Asians, we conducted a literature based meta-analysis of small candidate gene studies (Chapter 2) in order to increase power of the identified associations.

Pathway	Gene	Polymorphism
Coagulation	Prothrombin	G20210A
	Factor V leiden	C1619T
Fibrinolytic system	tPA	-7351C/T, I/D
	PAI-1	4G/5G
Renin angiotensin	ACE	I/D
aldosterone system	CYP11B2	−344C/T
	CYP4F2	1347 G/A
	ESR-1	Pvull and Xba morphism
Hom steine metabolism	MTHFR	C677T
eNOS metabolism	eNOS	27bp VNTR
Lipid metabolism	LPL	HindIII polymorphism
Extracellular matrix degradation	MMP-3	5A/6A
Inflammation pathway	IL-10	-1082G/A
	TNF-α	+488 G/A
cAMP degradation pathway	PDE4D	SNP41,83,87,56 and 37

Figure 1.4: Candidate gene polymorphisms and associated phenotypes studied in the Indian population (Source: Kaul and Munshi, 2012)

1.4.4 Blood Pressure

Hypertension is the biggest treatable risk factor for ischemic stroke (O'Donnell et al., 2010) and association of elevated blood pressure is a strong predictor of stroke. Small changes in blood pressure have serious effects on prognosis of stroke and treatment of HTN can reduce this risk by over 40% (WHO, 2002). The number of hypertensives across the world has risen from 600 million individuals in 1980 to over 1 billion in 2008, with higher prevalence in low and middle-income countries (WHO, 2013). In 2010, a comparative risk assessment study on the burden of disease attributable to 67 risk factors showed HTN to be the leading single risk factor accounting for 9.4 million (95% UI 8.6-10.1) deaths and 7% (95% UI 6.2-7.7) of the global disability adjusted life years (Lim et al., 2012). The study also showed that high blood pressure is one of the top five risk factors in individuals between 15-70 years of age and accounts for 15%-20% of all health loss in individuals between 50-70 years of age (Lim et al., 2012).

Another study by Johnston et al described the national level prevalence of stroke risk factors and found that raised mean systolic blood pressure predicted stroke mortality (albeit poorly) (Johnston et al., 2009).

The Framingham Heart Study has been at the head of several important contributions to the current understanding of the epidemiology of blood pressure and stroke (Romero and Wolf, 2013). The FHS investigators demonstrated the importance of several measures of blood pressure such as SBP, DBP and MAP in the development of stroke. A systematic review of case-control and cohort studies reporting family history of HTN as risk factors for stroke showed that the frequency of HTN was greater in stroke patients than controls. Authors suggested that the apparent heritability of stroke could be partly accounted for by the heritability of HTN (Flossmann et al., 2005).

Familial studies have long provided evidence of heritability (31%-68%) of blood pressure (BP) (Ehret et al., 2010). Since 2009, several large studies have contributed to the understanding of the genetics of various measures of blood pressure such as SBP, DBP, MAP and PP (Levy et al., 2009, Newton-Cheh et al., 2009, International Consortium for Blood Pressure Genome-Wide Association et al., 2011, Kato et al., 2011, Wain et al., 2011, Johnson et al., 2011), although the effect size accounting for phenotypic variance was small**.**

The genetics of long-term variability in blood pressure or episodic HTN remain unknown though its role in stroke is not new, with early studies (Ekbom et al., 1992) demonstrating the effect of antihypertensive drugs on decreasing blood pressure variability for the lowered risk of stroke.

In Chapter 3, genes influencing BP variability were identified in 3802 individuals from the ASCOT IR-UK cohort. GWAS were tested for association with ischemic stroke using a meta-analysis of an independent ischemic stroke population comprising 8624 cases and 12722 controls from 7 studies.

1.4.5 Carotid Disease

Carotid IMT measured by high-resolution B-mode ultrasonography, is a well-studied surrogate marker for subclinical atherosclerosis and a strong predictor for stroke. Ultrasonography measures of the carotid artery can be a useful intermediate phenotype for the identification of early markers of risk (Manolio et al., 2004, Bis et al., 2011). Presence of carotid stenosis or plaque is indicative of a diffuse vascular condition and can generate thrombo emboli in a specific arterial territory-causing stroke. Besides the arterial narrowing and embolic characteristics, stenosis can also produce downstream hemodynamic effects that impair cerebrovascular reserve and is strongly associated with the risk of ischemic stroke (Gupta et al., 2012). A 30% increase in the risk of recurrent stroke for the first 2 years following the first-time stroke is reported for symptomatic carotid stenosis (**Figure 1.5**)(Markus, 2003).

Variability of the carotid artery structure is suggested to be under genetic control and early family studies provided evidence supporting the presence of a strong independent genetic component to the burden of carotid IMT. Although there was evidence of heritability, initial candidate gene based studies were not successful in finding strong or consistent associations with common genetic risk variants (Manolio et al., 2004). The earliest evidence came from a candidate gene study, which found a risk association between *IL-6 174G/C* polymorphism with common carotid artery IMT although the effect size was small. The study showed that the CC genotype of *IL-6 174G/C* polymorphism was associated with a 4.8% increase in maximal carotid IMT and accounted for 0.6% of the observed variation in the trait, which is equivalent to 2.5% of the heritable component (Mayosi et al., 2005).

The CHARGE consortium GWA study on carotid disease phenotypes is the largest study of its kind with > 40,000 individuals of European ancestry (Bis et al., 2011). The study identified genetic risk variants associated with increased risk of cIMT and plaque although the effect sizes were small (0.9% IMT variance and 18%-22% increase in odds of plaque) (Bis et al., 2011). Despite the small effect sizes, the study robustly established the genetic burden carried by different stages of carotid disease.

In Chapter 5 we conducted an *a priori* GWA study on ~14,000 patients with advanced carotid disease (>50% stenosis) in an ischemic stroke population.
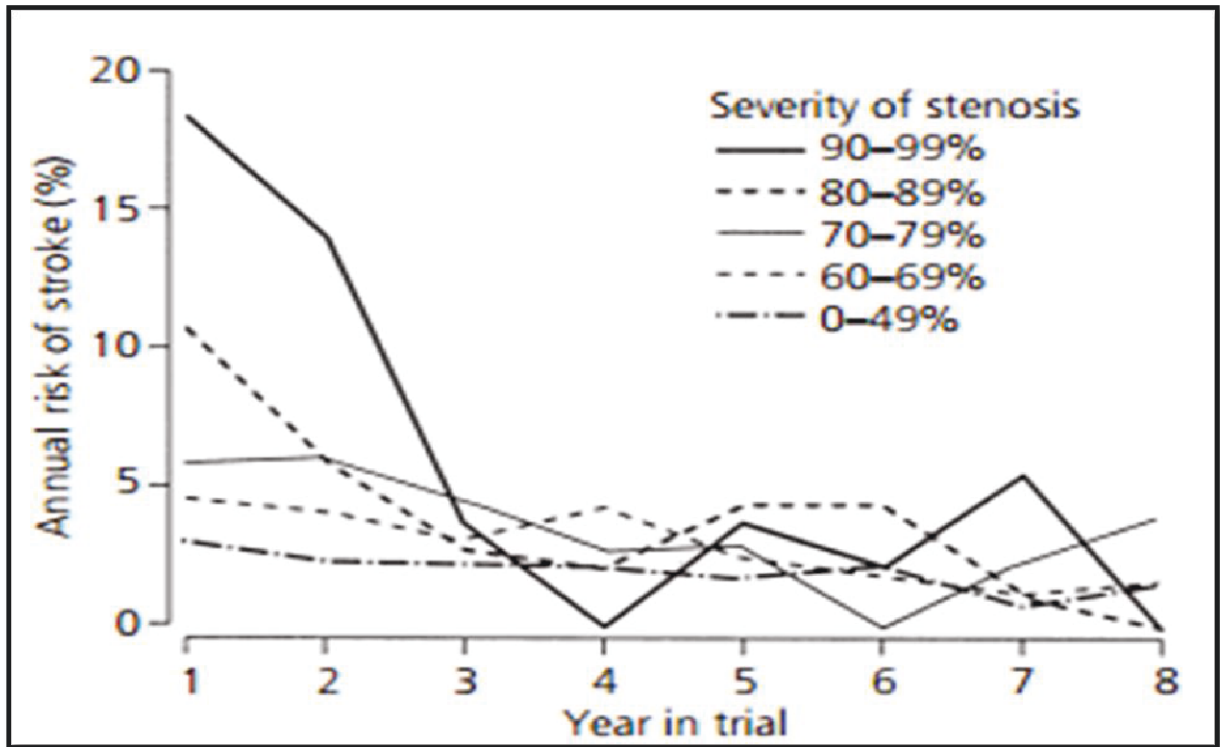
Figure 1.5: Risk of recurrent stroke with increasing carotid stenosis.

(Source: Markus, 2003, Originally published in European Carotid Surgery Trialist's Collaborative Group 1998)

1.4.6 Diabetes

Type 2 diabetes is a major public health concern and has risen to become the 5th leading cause of death across the globe (Shaw et al., 2010). The disease is also a major cause of morbidity and risk factor for stroke. A population study by Danaei et al studied the population distribution of fasting plasma glucose to measure exposure to higher-than-optimum blood glucose by collating individual patient level data, systematic reviews from 52 different countries. The study showed that in addition to 959,000 deaths attributed to diabetes, 709,000 deaths from stroke were attributable to high glucose levels (Danaei et al., 2006).

Family studies have attributed a heritable component to T2D (h2=49%) (Risch, 1990), with twin studies showing high concordance rate in monozygotic twins (96%) supporting an underlying genetic burden (Medici et al., 1999, Lo et al., 1991). Several large GWA studies have identified significant associations with genetic risk factors and T2D. A study by Voight et al comprising of 8,130 T2D cases and 38,987 controls of European descent identified 12 genetic risk variants influencing beta-cell function and insulin action (Voight et al., 2010). Another study by Zeggini et al identified more genetic variants influencing pancreatic beta cells (Zeggini et al., 2007) thereby providing insight molecular mechanisms of T2D.

1.4.7 Lipids

Stroke has historically been an age related disorder affecting elderly patients; however in recent times this boundary has been blurred. Due to a surge in risk factors such as cholesterol and HTN, stroke is now manifesting itself in young individuals between 35 to 64 years. Although elevated cholesterol has typically been a weak risk factor for ischemic stroke of all types, it is a significant risk factor for carotid disease (Musialek et al., 2013), which in turn is a major risk factor for stroke.

A huge evidence base implicating elevated plasma lipids as risk factor for stroke exists. A 2012 study by the American Heart Association found that obesity (body mass index >30 kg/m^2) was associated with increased mortality and morbidity in the US population which was in turn associated with increased incidence of risk factor such as diabetes mellitus, vascular disease (coronary heart disease, stroke, and heart failure), and other health conditions such as asthma, cancer and degenerative joint disease (Roger et al., 2012).

Large GWA studies have reliably identified > 100 genetic risk variants associated with lipids such as HDL, LDL, triglycerides and total cholesterol (Teslovich et al., 2010) supporting a genetic liability for their phenotypic variance.

1.4.8 Homocysteine

Elevated plasma Hcys levels have long been associated with stroke with contradicting reports on its causative or consequential roles. Notwithstanding this paradigm, stroke patients have significantly higher Hcys levels as compared to healthy individuals (Biswas et al., 2009) and are at a 19% increased risk of stroke for every 3μmol/L increase in plasma Hcys levels (Collaboration, 2002). Although the mechanistic understanding of the role of Hcys in stroke is still poorly understood, there is sufficient evidence available to tie it in as a major risk factor. Given the fact that Hcys levels can be effectively controlled through diet supplemented with folic acid and vitamin B6/B12 (Hankey et al., 2013), it is important for clinical consideration.

Several phenotypic and genetic studies have been conducted to examine the relationship between gene polymorphisms, Hcys and stroke, and unanimously support the association of Hcys with the increased risk of stroke. Although several genes play a role in Hcys metabolism, *MTHFR* remains the most widely studied. A study by Hsu et al investigated 86 SNPs in 9 candidate genes (*BHMT1, BHMT2, CBS, CTH, MTHFR, MTR, MTRR, TCN1, and TCN2*) coding for enzymes and cofactors involved in Hcys metabolism (Hsu et al., 2011). The study found that risk variants influenced post stroke plasma Hcys levels by altering the metabolic pathway, and also increased the risk of recurrent stroke in response to cofactor therapy (Hsu et al., 2011).

The largest study to so far examine the modulation of effect exerted by gene polymorphisms according to changes in dietary folate suggested that, the genetic effect of the *MTHFR 677CT* variant on plasma Hcys levels was larger in low folate regions such as Asia (Mean difference of 3.12 μmol/L, 95% CI 2.23-4.01 between TT versus CC genotypes in healthy individuals) than in areas with high folate fortification such as America, Australia, and New Zealand (Mean difference of 0.13 μmol/L, 95% CI -0.85-1.11)(Holmes et al., 2011). The study also reported increased odds of stroke for Asia (OR 1.68, 95% CI 1.44-1.97) as compared to America, Australia, and New Zealand (OR 1.03, 95% CI 0.84-1.25) for individuals carrying TT vs. CC genotypes. The findings of this study strongly argued in support of a gene-environment interplay, which is understandable since folic acid is a major player in Hcys metabolism.

Indians have considerably higher Hcys levels as compared to other ethnicities (Bentley et al., 2010, Yadav et al., 2013, Bhargava et al., 2012). In a study on an Indian population, Biswas et al found plasma Hcys levels to be significantly higher in stroke patients as compared to healthy individuals (12 μmol/L, range 5.3-39.1 μmol/L vs. 11.2 μmol/L, range 6.2-14.2 μmol/L; P =0.001). Similar results were found by a later study although the Hcys levels were considerably higher in both cases and controls (12.42μmol/L, SD ± 0.85 vs. 19.30 μmol/L, SD ± 2.41) (Bhargava et al., 2012). The study mirrored the findings of Holmes et al and reported lowering of patient Hcys levels in response to the increased folic acid dosage as well as a risk association between *MTHFR C677T* polymorphism and stroke (Biswas et al., 2009).

In Chapter 2, we examine the causal relationship between MTHFR C677T polymorphism and stroke via Hcys in a South Asian population using a Mendelian randomization strategy.

1.5. Strategies to Study Genetics of Stroke

With the emergence of large stroke consortia and developments in genotyping technology, statistical methods and computational power, researchers have finally begun to address the genetics of ischemic stroke effectively. Advances in our knowledge of the molecular underpinnings of stroke will enable scientists and clinicians to better understand the mechanistic workings of stroke and design effective treatments for it.

Evidence for stroke genetics can come from two different platforms; study of individuals and population based studies. Study of individuals can help identify genetic variants that causally affect stroke and provides concrete evidence for the genetic risk of stroke. Individual studies usually identify rare genetic variants with large effect sizes and high penetrance. Although such studies are of immense value, they rarely contribute to the prevention of stroke at a population level since it involves a large number of individuals at a small risk of stroke which gives rise to more cases of disease than a small number who are at high risk (Rose, 1985). Individual based studies have led to the identification of several monogenic forms of stroke such as CADASIL (Joutel et al., 1996) and enabled clinicians to use this information in their every day clinical practice.

Population based genetic association studies have found great popularity with genetic epidemiologists since the study samples are more representative of the general population and easier to recruit as compared to stroke families. Results from a large population study are useful in calculating PAR of a genetic variant, which can be extrapolated to the general population. Population studies also have greater power in detecting common genetic variants that affect >5% of the population.

The following sections discuss the current state-of-knowledge of population based genetic strategies available to researchers.

1.5.1 Candidate Gene Based Association Studies

In the pre-GWAS era, candidate gene based studies have enjoyed wide spread popularity in the study of stroke genetics. These studies are based on a prior hypothesis, primarily driven by the choice of a candidate gene which is based on the investigators research interest in a particular biological pathway such as coagulation, lipid metabolism, inflammation and blood pressure regulation (Matarin et al., 2008, Williams et al., 2013, Wang et al., 2009, Olsson et al., 2012), or candidate genes derived from related vascular conditions such as MI (Gschwendtner et al., 2009) or CAD (Arregui et al., 2012). This is not surprising as the pathophysiology of stroke and coronary disease are similar (Jashari et al., 2012). However replication of such candidate genes in other phenotypes have not always been successful (Cheng et al., 2012), with some candidates appearing to be organ-specific rather than pathophysiology-specific (Bentley et al., 2010, Helgadottir et al., 2012). Candidate genes found to be associated with stroke in one ethnic population are also routinely replicated in other ethnicities (Cheong et al., 2011, Kostulas et al., 2007).

Recently, findings from candidate gene association studies in stroke and other vascular phenotypes such as CAD and AF were replicated using statistically robust GWAS models. Using >3500 stroke cases and 5700 controls from the WTCCC 2ischemic stroke GWAS (International Stroke Genetics et al., 2012), Bevan et al tested association for 50 previously reported candidate genes (Bevan et al., 2012). Of the 32 stroke associated genes tested, 4 genes *ALOX5AP (*CE), *APOA (LPA)* (SVD), *Fibrinogen (*all ischemic stroke), and *Paroxonase-1* (SVD)) survived Bonferroni correction but failed when the Nyholt correction was applied (Bevan et al., 2012). The study also tested 18 genes associated with cardiovascular phenotypes and validated the association for3 genes at the modified Nyholt threshold: *PHACTR1*in LVD (P=2.63×10^{-6}), *PITX2*in CE stroke (*P*=4.78×10^{-8}), and *ZFHX3*in CE stroke (*P*=5.50×10^{-7}). Given the failure to replicate most stroke associated genes, the study concluded that the risk association is likely to be sub-type specific and success in identifying risk variants would continue to evade researchers unless the study populations are larger and extensively sub-typed.

Candidate gene studies have also been applied to test the progression of stroke through its intermediate phenotypes. Adib-Samii et al examined the 17q25 locus, which was

previously found to be associated with white matter hyperintensities in stroke-free individuals, and replicated the association with white matter hyperintensity volume in ischemic stroke patients to determine whether the 17q25 locus promotes small vessel arteriopathy. The study furnished evidence in support of an association between 17q25 and white matter hyperintensities (Adib-Samii et al., 2013).

Unique step back approaches have also been implemented to test association of candidate genes with stroke. A study by Krug et al performed gene expression profiling in peripheral blood mononuclear cells of 20 stroke cases and 20 controls and examined the differentially expressed genes between the two groups. Sixteen differentially expressed genes were then mapped to GWAS-derived regions associated with various vascular disorders. Using this approach the group was able to identify a risk association between stroke and the *TTC7B* gene locus (Krug et al., 2012).

Although candidate genes have largely produced unsatisfactory results and have not been successfully replicated in larger stroke association studies, these approaches have been helpful in indicating the effect size that a risk allele may exert. Candidate gene studies have also implicated disparity in the genetic burden of stroke for different stroke subtypes (Kostulas et al., 2009, Anderson et al., 2013) before this was discovered in large-scale GWA studies (Holliday et al., 2012). Genes involved in lipid metabolism and enzymatic activities are the most widely studied candidates for association with stroke (Stankovic et al., 2012) and provide an insight into the possible cellular processes that may underline stroke pathology.

1.5.2 Literature Based Meta-Analysis

In the current era of GWAS studies and whole genome sequencing, the literature-based meta-analysis has found its own unique utility. Where large GWAS studies have failed to identify risk associations with gene variants such as *MTHFR* and *ACE* literature based meta-analysis of comparable (or greater) power and sample size have been able to identify these genetic risk variants to be associated with stroke (Bentley et al., 2010, Bevan et al., 2012, Paternoster et al., 2009).

Although the stand-alone validity of individual candidate gene based studies and small meta-analysis (McColgan and Sharma, 2008) remains inconclusive, large-scale literature based meta-analysis may reveal a true association or confirm previous findings. The odds associated with candidate gene SNPs are varied due to small sample sizes, inter-study heterogeneity and publication bias; however meta-analyses of these studies pool the effect sizes in statistically sound models to provide robust results. Although there is a paucity of genetic studies for some ethnic groups, this methodology has been applied to candidate studies across different ethnic populations and the odds have broadly held true for all.

Results from most literature-based meta-analysis have suggested that common stroke has a genetic component with no single gene having a major effect. These studies have shown that common variants in several genes exert individual modest effects and contribute to the overall risk of stroke. They have also consistently demonstrated that genetic risk associations for ischemic stroke are similar across different ethnicities. Such findings have found great applicability in the design of large genetic studies and for predictive genetic testing for stroke.

A recent meta-analysis comprising of 41 published & unpublished studies (9,027 ischemic cases and 61,730 controls) established a small risk association between IS and different *Apolipoprotein E* genotypes (OR 1.05-1.12) (Khan et al., 2013). In an effort to establish causality for the observed effects between genotype-IS, the study further used individual patient level data from 16 studies (60,883 individuals). Mendelian randomization was used to translate the observed association to genotype-plasma lipids and a 33% increase in odds per 1 mmol/l increase in LDL plasma levels was reported (Khan et al., 2013).

Another comprehensive meta-analysis of 187 candidate genes in 37,481 European ischemic stroke cases and 95,322 controls, identified 5 genes to be associated with the risk of stroke; *Factor V ArgGln506* (OR 1.31), *ACE/ID* (OR 1.15), *MTHFR C677T* (OR 1.26), *Prothrombin G20210A* (OR 1.60), *PAI-15G* (OR 1.11), *Glycoprotein IIIa Leu33Pro* (OR 1.24) (Bentley et al., 2010). Although the effect size per gene was modest (OR 1.11-1.60), they accounted for accumulative population attributable risk of ~30%. This result is of great significance given the high frequency of stroke in the general population and translates to a large clinically observed effect.

A large meta-analysis conducted on 32,431 non-European individuals (Chinese, Japanese, and Korean), identified eight candidate genes, three of which were associated with ischemic stroke: *ACE (I/D)* polymorphism, OR 1.90 (95% CI 1.23–2.93) in the Chinese and OR 1.74 (95% CI0.88–3.42) in the Japanese; *MTHFR C677T*, OR 1.18 (95% CI 0.90–1.56) in Chinese and 1.34 (95% CI 0.87–2.06) in Koreans; and *APOE,* OR 2.18 (95% CI 1.52–3.13) in Chinese and 1.51 (95% CI 0.93–2.45) in Japanese(Ariyaratnam et al., 2007). The observed effect sizes were modest (OR 1.18-1.90) and largely similar amongst different ethnic groups.

A meta-analysis by Casas et al including 120 case-control studies and 32 genes, identified statistically significant risk associations with ischemic stroke for *Factor V Leiden Arg506Gln* (OR, 1.33; 95% CI, 1.12-1.58), *MTHFR C677T* (OR, 1.24; 95% CI, 1.08-1.42), *Prothrombin G20210A* (OR, 1.44; 95% CI, 1.11-1.86), and *ACE I/D* (OR, 1.21; 95% CI, 1.08-1.35)(Casas et al., 2004). Using a similar methodology, the group went on to establish causality between *MTHFR C677T* polymorphism, plasma Hcys levels and risk of stroke, thereby demonstrating the validity of literature based meta-analysis (Casas et al., 2005).

Literature based meta-analysis have also been applied to studying ischemic stroke subtypes (Rao et al., 2009), intracranial aneurysms (McColgan et al., 2010) and venous thromboembolisms (Gohil et al., 2009, Marjot et al., 2011). Although evidence for sub-type specific genetic risk variants is thin (Rao et al., 2009), robust associations have been identified for VTE. A large meta-analyses comprising of 126,525 cases and 184,068 controls from 173 case-control studies and including 21 genes was conducted (Gohil et al., 2009). Risk associations with VTE in Caucasian populations were identified for

Factor V G1691A (OR 9.45; 95% CI 6.72–13.30, p<0.0001), *Factor V A4070G* (OR 1.24; 95% CI 1.02–1.52, p=0.03), *Prothrombin G20210A*, (OR 3.17; 95% CI 2.19–3.46, p<0.00001), P*rothrombin G11991A*, (OR 1.17; 95% CI 1.07–1.27, p=0.0007), *PAI-1 4G/5G*, (OR 1.62; 95% CI 1.22–2.16, p=0.0008), A*lpha-fibrinogen Thr312Ala* (OR 1.37; 95% CI 1.14–1.64, p=0.0008). *MTHFR C677T* was associated with risk in Chinese/Thai populations (OR 1.57; 95% CI 1.23–2.00, p=0.0003), and *ACE I/D* in African American populations (OR 1.5; 95% CI 1.03–2.18, p=0.03). *Factor XIII Val34Leu* (OR 0.80; 95% CI 0.68–0.94, p=0.007) and *β-fibrinogen 455 G/A* (OR 0.84; 95% CI 0.72–0.97, p=0.02) both showed significantly protective effects (Gohil et al., 2009). Causal relationships with *MTHFR C677T* polymorphisms have also been suggested (Marjot et al., 2011).

1.5.3 Genome Wide Association Studies

Population based GWA studies have been made possible by the advent of the human genome project (Consortium, 2004) and the HapMap consortium (Consortium, 2005). With completion of the Human Genome Project in 2003, scientists identified regions of variation between individuals, the most common form of which is the single nucleotide polymorphism or SNP. The human genome is believed to consist of over 10 million SNPs and, with the efforts of the International HapMap project, 3.1 million SNPs have already been characterized (Frazer et al., 2007). Information provided by the international HapMap consortium has enabled the development of commercially available genotyping microarrays and heralded the era of the GWA study. In recent times the 1000Genomes project (http://www.1000genomes.org) has provided 4X deep sequencing data and added immensely to the knowledge base.

As technology used to unravel the genetic basis of disease has advanced, our ability to rapidly search for genetic risk variants on a throughput scale has dramatically improved. Statistically underpowered candidate gene studies have predominantly been replaced by whole-genome screening, which has been successfully conducted in a variety of disorders (Lettre and Rioux, 2008, Easton and Eeles, 2008). The emergence of genome wide approaches, which test variants without any priori hypothesis have presented investigators with an alternative and more powerful method to test the productivity of the candidate gene based, approached (Bevan et al., 2012).

GWAS provides a powerful statistical model to study millions of genetic variants in a single experiment. The methodology is also free from recall and temporal bias. Population based case control studies possess two distinct advantages over other study designs. The samples are easier and cheaper to recruit and large sample sets can be collected in relatively less time. Under the assumption of low population substructure, population based studies are also believed to be more powerful than family studies (Morton and Collins, 1998). One of the earliest applications of the GWAS model was to study age-related macular degeneration, which revealed associations with complement factor H gene (Edwards et al., 2005). In later years, this methodology has been successfully applied to several complex polygenic disorders such as bipolar disorder, CAD, carotid disease, Crohn's disease, HTN, rheumatoid arthritis, blood pressure, smoking and T2D (NHGRI catalogue, http://www.genome.gov/gwastudies).

GWA studies on stroke were few till 2003 when Gretarsdottir et al identified *PDE4D* to be significantly associated with risk of ischemic stroke in an Icelandic population (DeCODE) (Gretarsdottir et al., 2003). However several attempts to replicate these findings failed (Rosand et al., 2006, Bevan et al., 2005, Newcombe et al., 2009), while some studies reported conflicting results (Bevan et al., 2008, Matsushita et al., 2009, Bentley et al., 2010). These discrepancies were attributed to possible problems in study design, i.e. pooling of CE and LVD strokes in order to identify the risk association with PDE4D, which we now know are sub-type specific (Holliday et al., 2012).Lack of independent replication and limited experimental validation of results using B-cell lines was also suggested to be a major limitation of the original study (Worrall and Mychaleckyj, 2006). Several other GWA studies followed, but no single locus was identified at a genome wide level of significance (p value≤5x10^{-8}).

The WTCCC2 and the ISGC performed a GWAS involving 3,548 cases of ischemic stroke with replication of potential signals in 5,859 additional cases (International Stroke Genetics et al., 2012). The study demonstrated, as others had done previously (Gudbjartsson et al., 2009, Gudbjartsson et al., 2007), associations for CE stroke near *PITX2* and *ZFHX3*, which are known risk loci for AF (Ellinor et al., 2012). The study also confirmed the association for LVD and 9p21 locus. A novel finding was an association for large vessel stroke within *HDAC9* on chromosome 7p21.1 (OR 1.42). In a recent GWAS, the evidence for a stroke sub-type specific genetic influence became more compelling with the association of the 6p21.1 locus with large artery stroke subtype (Holliday et al., 2012).

Despite a number of genome-wide association studies reporting discoveries of novel genetic risk variants for stroke, these were rarely replicated implying that either these discoveries were fallacious or the effect is small such that much larger studies are required (Lanktree et al., 2010).The largest and most recent GWA study published so far, the METASTROKE meta-analysis, (~12,000 cases and ~60,000 controls) validated previous findings of genes *PITX2*, *ZFHX3*, and *HDAC9* suggesting that these are true associations (Traylor et al., 2012). All loci exhibited heterogeneous effect across subtypes, supporting distinct genetic architectures for each subtype.

Several other GWAS have been conducted in stroke mostly in those of European ancestry, with very little comparative data available in other ethnic populations. A few studies have been conducted in populations of South East Asian ancestry (Kubo et al., 2007, Yamada et al., 2009, Hata et al., 2007, Ding et al., 2010). The effect size is broadly similar across all GWAS's, ranging from 1.00 to 1.85, confirming that the effect sizes are small but the population attributable risks could be large given the common nature of this condition. Many studies however have failed to replicate their findings (Debette et al., 2010, International Stroke Genetics and Wellcome Trust Case-Control, 2010, Ikram et al., 2009).

Given that some ischemic stroke has a maternal heritability, a GWAS of common mitochondrial sequence variants failed to find a genome significance threshold, although this study was underpowered for GWAS (Anderson et al., 2011). GWA studies on stroke twins found no significant hits but were able to demonstrate significant correlation of age at stroke within pairs of affected siblings (r=0.83, 95% CI 0.78–0.86, p=2.2×10^{-16}) and high concordance of stroke subtypes among affected pairs (33.8%, kappa=0.13, p=5.06×10^{-4}) which did not differ by age at stroke in the proband (Meschia et al., 2011).

Some investigators have undertaken GWAS on surrogate markers of stroke such as white matter hypertensities, which are endophenotypes of small vessel ischemic disease without any reliable associations (DeStefano et al., 2006, Turner et al., 2005, Paternoster et al., 2009). Other intermediate phenotypes such as carotid intima-media thickness have also been explored with promising associations reported (Bis et al., 2011).

Reports of a new wave of GWA studies is underway, including the WTCCC2 and NINDS Stroke Genetics Network (Meschia, 2011, Meschia et al., 2013) which will utilize the CCS classification system (Ay et al., 2007).The studies will focus entirely on sub-typing large number of ischemic stroke cases. A total of 24 genetic research centers across Europe and America will participate in this global consortium amassing over 14,549 stroke cases (Meschia et al., 2013).

GWA studies on stroke, which reported GWAS significant associations (p≤5×10^{-8}) are summarized in **Table 1.1.**

Table 1.1: Genome wide association studies on ischemic stroke

Gene	SNP	Chr	Phenotype	Ethnicity	RA	RAF	OR (95% CI)	p	Study	Other studies
HDAC9	rs2107595	7p21.1	LVD	European	A	0.16	1.39 (1.27-1.53)	2.03×10^{-16}	(Traylor et al., 2012)	(International Stroke Genetics et al., 2012)
PITX2	rs6843082	4q25	CE	European	G	0.21	1.36 (1.27-1.47)	2.8×10^{-16}	(Traylor et al., 2012)	(Gudbjartsson et al., 2007, Gretarsdottir et al., 2008, International Stroke Genetics et al., 2012)
CDKN2A /B	rs2383207	9p21.3	LVD	European	G	0.52	1.15 (1.08-1.23)	3.32×10^{-5}	(Traylor et al., 2012)	(Matarin et al., 2008, International Stroke Genetics et al., 2012)
ZFHX3	rs879324	16q22.3	CE	European	A	0.19	1.25 (1.15-1.35)	2.28×10^{-8}	(Traylor et al., 2012)	(Gudbjartsson et al., 2009)
CDC5L	rs556621	6p21.1	LVD	European	A	--	1.21 (1.13-1.30)	4.70×10^{-8}	(Holliday et al., 2012a)	--
FMNL2	rs2304556	2q23.3	YS	European /African	G	--	0.69 (0.60, 0.79)	1.20×10^{-7}	(Cheng et al., 2011)	--
ARL6IP6	rs1986743	2q23.3	YS	European /African	A	--	0.69 (0.60, 0.79)	2.70×10^{-7}	(Cheng et al., 2011)	--
ROBO1	rs1383407	3p12.2	IS	European	C	0.44	0.96	7.63×10^{-5}	(Meschia et al., 2011)	--

NINJ2	rs11833579	12p13.33	IS	European	A	0.23	1.41 (1.27-1.56)	2.3×10^{-10}	(Ikram et al., 2009)	(Traylor et al., 2012)
CELSR1	rs6007897	22	IS	Japanese	G	--	1.85 (1.29-2.61)	6.00×10^{-4}	(Yamada et al., 2009)	--
AGTRL1	rs9943582	11q12	IS	Japanese	G	--	1.30 (1.14-1.47)	6.66×10^{-5}	(Hata et al., 2007)	--
PRKCH	rs1452	14q22–q23	SVD	Japanese	A	0.23	1.40 (1.23-1.59)	5.10×10^{-7}	(Kubo et al., 2007)	--
ALOX5AP	HapA (SG13S25, SG13S114, SG13S89 and SG13S32)	13q12-13	IS	European	--	0.09†	1.67	9.50×10^{-5}	(Helgadottir et al., 2004)	--
PDE4D		5q12	CE & cryptogenic	European	--	0.16	--	1.50×10^{-6}	(Gretarsdottir et al., 2003)	--

*All p values are the lowest ever reported for a SNP by the main study under the 'Study' heading.

**'Other Studies' are studies that have also reported a GWAS significant association for a SNP.

†RAF reported in controls only

1.5.3.1 Limitations of GWA Studies on Stroke

GWA studies are plagued with many limitations. The high throughput nature of GWA studies requires large sample sizes for analysis in order to boost power to detect modest effect sizes. Resultantly GWA studies are labor and finance intensive, taking years to compile data and running into millions of dollars per project.

Recruitment: The manner in which cases and controls are recruited is also questionable. Ideally controls should be individuals with similar gender, age and environmental exposures. Most cohorts, as a norm and not an exception, collect DNA for cases and controls from different recruitment sites. Although all efforts are made to use age and gender matched controls, it is common practice to use standard common-pool controls such as the WTCCC 1958 British Birth Cohort controls, where there is a lack of locally recruited controls. This is mostly done to increase power of a study or to reduce costs.

Experimental errors in DNA collection, genotyping and quality control are notorious for adding to the problem of spurious results. The sources of DNA can vary from blood, tissue, buccal and saliva cells. This is further compounded by disparities in the DNA isolation methods, division of labor between different laboratory personal and the use of different genotyping platforms. Because experimental methods vary over time and location, severe batch effects can be introduced for samples genotyped on separate sets of plates leading to false positives (Leek et al., 2010). An analysis by Lambert et al of commonly occurring GWAS mistakes, recommends the involvement of a statistician in the design of both sample collection and protocols for performing subsequent measurements (Lambert and Black, 2012).

Analytical error: The choice of analytical and quality control methods can also lead to false positive results as has been implicated for the DeCODE GWA study (Worrall and Mychaleckyj, 2006). In a meta-analysis, the practice of pooling of family-based, population-based cohort and case-control studies can introduce heterogeneity due to survival bias, population stratification and differences in risk allele frequency. The use of different quality control thresholds, analytical and imputation software can also lead to introduction of heterogeneity. Adjustment for different covariates across studies can also be a source of heterogeneity.

Population admixture is yet another limitation that can lead to spurious results. Despite stringent quality control measures, a population based case control study possesses several inherent disadvantages, which limit our ability to make reliable conclusions. Underlying population sub-structure due to systemic ancestral differences is one such limitation, which can act as a confounding factor and shield a true association or inflate a false one. Most of the stroke studies have been conducted on the European population that is known to be genetically stratified (Lao et al., 2008) and admixture of populations with different ancestry can lead to inflated statistics. Several methods to control population sub-structure exist, such as ancestry-informative principal components (Price et al., 2006) and genomic control (Devlin and Roeder, 1999) but population sub-structure cannot be completely ruled out. Unreliability of test results may also lie in the lack of statistical power combined with multiple testing which can inflate test statistics and produce spurious results. Cryptic relatedness in individuals manifesting the same disease can also produce false positive results (Devlin and Roeder, 1999).

Effect size: Another major limitation of the GWA study model has been the inability to identify common genetic variants (>5%) with large effect sizes that exert an effect measureable at the population level. Most published studies have identified common variants with small to modest effect sizes for dichotomous traits (OR <1.5) and variance of <1% for quantitative traits (de Bakker et al., 2008). However, results from large stroke GWA studies have yet to identify a risk variant for over-all stroke and have only shown the sub-type specific association of susceptibility loci (Holliday et al., 2012). This is not surprising, as stroke is more of a clinical syndrome multiple etiologies. Researchers have therefore increasingly begun to analyze stroke genetic datasets by subtype.

Misspecification of outcomes: One of the major assumptions of GWA studies is that of association of the genetic risk variant with stroke cases as compared to non-stroke controls. However, there can be several other confounding factors, which might be associated with the risk variant, giving the false appearance of an association with stroke.

Lack of causality: GWAS help identify and attribute genetic mutations to complex polygenic disorders such as stroke and MI, just as it has been done for single gene

Mendelian disorders. However the reality is far from the truth. Although these studies help identify genes which are part of molecular pathways that may be involved in the etiology of stroke they fail to pinpoint the causal variant and the exact biological mechanism relevant to the disease. The clinical utility of this genetic information for prognosis and treatment is also very thin and has not found any major applicability in everyday clinical practice. GWA studies also provide a limited understanding of the gene-environment interaction, which may play a major role in the differential gene expression.

It is clear from the current understanding of GWAS results that there is a need for denser coverage of the genome to identify possible causal variants followed by validation of results at a transcriptomic and proteomic level.

1.5.3.2 Genotyping Platforms

Completion of the international HapMap consortium has enabled the development of commercially available genotyping microarrays and heralded the era of the GWA study. DNA microarrays have been widely used by scientists since the early 1990's for gene discovery in diseased and healthy populations. Microarrays not only allow researchers the ability to analyze hundreds of thousands of genes in a single experiment but they also provide the flexibility of studying specific parts of the genome. High-throughput gene expression analysis has become a popular and standard method in many genetics and molecular biology laboratories.

SNPs form the foundation of current GWA studies. These are the most common type of genetic variation in the human genome, which occur approximately every 1000 base pairs. Mutations of the genome that are stable across several generations and occur in >1% of the population are called SNPs. SNPs usually lie outside protein coding regions and are phenotypically silent (Perkel, 2008). Currently available genotyping arrays provide ~87% genome coverage for common SNPs which are found in >5% of the population.

The International HapMap Project (http://hapmap.ncbi.nlm.nih.gov), a collaborative effort between Japan, United Kingdom, Canada, China, Nigeria, and United States, has identified most of the approximately 10 million SNPs estimated to occur commonly in the human genome. This has enabled the rapid growth in the genotyping tools and

companies such as Illumina (San Diego, USA) and Affymetrix (Santa Clara, USA) have become leading producers of high throughput genotyping technologies. The gene chips developed by these companies share similarities in application and format of the chips, however they differ considerably in SNP selection strategy and hybridization chemistry (Perkel, 2008). While Illumina uses50-mer oligonucleotides per SNP, Affymetrix uses 25-mer oligo's spotted as 4-6 replicate probes per allele. Additionally, probes designed by Illumina are based entirely on haplotype-tagging tag SNPs identified by the International HapMap consortium while only half of the Affymetrix probes are tag SNPs; the remaining being 'unbiased' SNPs chosen to cover the genome while accommodating sequence restraints imposed by the assay itself (Perkel, 2008). Studies have tested the technical reproducibility of Affymetrix and Illumina SNP arrays and validated the reliability of both genotyping platforms (Hong et al., 2012).

The BRAINS-UK DNA samples were genotyped on the IlluminaHuman610-Quad bead chip, which was introduced in 2005, and the technology has since become obsolete. The study aims to perform whole-genome genotyping using Illumina's brand new Omni Family of Microarrays, which can currently deliver up to 5 million markers per sample and are designed from next-generation sequencing data from international projects such as the 1000 Genomes Project, hence offering denser genome coverage as compared to previous gene chips.

1.5.3.3 GWAS Consortiums Boost Power

GWAS models have so far enjoyed over whelming popularity and encouraged researchers worldwide to form large collaborations to exchange data. Combining genetic data from multiple studies not only increases the sample size but also the statistical power to detect small effects caused by common genetic variants. Ideally, meta-analysis should combine raw genotype and phenotype data from various cohorts allowing tests for epistasis, gene effects and gene-phenotype interactions (de Bakker et al., 2008). However in the real world, several institutional and ethical limitations do not facilitate the above scenario and exchange of data between research groups mostly takes the form of an informal look-up request of sentinel hits from discovery GWAS's or more formal requests for summary statistics to be included in genome wide meta-analysis. Cohorts perform site-specific analysis (quality control, imputation using

reference datasets from HapMap or 1000 Genome and association testing) and forward the association results to the lead group for inclusion in the meta-analysis.

For example, the METASTROKE collaboration combined data from 15 ischemic stroke cohorts with a total of 12,389 individuals with ischemic stroke and 62,004 controls. This was followed by a replication analysis of top hits in 13,347 ischemic stroke cases and 29,083 controls from 18 cohorts. For a minor allele frequency of 0.25, the study had 80% power to detect variants with a per-allele OR greater than 1.11 for the all ischemic stroke analysis, 1.23 for CE stroke, 1.24 for LVD, and 1.26 for SVD at $p \leq 5 \times 10^{-8}$ in the discovery phase (Traylor et al., 2012).

1.5.4 TDT and S-TDT

Family based genetic association tests such as the TDT are robust against bias introduced by population stratification and varying environmental exposure, which are problems plaguing several population based case-control studies. However, due to the difficulty in recruiting sufficient number of stroke case-patient trios and the resultant loss of power, such study designs require very large sample sets (Hassan et al., 2002). An alternative to TDT is the S-TDT study design which is useful for late-onset or age related diseases where parent genotype data may not be available.

Linkage studies have been widely used to study clinical phenotypes associated with the heritability of stroke and monogenic stroke, although they have only recently found applicability in analyzing common genetic risk variants. The SWISS was the first GWAS to examine the associations between genetic variants and risk of stroke and its subtypes within sibling pairs (Meschia et al., 2011). The family-based association analyses used the S-TDT, which is a simultaneous test of linkage and association based on transmission of alleles from parents to offspring, free of any population sub structure (Spielman et al., 1993). Findings showed that stroke subtype and age at stroke in affected sibling pairs exhibit significant clustering of genetic risk variants although no one variants reached genome-wide significance. 10 significantly associated SNPs covering 8 genomic loci were identified, of which 2 were promising risk variants on chromosome 3p (neuronal nitric oxide synthase) and 6p (Meschia et al., 2011).

In another study by Kopyta et al, TDT analysis was used to study pediatric stroke genetics in 392 individuals (81 ischemic stroke children, 162 biological parents and 149 control children). The study reported no significant preferential distribution of risk alleles (*Factor II, Factor V, Factor VII*, and *Factor XIII* genes) from parents to the affected children (Kopyta et al., 2012).

1.5.5 CNV and Exome Analysis

The genetic architecture of stroke is complex and is likely to include non-SNP variations as disease causing. Current association studies test the 'common disease-common variant' hypothesis assuming that the risk variant is commonly found in >5% of the general population. The variants tested are usually intronic or intragenic SNPs that do not provide any information on the role of protein coding part of the genome i.e. exomes. However the lack of reliable associations in the recent stroke GWA studies point towards the possibility of rare or low frequency variants with high penetrance and large effect sizes. These non-SNP variations could be very informative about the genetic underpinnings of stroke, yet they remain under represented in most association studies.

The National Heart, Lung, and Blood Institute (NHLBI) Exome Sequencing Project recently identified rare genetic variations in HDL-associated paraoxonase-1 (*PON1*) gene to be associated with ischemic stroke. In a sub group analysis of different ethnic populations, the study also showed that the variants had the strongest association with African Americans (Kim et al., 2014).

Another study identified rare exonic variants to be associated with stroke suggesting that coding variations in the human genome need to be closely examined (Cole et al., 2012). The study was small, consisting of only 10 stroke cases (8 African-Americans and 2 Caucasians) and a non-standardized methodology. 48 genes that had at least one rare variant across all stroke cases were identified. Gene*CSN3*,was found to contain an excess of rare variations as compared to other genes (Cole et al., 2012).

CNVs are large structural variations of the genome that include deletions, insertions, translocations, inversions and variable number repeats. CNVs alter the gene dosage without affecting function and are known to play a role in monogenic disorders although their role in complex traits such as stroke is unclear (McCarroll and Altshuler, 2007). GWA of CNVs associated with ischemic stroke has not identified any unique structural genomic variations that may contribute to the risk for stroke (Matarin et al., 2008). Smaller candidate gene based studies have provided some evidence of a unique genomic structure in ischemic stroke patients (Tiszlavicz et al., 2012), however large well-powered studies have failed to do the same (Norskov et al., 2011). The Human

Genome Structural Variation initiative started by the NHGRI to map structural variations within the human genome, is currently in the process of genotyping human CNVs and will add more information to the existing reference genome which may aid future stroke CNV analysis in identifying a disease-causing mutation.

1.5.6 Next Generation Sequencing and Rare Variants

GWAS studies function under the 'common disease common variant' hypothesis and have succeeded in identifying common genetic risk variants for stroke to which only a small effect size (<1.5) can be attributed. Although HapMap database has some rare variants, it is mostly the common SNPs that are genotyped. The general perception is that the 'missing heritability' lies with rare genetic variants, which are too infrequent to be picked up by commercially available genotyping platforms. Next generation sequencing is the latest technology to appear in the field of genetics, which offers a solution to this problem. The underlying hypothesis for NGS is that of 'rare variant common disease' (Bevan et al., 2011), which assumes that a rare causal genetic variant has a large effect on the development of stroke. The availability of the entire human genome via HapMap aided by advances in statistical computation makes it a promising strategy for studying genetics of stroke. The 1000Genomes project with whole genomes of 1000 healthy individuals will further provide dense coverage of both common and rare variants and add important information to the current knowledge base.

Although large-scale NGS approaches are already in the pipeline, some researchers are using targeted approaches to finding rare variants. Perkel et al report that companies such as Perlegen Sciences, Mountain View, California, USA, are using 454 Life Sciences (Roche) sequencing platforms to identify very rare genetic variants to study side effects to PPAR-gamma agonists, identifying two candidate genes for further analysis (Perkel, 2008).

Currently there are no NGS studies on stroke.

1.6. Aims of Thesis

This PhD thesis examined the genetic epidemiology of three unique risk factors that predispose individuals to ischemic stroke (South Asian ethnicity, blood pressure variability and carotid disease) using different study models: literature based meta-analysis, candidate gene based study and genome wide association study. Following are the aims of the thesis:

Aim 1 (Chapter 2): South Asian ethnicity predisposes individuals to the risk of stroke. However, the lack of large well-powered studies has obstructed the estimation of reliable risk associations. As the first aim of the study, a comprehensive systematic review and literature based meta-analysis of candidate gene studies examining gene polymorphisms associated with ischemic stroke in South Asians was conducted and the risk association was quantified by calculating pooled odds ratios. In an effort to establish causality, Mendelian randomization strategy was used to quantitate the level of risk for biomarkers.

Aim 2 (Chapter 3): A genome wide association study in blood pressure variability by the multinational ASCOT study identified a cluster of 17 correlated SNPs within the *NLGN1* to be associated with blood pressure variability. Since HTN and blood pressure variability are major risk factors for ischemic stroke, the study hypothesized that *NLGN1*SNPs may also be associated with stroke. In an effort to prove the hypothesis, the second aim of this study replicated the *NLGN1* SNPs in an independent European ischemic stroke population comprising of 8295 ischemic stroke cases and 12722 healthy controls from 7 different cohorts.

Aim 3 (Chapter 4): Carotid stenosis of >50% is associated with high risk of ischemic stroke and nearly 20% of all ischemic strokes and TIA's are caused by emboli originating from unstable plaque. As the final aim of the study, a GWAS was conducted to identify novel gene variants associated with advanced carotid disease in ischemic stroke. Seven independent cohorts comprising of >14,000 individuals were meta-analyzed using fixed-effect model and inverse variance method of weighted beta coefficients.

CHAPTER 2: GENE POLYMORPHISMS ASSOCIATED WITH ISCHEMIC STROKE IN SOUTH ASIANS: A LITERATURE BASED META-ANALYSIS

Abstract

The burden of stroke is disproportionately high in the South Asian subcontinent with South Asian ethnicity conferring a greater risk of ischemic stroke than European ancestry regardless of country inhabited. While genes associated with stroke in European populations have been investigated, they remain largely unknown in South Asians. A comprehensive meta-analysis of known genetic polymorphisms associated with South Asian ischemic stroke was conducted, and effect size of the *MTHFR C677T*-stroke association was compared with effect sizes predicted from Hcys-stroke association.

Electronic databases were searched up to August 2012 for published case control studies investigating genetic polymorphisms associated with ischemic stroke in South Asians. Pooled ORs for each gene-disease association were calculated using a random-effects model. 26 studies (approximately 2529 stroke cases and 2881 controls) interrogating 33 independent genetic polymorphisms in 22 genes were identified. Ten studies described *MTHFR C677T* (108 with TT genotype and 2018 with CC genotype) - Hcys relationship and six studies (735 stroke cases and 713 controls) Hcys-ischemic stroke relationship.

Risk association ORs were calculated for *ACE I/D* (OR 5.00; 95% CI, 1.17-21.37; p=0.03), *PDE4D SNP 83* (OR 2.20; 95% CI 1.21- 3.99; p=0.01), *PDE4D SNP 32* (OR 1.57; 95% CI 1.01-2.45, p=0.045) and *IL10 G1082A* (OR 1.44; 95% CI, 1.09-1.91, p=0.01). Significant association was observed between elevated plasma Hcys levels and *MTHFR/677* TT genotypes in healthy South Asians (Mean difference (ΔX) 5.18 μmol/L; 95% CI 2.03-8.34; p=0.001).

Findings from this study demonstrated that the genetic etiology of ischemic stroke in South Asians is broadly similar to the risk conferred in Europeans, although the dataset is considerably smaller and warrants the same clinical considerations for risk profiling.

2.1 Introduction

South Asia comprising of India, Pakistan, Sri Lanka and Bangladesh, forms 20% of the world's populous and shoulders much of the global death burden from cardiovascular disease (Ezzati and Riboli, 2012) with India reporting 930,985 cases of stroke in 2004 leading to 639,455 deaths and loss of 6.4 million DALYs (Goverment of India, 2008). The past decade has seen prevalence rates of stroke rise in South Asia from between ~200 (Banerjee et al., 2006, Saha et al., 2003, Gourie-Devi et al., 2004) to 545 per 100,000 persons (Das et al., 2007), while incidence studies demonstrate exponential increases in stroke incidence rates of >800% over the past 30 years (e.g. from 13 to 123 per 100 000 persons per year between 1969 and 1993) (Rao et al., 1971, Bhattacharya et al., 2005). Studies of South Asian Diasporas in the West have also shown a heightened prevalence and incidence of stroke and coronary artery disease compared to Caucasians (Potluri et al., 2009, Baweja et al., 2004). This is compounded by the fact that South Asians are developing stroke at a relatively young age, despite lower rates of alcohol and tobacco use (Biswas et al., 2009). The burden of stroke is high in South Asia and is likely to increase further due to demographic and epidemiological transitions in the South Asian populations. As lifestyles change and urban and rural populations are re-structured, South Asia is set to witness a surge in stroke occurrence.

Ethnic differences in the incidence of stroke (Tsai et al., 2013, Cappuccio et al., 1997) indicate differences in environmental exposure or genetic makeup underpinning the disease. Although 90% of the PAR for stroke is attributed to ten modifiable risk factors (O'Donnell et al., 2010), this does not account for the occurrence of stroke in young unexposed populations and also fails to explain the development of stroke in only some individuals within a population that is uniformly exposed to environmental risk factors. As with Europeans (Bentley et al., 2010, Bevan et al., 2012), it is likely that the causality of stroke in South Asians involves the complex interaction between genetic and environmental risk factors.

We undertook a comprehensive meta-analysis of all known genetic associations with ischemic stroke in South Asians and compared it to published results in different ethnic groups. We further sought to establish whether Hcys, the putative biochemical intermediary of the *MTHFR* gene is associated with quantitative levels of risk (Casas et

al., 2005) in South Asians as similarly shown in Europeans (Casas et al., 2005, Bentley et al., 2010). Our meta-analysis strategy was to pool candidate gene based genetic studies into a statistical model and calculate more reliable pooled odds ratios, bearing in mind that most such are underpowered as stand-alone studies. Such candidate gene studies are based on a pathophysiological *priori* hypothesis tested by investigators. This was the model upon which we sought to provide further and more robust clarification.

2.2 Methods

2.2.1 SNP Analysis

2.2.1.1 Data Searching

We identified all published case-control studies evaluating any gene polymorphisms and ischemic stroke in South Asians residing in their native countries and Diasporas around the world. Electronic searches were conducted using Medline, EMBASE and Google Scholar. All published manuscripts up until and including August 2012 as well as letters, previous meta-analyses and abstracts were included. The retrieved studies were examined thoroughly to assess their appropriateness for inclusion in our study. The references of all identified publications were manually reviewed for additional studies and the PUBMED *'relevant articles'* option was utilized. The following index terms along with *'and/or'* as a Boolean operator were used: ''South Asia'' ''India'' ''Pakistan'' ''Sri Lanka'' ''Bangladesh'' for ancestry and ''stroke genetics'' ''gene polymorphism'' ''gene mutation'' ''stroke genes'' for genetics and ''stroke'' ''cerebrovascular disease'' ''ischemic stroke'' ''brain infarction'' ''brain ischemia'' for clinical phenotype.

2.2.1.2 Study Selection

Study inclusion criteria were: (1) studies in populations of South Asian descent defined as Indian, Pakistani, Sri Lankan or Bangladeshi; (2) case-control studies where ischemic stroke was analyzed as a dichotomous trait; (3) stroke was confirmed using brain imaging with sub-acute (within 10 days) CT or MRI, and; (4) genotype frequency for both cases and controls was reported. Studies were excluded if: (1) subjects were <18 years age; (2) the genotype frequency was not reported and could not be obtained from authors and (3) stroke other than ischemic.

2.2.1.3 Data Extraction

Data extracted from each study included: first author, journal, year of publication, stroke sub-type, and number of cases and controls for each genotype and SNP. Baseline characteristics for cases and controls were documented including mean age, gender, ethnicity and geographical location.

2.2.1.4 Data Analysis

For each genetic variant for which data were available from at least two studies, a meta-analysis was carried out. Data was analyzed using Review Manager *v*5.0 and Comprehensive Meta-Analysis *v*2.0. Using a recessive genetic model we tested the carriers of mutant homozygous genotypes (mt/mt) versus wild type heterozygous and wild type homozygous (mt/wt + wt/wt) genotype carriers. To test for strength of association for each gene variant, pooled OR and 95% CI were calculated using a random effects model (DerSimonian and Laird, 1986). The OR is the odds of an event occurring in one group divided by the odds of the event occurring in another group. For a recessive model, an OR was calculated using the following formula:

$$Odds Ratio = \frac{Odds\, of\, subject\, with\, stroke\, when\, mt/mt}{Odds\, of\, subject\, with\, stroke\, when\, \frac{mt}{wt}\, and\, wt/wt}$$

If the mutant and wild type alleles have an identical effect the OR will be 1. If the mutant allele increases the risk of stroke, the OR will be greater than 1; if it reduces the risk of stroke the OR will be less than 1. In a recessive genetic model an OR of > 1 implies that carriers of the homozygous mutant (mt/mt) genotype have a greater risk of ischemic stroke than carriers of the wild type (wt/mt and wt/wt) genotypes. Effect sizes can be inflated or underestimated and 95% confidence intervals were calculated as a way of representing the uncertainty in the estimation of the OR. The 95% CI contains a range of values which are above and below the effect size (OR) within which we can say with 95% certainty that the real effect lies. The results is considered statistically significant if the 95% CI does not contain the value 1 which implies that the mutant and wild type alleles have an identical effect. The strength of genetic association (risk or protection) for each polymorphism was tested using a chi square test and considered statistically significant if a p-value of <0.05 was obtained (Cochran, 2010).

Statistical heterogeneity was determined in 2 ways. First forest plots were observed for overlapping confidence intervals. If the CI's of studies were far apart from each other and did not overlap, heterogeneity was reported. Next, for each analysis a chi squared

and I^2 index test for heterogeneity (Deeks, 2008) was performed, with significance set at $p \leq 0.10$ rather than the conventional $p \leq 0.05$. This is because the test is not sensitive enough to detect heterogeneity in low powered studies. The chi-squared test generates a chi-squared statistic; degrees of freedom and a p value obtained by referring the first two numbers to statistical tables and measure the presence vs. absence of heterogeneity. Heterogeneity was reported if the chi squared statistic was greater than the df. The I^2 index test compliments the chi-squared test and quantifies (in percentage) the extent of heterogeneity in a meta-analysis (Cochran, 2010).

Finally, funnel plots and Egger linear regression intercept p-values (two-tailed) were used to determine probability of publication bias (Egger et al., 1997). In funnel plots, study specific standard error was plotted against log OR's and if the plot were not symmetrical i.e. did not resemble an inverted funnel, publication bias was reported. Observing whether 95% confidence intervals overlapped made comparison of ORs of genes with significant associations.

2.2.2 MTHFR C677T – HcysPhenotype Comparison

For MTHFR C677T gene variant which had an associated biomarker Hcys, we performed a separate analysis that produced an estimate of expected risk based upon genotype–biomarker, and biomarker–stroke, association studies, using Mendelian randomization (Bentley et al., 2010, Casas et al., 2005, Wald et al., 2003).

2.2.2.1 Data Search Criteria

Electronic searches were conducted using Medline, EMBASE and Google Scholar and all published manuscripts up until and including August 2012 were considered. The following index terms along with *'and/or'* as Boolean operators were used for MTHFR: (MTHFR OR *Methylenetetrahydrofolatereductase*) AND (gene OR genetic OR genotype OR polymorphism OR mutation), in combination with (MTHFR OR *Methylenetetrahydrofolatereductase*) AND (Hcys OR activity OR level); and 2) (MTHFR OR *Methylenetetrahydrofolatereductase*) AND (activity OR level) in conjunction with (CVDOR brain infarction OR stroke OR cerebral ischemia) and (South Asia or India or Pakistan or Sri Lanka or Bangladesh).

2.2.2.2 Study Selection

The literature was searched for two types of studies: (Study 1) case-control studies reporting dichotomous and continuous data for plasma Hcys levels in South Asians with ischemic stroke, and; (Study 2) studies linking plasma Hcys levels with the MTHFR 677 wild type (CC) and homozygous mutant (TT) genotype in healthy South Asians. Control populations with genotype-Hcys data from case-control studies on psoriasis, coronary artery disease and glaucoma were also included.

2.2.2.3 Data Extraction

Data extracted from each study included: first author, journal, year of publication, stroke sub-type, and total number of participants. Baseline characteristics for cases and controls were documented including mean age, gender, ethnicity and geographical location. In addition for Study 1, plasma Hcys levels as means with standard deviations and median with ranges in ischemic stroke cases and controls, and; for Study 2, plasma Hcys levels as means with standard deviations and median with ranges in *MTHFR* 677 wild type (CC) and homozygous mutant (TT) genotype in healthy South Asians were extracted.

2.2.2.4 Data Analysis

The methodology for a literature based Mendelian randomization for MTHFR and Hcys is well described (Bentley et al., 2010, Casas et al., 2005). Where values of plasma Hcys were reported as medians and ranges, the mean and standard deviation were estimated using established models dependent upon sample size (Hozo et al., 2005). For both types of studies, mean Hcys difference between ischemic stroke cases *vs.* controls and MTHFR TT *vs.* CC genotypes were calculated using the continuous data type inverse variance method in Review Manager *v*5.0. A pooled mean difference (ΔX)with 95% confidence interval was calculated for *MTHFR* TT *vs* CC genotypes using a random effects model. Mean difference in Hcys levels between ischemic stroke cases *vs.* controls were converted to ORs and 95% confidence interval using Comprehensive Meta-Analysis (CMA) *v*2.0. Assuming a log-linear relationship, this OR was scaled with the ΔX change in Hcys levels conferred by the *MTHFR* TT genotype to give the expected OR (Casas et al., 2005).

For each meta-analysis an I^2 test for heterogeneity was performed, with significance set at $p<0.05$. Funnel plots and Egger regression intercept p-value (two-tailed) were used to determine probability of publication bias.

2.3 Results

2.3.1 Meta-Analysis of Gene Variants Associated with Ischemic Stroke

Our search strategy identified ~4500 potentially relevant studies of which 26 met the inclusion criteria allowing interrogation of 33 independent genetic polymorphisms in 22 different genes across 2529 stroke cases and 2881 controls (**Figure 2.1**). The majority of studies were from India (North and South) followed by Pakistan, Bangladesh, Malaysia and the United Kingdom (**Table 2.1**).

Relevant studies identified were; five for *MTHFR C677T* (Biswas et al., 2009, Alluri et al., 2005, Panigrahi et al., 2006, Somarajan et al., 2011, Mejia Mohamed et al., 2011), three for *PDE4D SNP 83* (Saleheen et al., 2005, Munshi et al., 2009, Banerjee et al., 2008) and 2 studies each for *eNOS 4a/4b* (Majumdar et al., 2009, Munshi et al., 2010), *ACE* (Munshi et al., 2008, Kalita et al., 2011), *ApoE E4/E4* (Luthra et al., 2002, Chowdhury et al., 2001), FVL *G1691A* (Dindagur et al., 2007, Dindagur et al., 2006), *PDE4D SNP 87* and *32* (Saleheen et al., 2005, Munshi et al., 2009) and *IL10* (Munshi et al., 2010, Sultana et al., 2011) genes. Two studies each reporting different stroke sub types (CVT and arterial pediatric ischemic stroke) were found for genes *MTRR G66A* (Biswas et al., 2009, Biswas et al., 2009), *MTR A2756G* (Biswas et al., 2009, Biswas et al., 2009), *MTHFR A1298C* (Biswas et al., 2009, Biswas et al., 2009) and *FVL A4070G* (Biswas et al., 2009, Biswas et al., 2009, Dindagur et al., 2007), and therefore ORs for risk were not estimable. *Prothrombin G20210A* polymorphism was investigated in two studies (Dindagur et al., 2006, Munshi et al., 2009). Homozygous or heterozygous variants were completely absent in all individuals studied and with only the wild type genotype being expressed an OR for risk was not estimable.

One relevant study was identified for genes variants *IL-1α C889T* (Banerjee et al., 2008), *CYP11B2 C344T* (Munshi et al., 2010), *ESR1PVUII and XbaI* (Munshi et al., 2010), *α ADD1 WG* (Kalita et al., 2011), *TNF α G488A and G308A* (Munshi et al., 2011), *CYP4F2 G1347A* (Munshi et al., 2012), *MDR-1* (Sharma et al., 2011), *t-PA C7351T and I/D* (Babu et al., 2012), *PAI-1 4G/5G* (Babu et al., 2012), *CBS T833C* (Chandra, 2006), *Klotho KL-VS and C1818T* (Majumdar et al., 2010), *Factor XIIIB V34L* (Kain et al., 2005), *α1antichymotrypsinAla15Thr* (Somarajan et al., 2010) and *MMP3 5A/6A* (Munshi et al., 2011). Of these, *IL-1α*, *CYP11B2*, *ESR1 PVUII*, *α ADD1*, *TNF α G488A*), *CYP4F2*, *MDR-1* and

t-PA I/D were found to be significantly associated with ischemic stroke (**Table 2.1**). The following are the results for the most significantly associated (p≤0.05) gene variants with 2 or more studies.

2.3.1.1 PDE4D SNP 83 (rs966221)

The *PDE4D SNP83* polymorphism was investigated in three studies covering Pakistan and North India (Saleheen et al., 2005, Munshi et al., 2009, Banerjee et al., 2008) in a total of 1338 subjects (626 Ischemic stroke cases; 712 controls). A pooled OR of 2.20 (95% CI, 1.21-3.99; p=0.01) was generated with a recessive random-effects model. There was no evidence of inter-study heterogeneity ([P_{HET}] = 0.06, I^2 = 65%) **(Fig. 2.2)**.

2.3.1.2 ACE I/D (rs4646994)

The homozygous *ACE D/D* polymorphism was investigated in a total of 2 studies covering India (Munshi et al., 2008, Kalita et al., 2011) in 693 subjects (355 ischemic stroke cases; 338 controls) providing a pooled OR of 5.00 (95% CI, 1.17-21.37; p=0.03) with a recessive random-effects model. There was evidence of inter-study heterogeneity ([P_{HET}] = 0.002, I^2 = 90%) **(Fig. 2.3)**.

2.3.1.3 IL10 G1082A (rs1800896)

The homozygous *IL10 G1082A* polymorphism was investigated in a total of 2 studies (Munshi et al., 2011, Sultana et al., 2011) covering India in 1414 subjects (718 ischemic stroke cases; 696 controls) providing a pooled OR of 1.44 (95% CI, 1.09-1.91; p=0.01) with a recessive random-effects model. There was no evidence of inter-study heterogeneity ([P_{HET}] = 0.12, I^2 = 59%) **(Fig. 2.4)**.

2.3.1.4 MTHFR/C677T (rs1801133)

The homozygous MTHFR C677T polymorphism was investigated in seven studies (Biswas et al., 2009, Alluri et al., 2005, Panigrahi et al., 2006, Somarajan et al., 2011, Dindagur et al., 2007, Biswas et al., 2009, Kalita et al., 2006), of which five studies covering North Indians and Malaysian Indians (Biswas et al., 2009, Alluri et al., 2005, Panigrahi et al., 2006, Somarajan et al., 2011, Dindagur et al., 2007) were included in the final analysis. These totaled 615 subjects (309 ischemic stroke cases and 356 controls) providing a pooled OR of 2.50 (95% CI, 0.89-6.97; p=0.08) with a recessive genetic and

random-effects model **(Fig. 2.5)**. There was no evidence of inter-study heterogeneity ($[P_{HET}]=0.33, I^2=14\%$).

2.3.2 Comparison with Different Ethnic Groups

Comparisons of genetic risk of ischemic stroke for our top four gene polymorphisms were made with different ethnic populations reported in published literature (**Figure 2.6**). *PDE4D (SNP 83)* was associated with risk for South Asians and Chinese but conferred protection for the Japanese. The gene showed a (non-significant) trend for protection against risk in a European descent population (Bevan et al., 2008). Odds ratios for *MTHFR C677T* gene polymorphism had relatively uniform association of risk between the different ethnic groups with overlapping 95% confidence intervals. Except for the current study, odds ratios for *ACE I/D* gene polymorphism were also uniformly distributed. We discuss the various reasons for the possible inflation of effect size in our study in the discussion section. *IL10 (G1082A)* gene polymorphism was associated with an increased risk of stroke for Caucasians as compared to South Asians; however the result is probably inflated due to inclusion of only 2 studies to calculate the pooled effect size.

2.3.3 Biochemical Marker of Risk: MTHFR C677T and Hcys

Seven studies (Modi et al., 2005, Kalita et al., 2009, Dhamija et al., 2009, Biswas et al., 2009, Mamatha et al., 2011, Mejia Mohamed et al., 2011, Narang et al., 2009) that presented plasma Hcys levels in South Asian ischemic stroke patients were subjected to a meta-analysis. One study (Dhamija et al., 2009) did not have normally distributed data with a threefold difference in mean Hcys levels between cases and controls, and the study was excluded from the final analysis. The remaining six studies presented data as means and standard deviations as well as median with their lower and higher range.

A mean difference in Hcys between ischemic stroke cases and controls was found to be 2.90 µmol/L (p=0.02, 95% CI 0.45-5.34) and this corresponded to an OR of 1.68 (95% CI 1.10-2.58), calculated using the CMA *v*2.0 software. There was evidence of inter-study heterogeneity ([P_{HET}] < 0.0001, I^2 = 95%) **(Fig 2.5)**. We found 10 studies (Dhillon et al., 2007, Chambers et al., 2000, Kumar et al., 2005, Liew et al., 2012, Iqbal et al., 2005, Misra et al., 2010, Tripathi et al., 2010, Micheal et al., 2009, Sukla et al., 2012, Rama Devi et al., 2004) linking Hcys levels to the *MTHFR 677* wild type (CC) and homozygous mutant (TT) genotypes in healthy South Asians and calculated a mean difference (ΔX) of 5.18 µmol/L (p= 0.001, 95% CI 2.03-8.34) between the TT vs. CC genotypes in healthy

individuals **(Fig. 2.5)**. There was evidence of inter-study heterogeneity ([P$_{HET}$] < 0.0001, I^2 = 78%).

The expected odds ratio was calculated using the following formula:

$$ExpectedOR = 1.68^{5.18/2.90}$$

Where 1.68 = OR associated with 2.90 μmol/L difference in Hcys levels between stroke cases and controls

ΔX = 5.18 μmol/L mean difference in Hcys levels between the *MTHFR* TT vs. CC genotypes in healthy individuals

MD = Mean difference of 2.90 μmol/L associated with an OR of 1.68

The above calculation revealed an expected OR of 2.52. A 95% confidence interval for the logged odds ratio of 0.92 (LN 2.52) was obtained as 1.96 standard errors on either side of the point estimate as previously described by Bland et al (Bland and Altman, 2000). Standard error (SE) was calculated as the square root of the sum of reciprocals of the frequencies i.e. number of cases and controls. The 95% confidence interval was calculated using the formula LN (OR) ±1.96×SE, which generated a range of -1.17 to 3.02. The antilog of these limits generated a 95% confidence interval for OR 2.52 as Exp (-1.17) =0.31 to Exp (3.02) =20.65. A meta-analysis relating MTHFR 677/TT vs. CC risk genotype to disease in our current study generated a pooled observed OR of 2.50 (95% CI, 0.89-6.97). The observed OR was close to the expected OR and its 95% confidence interval fell entirely within the confidence interval for the expected OR **(Fig. 2.7)**.

2.4 Discussion

The current study is the most comprehensive genetic meta-analysis of ischemic stroke in South Asians from India, Pakistan, Sri Lanka and Bangladesh. Findings suggest three genes (*PDE4D, ACE I/D* and *IL10*) to have statistically significant ORs for risk of ischemic stroke in a South Asian adult population and the totality of data supported MTHFR C677T as a further likely risk factor. Results from 12 other gene variants, although significant, were of insufficient power to allow robust conclusions. The remaining 17 gene polymorphisms **(Table 2.1)** failed to support any significant association either due to lack of sufficient studies, paucity of subjects studied and/or heterogeneity in the stroke sub-types investigated.

The two most significant risk associations, *PDE4D SNP 83* and *ACE I/D* identified, were previously described as genetic risk factors for stroke in other ethnic groups (Xu et al., 2010). There was a doubling in the odds (OR 2.20; 95% CI, 1.21-3.99) or a ~120% increase in the risk of developing stroke per copy of the risk allele of PDE4D SNP 83 for overall ischemic stroke. *PDE4D* gene encodes a phosphodiesterase enzyme that regulates cAMP levels in the body (Conti et al., 2003) and was first identified as a candidate gene for stroke by the DeCODE study (Gretarsdottir et al., 2003). However several attempts to replicate these findings failed (Rosand et al., 2006), while some studies reported conflicting results (Bevan et al., 2008, Matsushita et al., 2009, Bentley et al., 2010). These discrepancies have been attributed to possible problems in study design and analytical methods (Worrall and Mychaleckyj, 2006). A recent meta-analysis by Yoon et al (Yoon et al., 2011) identified SNP 83 as having a protective association with stroke in Asians (OR 0.79, 95% CI 0.69-0.90; p=0.0005). The meta-analysis combined 4 studies of South East Asian origin and 2 studies of South Asian origin. While combining different ethnic populations for the analysis maybe a reasonable approach for an exploratory analysis of genetic risk factors, such assumptions may lead to fallacious results as studies have already proved that India has a unique genetic population sub-structure which cannot be imputed from other ethnic groups (Reich et al., 2009). Other studies have also identified race-ethnic disparities for stroke risk factors such as HTN and diabetes (Sacco et al., 2001) as well as genes affecting stroke in different ethnic groups (Hall et al., 2000). The current study included individuals of only

South Asian ancestry from the Indian sub-continent and though the population size is small, it takes into account the need for independent analysis of South Asians.

ACE gene plays an important role in vascular physiology and structural integrity. This study finds a high risk of *ACE/DD* variant in South Asians (OR 5.00: 95% CI, 1.17-21.37) which accounts for a five-fold increase in the risk of overall ischemic stroke per copy of the risk allele of the *ACE DD* variant relative to Europeans (OR 1.15: 95% CI, 1.06-1.25) (Bentley et al., 2010). Closer examination of this comparison revealed a very large confidence interval for our study largely due to low sample size (355 cases/338 controls vs. 4897 cases/13949 controls in Europeans) suggesting that the measured effect size is probably inflated. Partial overlapping confidence intervals further suggested that there might not be a statistical difference between the two groups. An alternate explanation could be the difference in the prevalent stroke sub type in South Asians as compared to other ethnic groups. The homozygous *D* allele is associated with HTN in Indians (Tao et al., 2009) and is also associated with preferential risk of small vessel disease (SVD) (Rao et al., 2009), which is the most common subtype of stroke found in South Asians (Gunarathne et al., 2009, Biswas et al., 2009). The polymorphism accounts for 47% of the total phenotypic variance of serum ACE (Rigat et al., 1990) that is linked with quantitative levels of risk of disease (Munshi et al., 2010, Bentley et al., 2010).

Interleukin 10 is an anti-inflammatory cytokine produced primarily by monocytes and type 2 T helper cells. The study found *IL10* to be associated with a 44% increase (OR 1.44; 95% CI 1.09-1.91) in the risk of overall ischemic stroke per copy of the risk allele of the *IL 10 G1082A*variant. IL10 is involved in various cellular processes such as inhibition of pro-inflammatory cytokines, suppression of antigen-presenting capacity of antigen presenting cells (APC) and stimulation of B cell maturation. IL10 forms part of an inflammatory genetic profile and elevated levels of IL10 post stroke have been implicated in severe neurological impairment and major adverse clinical outcomes (Chang et al., 2010, Dziedzic et al., 2002). Since acute ischemic brain insult is known to trigger anti-inflamatories, mutations in IL10 may result in uninhibited effects of pro-inflammatory cytokines.

MTHFR C677T genotype is associated with hyperhomocystinemia among South Asians and Europeans (Kelemen et al., 2004) and has been shown to have a larger effect on Hcys concentration in geographical regions of low folate consumption than in regions with high dietary folate intake (Holmes et al., 2011). Knowing whether there is an ethnic predisposition to hyperhomocystinemia is necessary as homocystiene levels can be effectively lowered by supplementing diet with folic acid, vitamin B12 and vitamin B6. This is important as the majority Hindu South Asian population has low plasma folate and Vitamin B12 levels (Hughes and Ong, 2000, Chandalia et al., 2003) which can be partly accounted for by their predominantly vegetarian diets or high temperature cooking methods which destroy folate. The study sought to establish whether the putative biochemical intermediary, Hcys, of the *MTHFR C677T*gene variant was associated with equivalent quantitative levels of risk (Casas et al., 2005) in South Asians.

Effect sizes of gene-stroke associations were compared with those predicted from independent biochemical data using a Mendelian randomization strategy which utilizes genetic variants or their cumulative risk allele scores (Burgess & Thomson, 2013) as instrumental variables (IV) for exposures of interest (in this case ischemic stroke) to overcome problems of confounding and reverse causality. MR strategies are widely used for analyzing causal relationships in genetic epidemiological studies using both summary data from published studies (Kunutsor et al., 2014, Huang et al., 2013, Bentley et al., 2010) and individual patient level data from prospective studies (Holmes et al., 2014a, Holmes et al., 2014b, IL6R MR Consortium, 2012).

Although the genetic association between *MTHFR C677T* and ischemic stroke was weak (p=0.08), and study results did not confirm causality, the totality and direction of effect of data supported a causal relationship, with support from previously documented strong association in Europeans (Casas et al., 2004). A Mendelian randomization strategy for a causal association was concordant with literature on Europeans (Casas et al., 2005, Bentley et al., 2010) with similar observed and expected odds ratios. Findings from the study also showed a significant association (p=0.02) between Hcys levels and ischemic stroke where a 2.90μMol/L elevation in plasma Hcys levels increased the odds of stroke risk by 68%. This is in contrast to reports from European studies, which have observed much lower odds of risk (24%) for a similar increment in plasma Hcys levels (Wald et al., 2003). Interestingly, the biochemical change C *vs* T also showed a high

difference in Hcys levels, ΔX value of 5.18 µMol/L, for South Asians, compared to Europeans ΔX of ~2.30µMol/L (Bentley et al., 2010). While it is possible this is an artefactual difference, another explanation is that healthy South Asians with CC genotype have overall higher levels of Hcys than healthy CC genotype Caucasians resulting in a greater ΔX. Alternately, the study population maybe largely vegetarian and resultantly have high Hcys levels (Bissoli et al., 2002). Epidemiological studies (Chandalia et al., 2003) have already established ethnic differences in plasma Hcys levels and these need to be followed up with comparative analysis of Hcys levels in healthy South Asians and Europeans with *MTHFR C677C* genotypes.

Study Strengths

Although the ideal data to address the genetics of ischemic stroke in South Asians would be experimentally genotyped from large study populations, the next best place to begin is published candidate gene based studies. The meta-analysis uses statistical methods to combine results of individual studies that are similar in their metrics and outcome measures, thereby increasing the power of the analysis. The method does not simply pool the events and non-events in different studies to generate an effect size but keeps each study discrete so as to preserve the effects of randomization and compare like with like. Single effect measure for each study (OR) were calculated and pooled to generate estimates across studies using a weighted average method that takes into account the sample size of each study. The larger studies are assumed to have a greater effect on the outcome measure and are given more weightage in the analysis. A random effect model, which takes inter-study heterogeneity into account based on the assumption that the true effects in individual studies differ from each other and are not common or fixed, was used. A random effects model also assumes that different true effects are normally distributed and based on this our meta-analysis calculated means and standard deviations which were converted to OR's.

Study Limitations

The reliability of a meta-analysis depends on the accuracy and quality of the included data. Systemic reviews bring together studies from various sources, which are conducted by different analysts in varied ways. These differences can introduce

heterogeneity in the meta-analysis and lead to spurious results. As with all meta-analyses, a number of limitations to interpreting results need to be reported.

1. Publication bias: Possible publication bias (Lin et al., 2006) could be introduced by selective outcome reporting (only positive results) in the English language (language bias) as well as an unconscious exclusion of studies that may have reported contradictory results. Small studies are more vulnerable to publication bias since smaller studies are less likely to be published unless they demonstrate a positive statistically significant result. On the other hand, irrespective of the outcome larger studies are more likely to be published. An excess of positive studies in the meta-analysis can lead to miscalculation of an effect size associated with a polymorphism. Therefore it is critical to make all possible efforts to identify unpublished, negative and non-significant genetic association studies in order to make the analysis well balanced and reliable. Other related biases such as time lag bias (more likely to be published rapidly) (Stern & Simes, 1997), multiple publication bias (more likely to be published many times), location bias (published in high impact indexed journals) and citation bias (more likely to be cited by others) can also lead to over-representation of significant positive results over non-significant negative findings.

In order to overcome limitations introduced by bias, the current study made every possible effort to identify all published and unpublished genetic association studies in South Asian ischemic stroke. Authors were contacted if any reported study results were unclear or insufficient for analysis.

2. Small study effects: Estimates from small genetic association studies are generally viewed with suspicion unless replicated in larger studies. Small studies are vulnerable to higher sampling random errors and their observed effect sizes may be markedly different (higher or lower) from the true effect. Such studies are also known to be consistently more positive or negative in their findings than larger association studies. A meta-analysis consisting of small genetic association studies may lead to an over or under estimation of the actual effect of the risk allele on a phenotype.

RevMan 5 software offers several ways to test whether the results from a meta-analysis are associated with the presence of small study effects. This includes visual inspection of funnel plots, statistical testing and sensitivity analysis (comparing fixed and random

effects models). Funnel plots plot the individual study effect sizes against the study size, which is represented by the standard error of the study. The studies are scattered around the effect size and a symmetrical funnel plot indicates lack of study bias. Unfortunately funnel plots are not very meaningful for meta-analysis of ≤ 10 studies which was the case with the current study.

There is considerable paucity of genetic studies on stroke in South Asians and our meta-analysis highlights the need for larger statistically well-powered prospective and retrospective studies in this population. Studies were smaller and much fewer in numbers leading to less robust conclusions compared to similar work in Europeans (Ioannidis and Panagiotou, 2011). As an example, for *ACE I/D* polymorphism the availability of only 2 studies for the meta-analysis prevented us from conducting an iterative sensitivity analysis to identify and remove the source of inter study heterogeneity ([P_{HET}] = 0.002, I^2 = 90%). The small sample size meant a large confidence interval and less statistical reliability of data.

3. Case control study design: Although case control genetic epidemiology studies have several advantages compared to cohort or family studies including being less time and resource intensive as well as being better models for studying outcomes with long dormant periods following exposure (such as stroke), they also have several limitations. Case control studies provide an estimate of relative risk and no information on the gene-environment interaction. Resultantly, detected association may be prone to reverse causation. Case control studies are also prone to various forms of bias (selection bias, observation bias, recall bias and misclassification) and do not control for confounding risk factors.

4. Confounding: Confounding due to environmental stroke risk factors such as hypertension, diabetes and smoking can adversely affect the outcome measure of a meta-analysis. Alternative methods such as meta-regression can take such confounders into account however the small number of studies included in this study did not allow for such an analysis. Adjustment for population stratification is another important consideration with individual patient level data but it was not possible to adjust a literature based meta-analysis for confounders. Fewer studies also limited our ability to take the genetic stratification of the South Asian population (Reich et al., 2009) into

account and a pooled analysis of both South and North Indian stroke cases was conducted. The analysis was primarily based on a mixture of published studies from North and South India. The analysis attempted to control for heterogeneity by setting strict inclusion and exclusion criteria for studies and the use of a random effects model.

5. Inadequate reporting by studies: Another limitation was that results of candidate SNPs from large stroke GWAS's could not be tested although this was considered in the initial analysis plan. To the best of our knowledge no GWA studies specifically on South Asian stroke cases have been conducted and only a few studies have been conducted where some of the population included South Asian stroke, although these were not recruited in a systematic manner as that was not the primary purpose of those studies.

6. Stroke sub-typing: The genetic burden of ischemic stroke is known to lie within specific sub-types (Holliday et al, 2012). Lack of stroke TOAST classification amongst the selected studies hampered sub-group analysis that is essential due to the heterogeneous etiology of ischemic stroke sub-types.

7. Use of published summary data: There are several disadvantages of using published data for a meta-analysis as compared to individual patient level data. With published studies the methods used for data generation and quality control are unknown which may lead to the inclusion of unsuitable data. Since the reported summary statistics cannot be altered, downstream analysis options are limited. Sensitivity and sub-group analysis may also not be available with meta-analysis of published studies.

8. Mendelian Randomization: Mendelian randomization strategy used in this chapter is useful to assess causal effects, however this methodology has several limitations that merit a detailed discussion. MR makes several assumptions such as, the genetic marker is associated with the exposure (Hcys), the genetic marker is independent of the outcome (stroke) (also called exclusion restriction) and the genetic marker is independent of factors that confound the exposure–outcome relation (VanderWeele et al, 2014). An ideal situation would be where the genetic marker affects the outcome only through the exposure; however, several scenarios violating the above assumptions can lead to biased results.

8a. Inadequate selection of instrumental variable: The inability to calculate the F statistic to determine the strength of the genetic instrumental variable is a major limitation and can lead to severe bias. A Mendelian randomization strategy on individual patient level data involves the assessment of a causal effect of a phenotype on an outcome by using genetic IV (Burgess and Thompson, 2011). The magnitude of bias is determined by the F statistic for the strength of association between the IV and phenotypes and is essential to assess the suitability of the IV. The F factor is calculated from the regression of the risk factor on the IV and a value of < 10 is deemed as a weak IV. A weak instrument has little effect on the exposure and bias is caused due to over fitting of the genetic model where the IV is forced to account for not only the systematic variation but also the chance variation in the confounders (Burgess & Thomson, 2013).In a literature-based meta-analysis pooled F statistics can be calculated if each study included in the meta-analysis reports the study specific values. However none of the candidate gene based studies included in our meta-analysis conducted a Mendelian randomization or reported the study specific F statistics and only reported the allele and genotype frequencies with baseline population phenotypes. Previously published literature based Mendelian randomizations on the *MTHFR C677T*-Hcys-stroke have also not reported F statistics due to similar reasons (Casas et al., 2005, Bentley et al., 2010).

8b. Lack of statistical power: Genetic risk variants typically exert a very small effect on the phenotypic variation of a trait and this makes power a very important consideration for a Mendelian randomization. As with genetic association studies, large samples sizes are required to meet the power requirements of MR studies. Previous studies have calculated power estimates for MR of continuous variables and a single IV using the F statistic and R^2 from the first stage regression analysis of the exposure phenotype and genetic variant (Brion et al., 2013).

8c. Gene-environment interaction: A major assumption of Mendelian randomization is that the genetic variants affect the outcome (stroke) only through the exposure (Hcys) and not external environmental factors (such as folic acid and vitamin supplements). However, gene–environment interactions are common and may involve both the genetic risk variant and environmental factor in causing the outcome.

8d. Canalization: Canalization is a process of developmental compensation by which the effect of potentially disruptive genetic and/or environmental influences on normal development is reduced. Several molecular mechanisms including genetic redundancy, feedback regulation and cooperative biochemical interactions are known to play a role in canalization. This phenomenon could invalidate findings from Mendelian randomization studies by altering the effect of a genotype on the outcome of interest in adulthood without any effect on the association between genotype and the modifiable exposure of interest (Lawlor et al, 2008).

8e. Linkage disequilibrium and pleiotropy: One of the most common limitations of Mendelian randomization includes confounding by SNPs in high LD with the polymorphism being studied. If the genotype being used as an instrument and other polymorphisms associated with the outcome are in high LD, then there will be a violation of the MR assumptions, which will lead to a confounded estimate of the causal association. Plasma Hcys levels are controlled by other genetic risk markers such as MTRR and MTR which may be in strong LD with the genetic variant used in this analysis, MTHFR, and violate the assumptions of Mendelian Randomization.

Pleiotropy, a phenomenon by which a single gene can have multiple effects, can also confound or invalidate the Mendelian randomization approach only if the genetic variant is associated with pleiotropic effects that influence the outcome.

Conclusion

In the current climate of GWAS studies and whole genome sequencing, the literature-based meta-analysis has its own unique utility. Where large GWAS studies have failed to identify risk associations with gene variants such as *MTHFR* and *ACE*, literature based meta-analysis of comparable (or greater) power and sample size have been able to identify these genetic risk variants to be associated with stroke (Bevan et al., 2012, Bentley et al., 2010). Although the stand-alone validity of individual candidate gene based studies remains inconclusive, a meta-analysis may reveal a true association. Empirical comparison of meta-analyses of published gene-disease associations versus consortium analyses in cancer genetics has shown that although the two methodologies are based on significantly different data types, both approaches compliment each other and have been successful in identifying statistically significant decreased breast cancer

risks for gene *CASP8 D302H* (Janssens et al., 2009).

The current meta-analysis included studies that examined candidate genes previously identified in other vascular disorders such as MI in Europeans and South East Asians. Though this is a reasonable exploratory approach to understanding the genetics of stroke in South Asians, there is a clear need for larger GWAS strategies to truly understand the genetic underpinnings of stroke. A good starting point could be replication in an independent South Asian stroke population of gene variants identified in large statistically powered GWAS studies conducted in Europeans. The WTCCC2 and the ISGC have successfully identified novel genes associated with stroke risk (*PITX2, ZFHX3, 9p21* locus and *HDAC9*) (International Stroke Genetics et al., 2012) while other GWAS studies have highlighted the subtype specific nature of these genetic effects (Holliday et al., 2012). The recent METASTROKE (Traylor et al., 2012, Holliday et al., 2012) meta-analysis which included 15 stroke cohorts comprising of 12,000 cases and 60,000 controls validated these findings but failed to identify any new genetic risk variants. Replication of SNPs from related cardiovascular GWAS studies found one novel association with gene *PHACTR1*that suggests that detection of any new genetic risk variants will rest on proper stroke sub typing. Though the effect sizes for these genes are small, it is likely that they are true associations as compared to genes identified in smaller underpowered candidate gene based studies.

Findings from the current study support a genetic etiology of ischemic stroke in South Asians but the dataset is considerably smaller compared to those of European descent. We show no major differences in risk associations for four previously studied stroke susceptibility genes between South Asians, Europeans and South East Asians. However our study evaluated a limited number of risk variants and therefore the presence of genetic variants with differential effects by ethnicity cannot be completely ruled out. It would be fallacious to assume a literal comparison between studies that are not statistically at power and hence there is a clear need for large prospective well powered GWAS studies in South Asians, as has been done in Europeans.

2.5 Figures and Tables

Figure 2.1: Flow diagram illustrating search strategy and studies included in the meta-analysis.
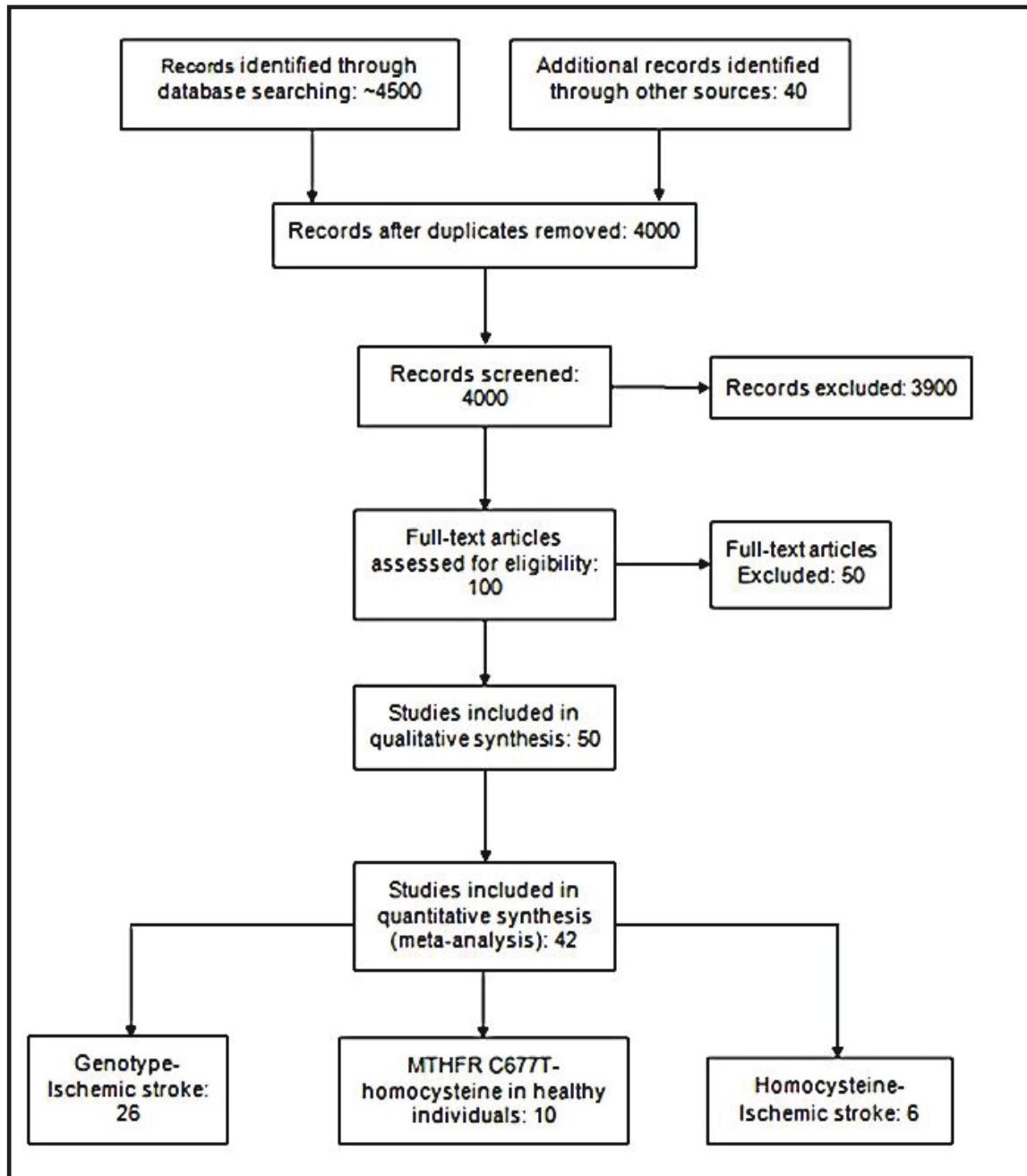
Figure 2.2: Meta-analysis forest plot and pooled ORs of risk from studies investigating *PDE4D SNP 83* polymorphism
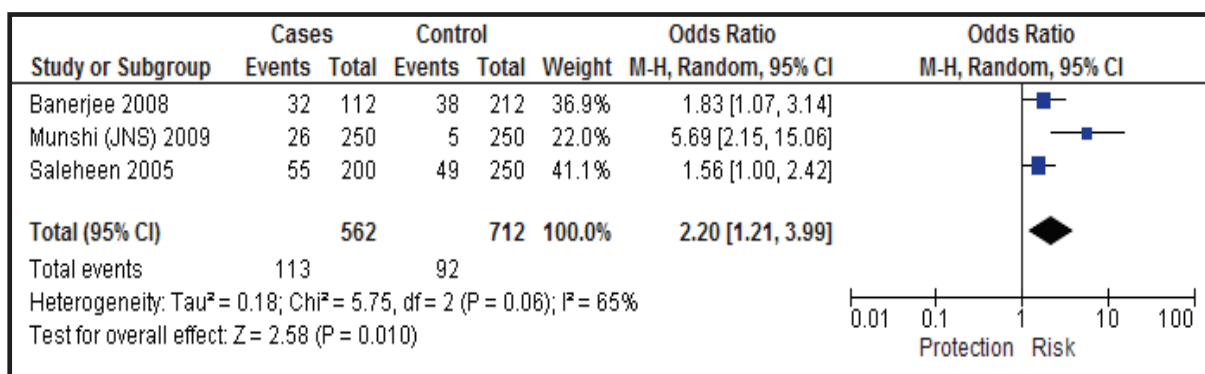
Figure 2.3: Meta-analysis forest plot and pooled ORs of risk from studies investigating *ACE I/D* polymorphism
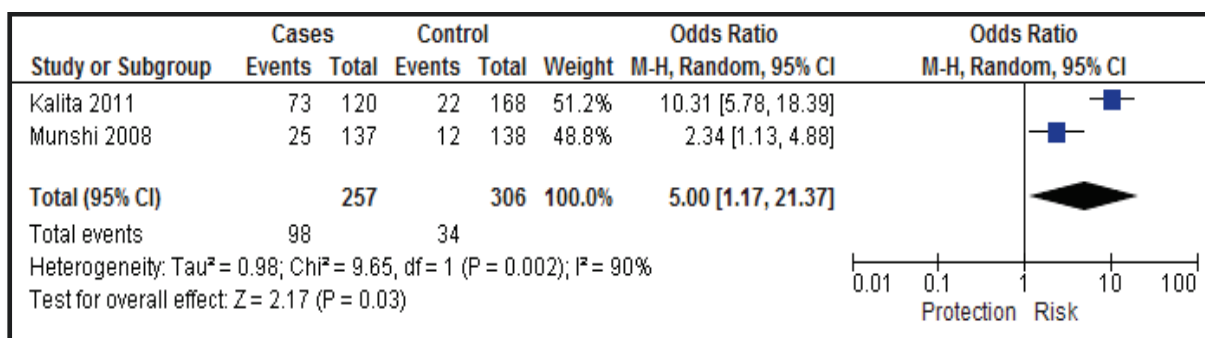
Figure 2.4: Meta-analysis forest plot and pooled ORs of risk from studies investigating *IL10 G1082A* polymorphism
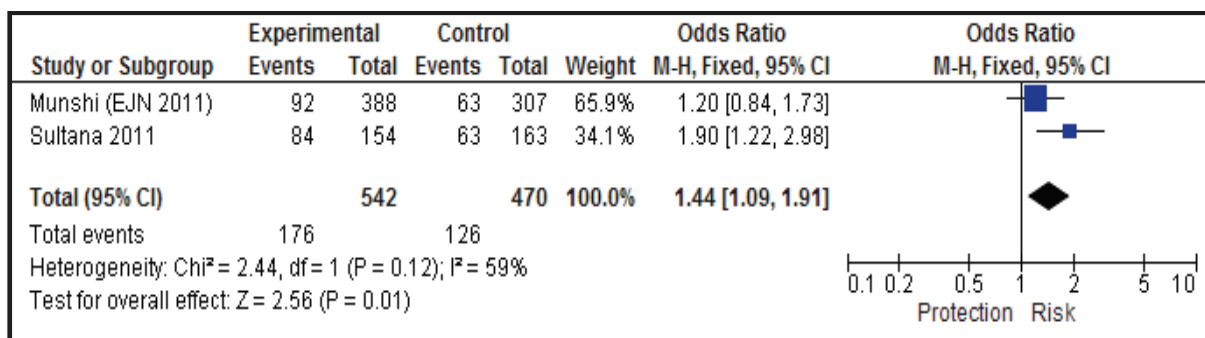
Figure 2.5: Meta-analysis, forest plots and pooled ORs of risk from studies investigating *MTHFR C677T* polymorphism (TT vs CC-stroke, Hcys-stroke and Hcys-MTHFR C677T.
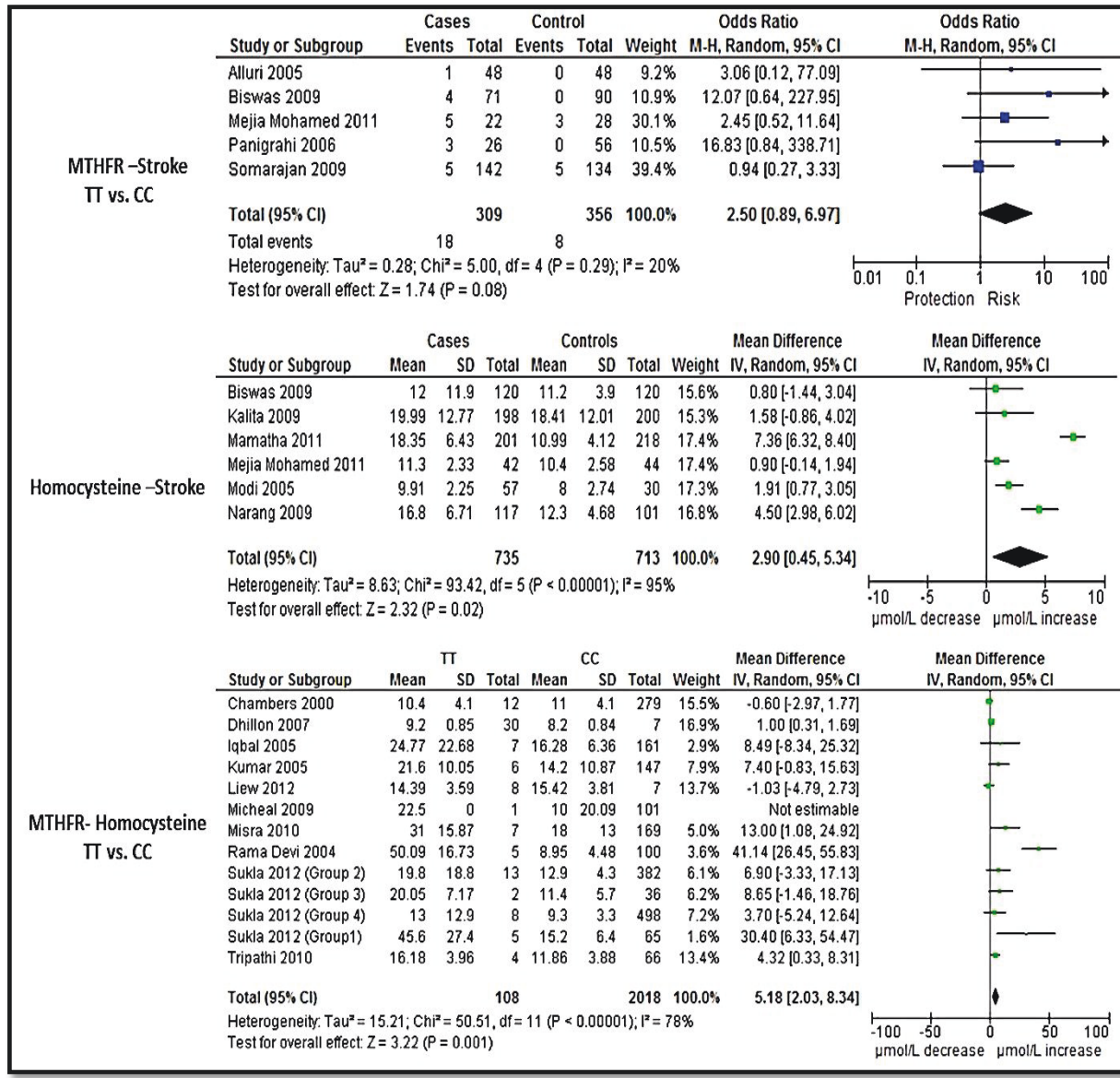
Figure 2.6: Comparison of effect sizes for risk of ischemic stroke for genes *PDE4D (SNP83), ACE (I/D), IL10 (G1082A)* and *MTHFR (C677T)*, among different ethnicities.
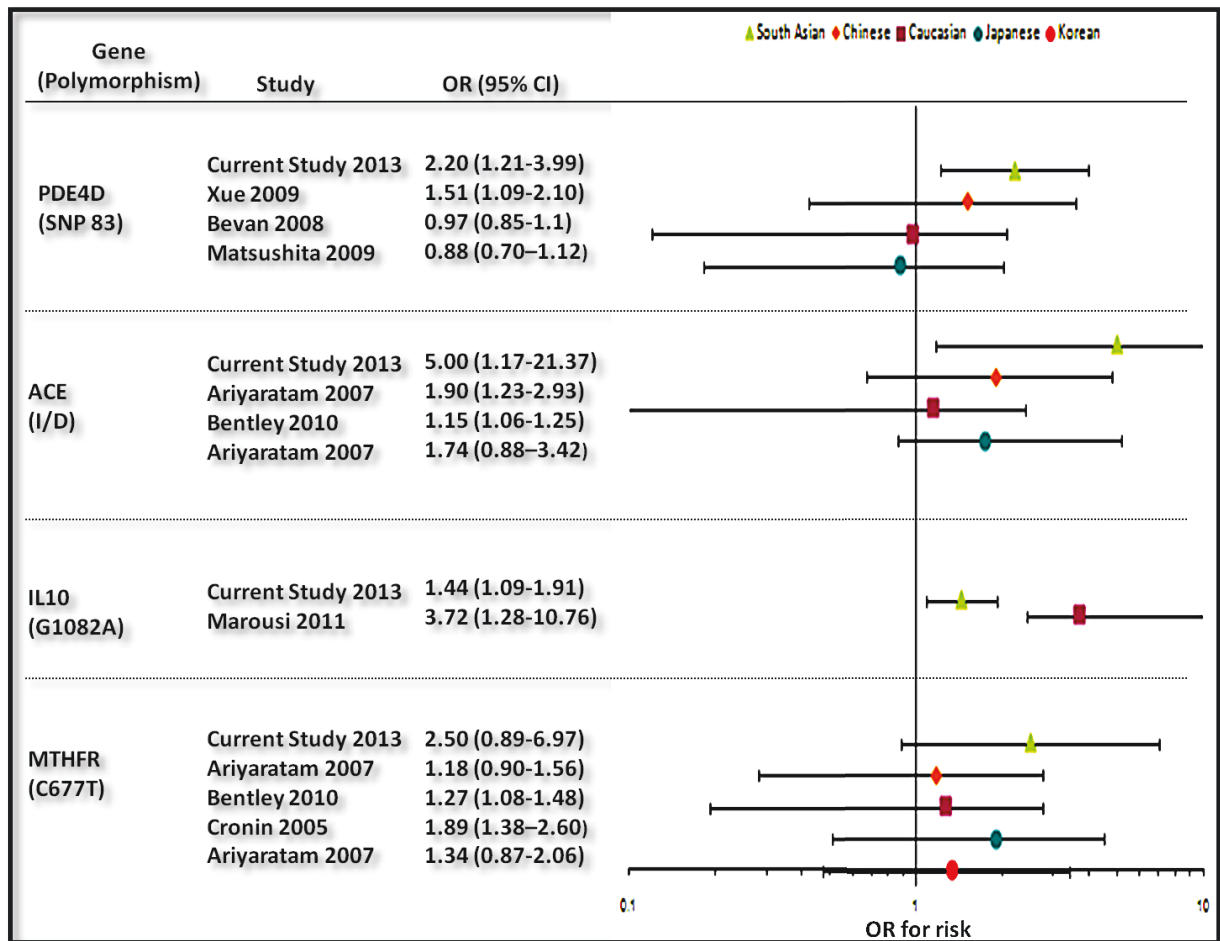
Figure 2.7: Mendelian Randomization to compare estimated risk with observed risk for gene polymorphisms associated with ischemic stroke.
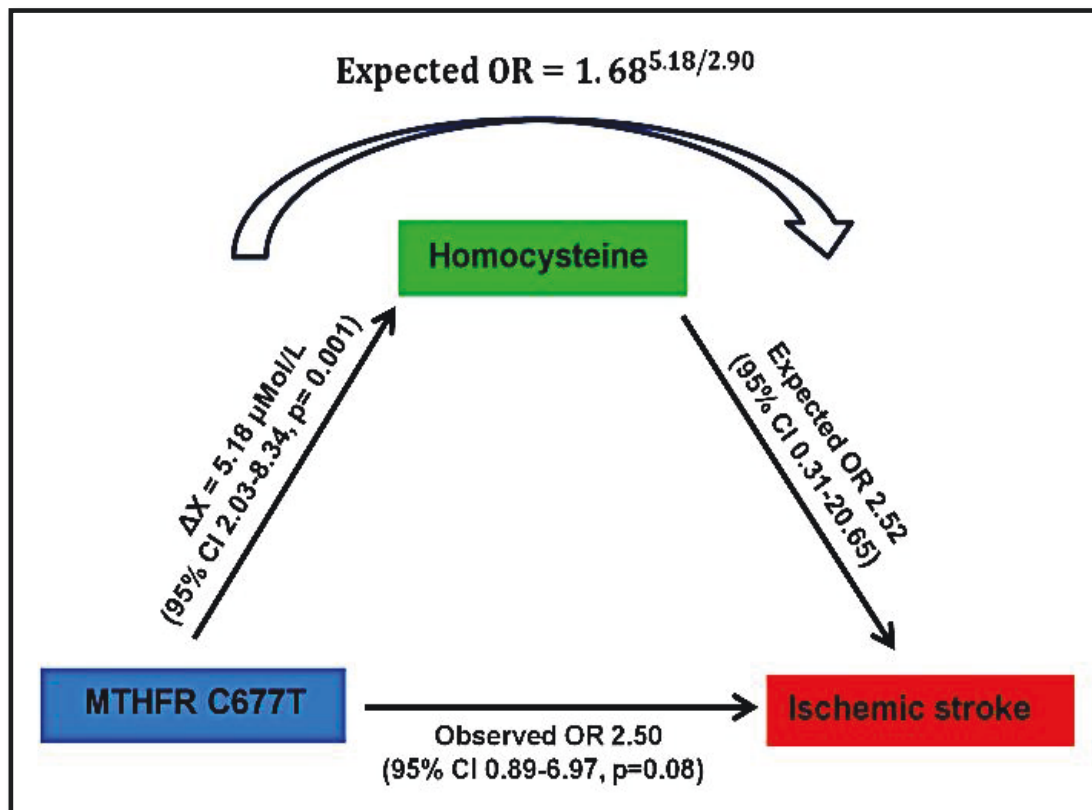
Table 2.1: Summary table of gene polymorphisms associated with risk of ischemic stroke in South Asians.

Gene	Polymorphism (R A)	Study	Cases			Control			Pooled OR (95% CI)	p
			N	Age (years)	Gender (F/M)	N	Age (years)	Gender (F/M)		
MTHFR	C677T (T)	Biswas 2009 (J of Stroke and CVD) ∞	71	< 40	NR	90	< 40	NR	2.50 (0.89-6.97)	0.08
		Alluri 2005∞	48	7-78	9/60	48	NR	NR		
		Somarajan 2011∞	142	54±15.9	55/152	134	55.25±10.65	62/126		
		Panigrahi 2006∞	26	12 (1-42)	10/32	56	NR	NR		
		Mejia Mohamed 2011⌂	22	61.0±10.1	58/92	28	60.6±7.1	71/71		
ACE	I/D (D)	Kalita 2011 (Clin. Chim. Acta) ∞	193	56.5±2.8	52/141	188	55.7±12 (F) 54.3±9.7 (M)	69/129	5.00(1.17-21.37)	0.03
		Munshi 2008 (J of Neur Sc.) ◊	162	52.83±4.04	42/120	150	54.76±1.04	39/111		
ApoE	E4/E4	Luthra 2002 (Clin Gen.)	63	56.4±13.1	NR	57	39.4±8.0	NR	2.15 (0.75-6.19)	0.16
		Chowdhury 2001 (J of Epid.)∘	147	57.9±11.1	30/117	190	60.3±9.6	61/129		

ENOS	4a/4b (aa)	Majumdar 2010 (JAT) ◊	175	29.71±10.82	54/123	214	28.42±7.87	72/147	1.87 (0.77-4.59)	0.17
		Munshi (BRB 2010) ◊	357	48.01±15.25	93/264	283	47.05±16.75	73/210		
	T786C (C)	Majumdar 2010 (JAT) ◊	129	29.71±10.82	54/123	129	28.42±7.87	72/147	1.93(0.63-5.93)	0.25
	G894T (T)	Majumdar 2010 (JAT) ◊	172	29.71±10.82	54/123	214	28.42±7.87	72/147	0.64 (0.18-2.27)	0.49
PDE4D	SNP 83	Munshi 2009 (JNS) ◊	250	48.53±16.34	62/188	250	47.01±17.78	65/185	2.20 (1.21-3.99)	0.001
		Saleheen 2005 (Stroke) ▫	200	62.4±12.4	82/118	250	54.1±8.87	87/163		
		Banerjee 2008 (Brain Res Bul.) ◊	176	58.6±14.2	63/113	212	57.4±8.8	69/143		
	SNP 87	Munshi 2009 (JNS) ◊	250	48.53±16.34	62/188	250	47.01±17.78	65/185	1.17 (0.71-1.94)	0.54
		Saleheen 2005 (Stroke) ▫	200	62.4±12.4	82/118	250	54.1±8.87	87/163		
	SNP 32	Munshi 2009 (JNS) ◊	250	48.53±16.34	62/188	250	47.01±17.78	65/185	1.57 (1.01-2.45)	0.045
		Saleheen 2005 (Stroke) ▫	200	62.4±12.4	82/118	250	54.1±8.87	87/163		
	SNP 41 (T)	Munshi (Gene 2012) ◊	516	49.3± 17.34	156/360	513	49.01±16.78	16.78	1.76 (1.25-2.48)	0.001
	SNP 56 (T)	Munshi (Gene 2012) ◊	516	49.3± 17.34	156/360	513	49.01±16.78	16.78	3.25 (2.30-4.59)	<0.0001
	SNP 59736747 T>G (G)	Munshi (Gene 2012) ◊	516	49.3± 17.34	156/360	513	49.01±16.78	16.78	1.23 (0.50-2.99)	0.66

IL 10	G1082A	Munshi 2010 (Cytokine) ◊	480	49.3±17.34	123/357	470	47.01±16.78	155/351	1.44 (1.09-1.91)	0.01
		Sultana 2011 ◊	238	53.72±11.11	74/164	226	54.06±10.98	105/121		
PAI-1	4G/5G	Babu 2012 (Gene) ◊	516	49.3	156/360	513	49.01	155/358	1.33 (0.83-2.15)	0.24
CBS	T833C/844ins68 (C)*	Chandra 2006 (Neur. India)	30	NR	NR	138	NR	NR	0.40 (0.049-3.2)	0.39
IL-1 α	C889T (T)	Banerjee (BRB 2008) ∞	176	58.6±14.2	63/113	212	57.4±8.8	69/143	0.53 (0.33-0.86)	0.009
Factor XIIIB	V34L	Kain 2005	143	63±34	40/40	146	61±35	42/38	3.16 (0.32-31.08)	0.32
α1 ACT	Ala15Thr	Somarajan 2009∞	272	53±16.7	75/197	188	54.8±10	65/	0.68 (0.40-1.17)	0.17
ADD1	G/W460	Kalita (CCA 2011)	193	56.5 (2-83)	52/141	188	54.3±9.7 (M) 55.7±12 (F)	69/129	1.38 (0.43-4.42)	0.59
CYP11 B2	C344T	Munshi (JNS 2010) ◊	403	49.3±17.34	116/287	394	47.01±16.78	115/279	0.43 (0.29-0.63)	0.0001
ESR	Pvu II	Munshi (CCA 2011) ◊	400	49.3±17.34	115/285	380	47.01±16.78	102/278	1.92 (1.24-2.99)	0.004
	Xba I	Munshi (CCA 2011) ◊	400	49.3±17.34	115/285	380	47.01±16.78	102/278	1.59 (0.92-2.75)	0.09
Klotho	KL-VS	Majumdar 2010 ◊	460	41.76±16.34	258/202	574	40.36±11.42	321/253	1.26 (0.56-2.84)	0.57
	C1818T	Majumdar 2010 ◊	460	41.76±16.34	258/202	574	40.36±11.42	321/253	0.81 (0.48-1.38)	0.44

α-ADD1	WG	Kalita (CCA 2011)∞	193	56.5±2-83	52/141	188	55.7±12 (F) 54.3±9.7 (M)	69/129	1.37 (0.43-4.39)	0.60
TNF α	G308A	Sultana 2011 ◊	238	53.72±11.11	74/164	226	54.06±10.98	105/121	1.39 (0.52-3.72)	0.5
	G488A	Munshi (EJN 2011) ◊	525	49.3	NR	500	47.01	NR	1.91 (1.32-2.76)	0.0006
MMP-3	5A/6A	Munshi (EJN 2011) ◊	525	49.3	NR	500	47.01	NR	0.87 (0.58-1.29)	0.9
CYP4F2	G1347A	Munshi (M B R 2012) ◊	507	49.3±17.34	144/363	487	49.01±16.78	131/356	1.58 (1.16-2.15)	0.004
MDR 1	C3435T	Sharma (Neur. Sciences 2011) ◊	560	49.3±17.34	167/393	560	49.01±16.78	173/387	2.66 (1.98-3.55)	<0.0001
LPL	Hind III	Munshi (JNS 2012) ◊	525	49.3±17.34	151/374	500	49.01±16.78	143/357	0.65 (0.43-0.97)	0.03
tPA	C7351T	Babu (Gene 2012) ◊	516	49.3±17.34	156/360	513	49.01±16.78	155/358	1.14 (0.74-1.77)	0.56
	I/D	Babu (Gene 2012) ◊	516	49.3±17.34	156/360	513	49.01±16.78	155/358	1.74 (1.16-2.62)	0.008

*All comparisons are based on a recessive genetic model i.e. homozygous mutant alleles (MM) vs. total alleles (W+) where W and M denote wild type and mutant alleles respectively and + can denote either W or M.

** ∞ Studies from North India, ⌂Studies from Malaysia, ◊ Studies from South India, ∘ Studies from Bangladesh, ▯ Studies from Pakistan

CHAPTER 3: GENOME WIDE ANALYSIS OF BLOOD PRESSURE VARIABILITY AND ISCHEMIC STROKE

Abstract

Visit-to-visit variability in BP is associated with ischemic stroke and is likely to be a heritable trait as are other measures of elevated blood pressure (SBP, DBP, MAP, and PP).

The study sought to determine whether genetic variants associated with long term BP variability are also associated with ischemic stroke. A GWAS for loci influencing BP variability was undertaken in 3,802 individuals from the multinational ASCOT study where long-term visit-to-visit and within visit BP measures were available. Since BP variability is strongly associated with ischemic stroke, the sentinel SNP from the BP analysis was tested for association in an independent ischemic stroke population comprising of 8,624 cases and 12,722 controls from 7 cohorts.

The ASCOT discovery GWAS identified a cluster of 17 correlated SNPs within the *NLGN1* gene (3q26.31) associated with BP variability. The strongest association, reaching genome wide significance, was with rs976683 (p=1.4x10^{-8}). Conditional analysis on rs976683 provided no evidence of additional independent associations at the locus. Analysis of rs976683 in a large group of patients with ischemic stroke found no association for overall stroke (OR 1.02; 95% CI 0.97-1.07; p=0.52) or its sub-types: cardio embolic stroke (OR 1.07; 95% CI 0.97-1.16; p=0.17), large vessel disease (OR 0.98; 95% 0.89-1.07; p=0.60) and small vessel disease (OR 1.07; 95% CI 0.97-1.17; p=0.19).

Findings from the study identified a cluster of SNPs within the *NLGN1* gene showing significant association with BP variability. However, these SNPs were not associated with an increased risk of ischemic stroke or its subtypes.

3.1 Introduction

Hypertension is the largest modifiable risk factor for stroke resulting in death and disability (WHO, 2013). Association of elevated BP is a strong predictor of ischemic stroke and small changes in BP can have serious effects on the prognosis of stroke. In 2004, the WHO attributed approximately 7.4 million deaths, 12.5% of the global death burden, to high blood pressure.

Familial studies have long provided estimates of heritability ranging from 31%-34% (single measure SBP and DBP averaged over 3 studies) to 56%-57% (long-term SBP and DBP average) to 63%-68% (24 hour SBP and DBP profile)(Ehret et al., 2011). However, the genetics of blood pressure remained largely unexplained until the year 2009, when two big GWAS studies (Newton-Cheh et al., 2009, Levy et al., 2009) involving around ~40,000 individuals identified 13 gene loci to be significantly associated with blood pressure. Cross-population GWAS's in South-east Asians (Kato et al., 2011) and large global consortia (Ehret et al., 2011) soon followed and identified an additional 21 gene loci. Genome wide association studies have since identified several genetic loci to be significantly associated with blood pressure traits such as SBP, DBP, MAP, PP and presence or absence of HTN (Johnson et al., 2011).

However, episodic HTN or variability in BP remains understudied despite evidence supporting their role as risk factors in vascular events (Rothwell et al., 2010b). Studies as early as 1992 (Ekbom et al., 1992) demonstrated the effect of antihypertensive drugs on decreasing blood pressure variability for the lowered risk of stroke. Other studies (Frattola et al., 1993, Mancia et al., 2007, Kikuya et al., 2000) also provided strong evidence supporting the association of blood pressure variability and cardiovascular risk. However, these early studies were limited by the fact that they measured only short-term blood pressure variability using 24 hour ambulatory monitoring or home blood-pressure recordings. Recent work by Peter Rothwell and colleagues has shown that visit-to-visit variability in systolic BP is a strong predictor of ischemic stroke independent of mean BP (Rothwell et al., 2010), with hypertensives showing the most BP variability over a series of visits having the greatest risk of a cardiovascular event (Rothwell et al., 2010).

Further support was provided in a meta-analysis published on the effects of β-blocker selectivity on blood pressure variability and stroke (Rothwell et al., 2010), where authors concluded that nonselective β-blockers increase variability in blood pressure which may explain their non-effectiveness in preventing stroke and MI (Webb et al., 2011). Since most of the studies on blood pressure variability have been conducted on stroke cohorts, it is difficult to rule out pre-existing ischemia. Cerebral auto regulation is an important protective mechanism of the brain against ischemic changes and regulation of blood pressure maybe associated with it (Jordan and Powers, 2012).

Determining whether BP variability has a genetic basis is difficult given the lack of prospective cohorts with visit-to-visit BPs recorded and accompanying GWAS data. The ASCOT study is a longitudinal study investigating the impact of a calcium channel blocker against a beta-blocker regime in hypertensive individuals at moderate risk of cardiovascular outcomes, recruited in the United Kingdom, Ireland and Nordic countries from 1998-2000 (Sever et al., 2001). Unusually, long-term BP variability measurements and genotyped data were available for the ASCOT cohort allowing a genome wide analysis to be conducted on the genetic risk variants of BP variability. A cluster of 17 correlated SNPs near the *NLGN1* gene locus (Chr 3) were found to be associated with blood pressure variability at genome wide significance (p=1.7x10^{-8}) in the ASCOT Anglo IR-UK cohort (N=3764).

Based on the ASCOT GWAS results, there was a clear need for independent replication and further data analysis in another sample population. The study hypothesized that since visit-to-visit BP variability is associated with risk of ischemic stroke more than hemorrhagic stroke (Rothwell et al., 2010) and HTN is a major modifiable risk factor, any genetic variants associated with BP variability may also be associated with ischemic stroke. Based on recently published GWAS studies (Traylor et al., 2012, Holliday et al., 2012), which show the genetic risk of stroke to be subtype specific, the genetic variant in stroke subtypes were also tested.

3.2 Methods

3.2.1 Study Populations

3.2.1.1 ASCOT

The ASCOT study was conducted by the ASCOT group. The methodology and results are reproduced with permission from Dr. Patricia Munroe.

The ASCOT Blood Pressure Lowering Arm (ASCOT-BPLA) is an investigator-led multi-center trial, which included over 19,000 hypertensive patients, aged 40-79 years at baseline, with an average SBP of 140/90-mmHg on-treatment and 160/100-mmHg off-treatment. Patients had no history of CHD but had at least three other risk factors for cardiovascular disease such as LVH, T2D, peripheral artery disease, previous stroke/TIA, male, ≥ 55 years of age or cigarette smoking. The study tested the impact of a contemporary calcium channel blocker based regimen against an older beta-blocker based regime in hypertensives at moderate risk of a cardiovascular outcome. The primary objective of the blood pressure-lowering arm (BPLA) was to assess and compare the long-term effects of two blood-pressure-lowering regimens on the combined endpoint of non-fatal MI (including silent MI) and fatal CHD. Blood pressure was measured in a seated position by a uniform automated device (Omron HEM705CP) in all participants over an average of 13 visits across 5.5 years.

The ASCOT GWA study population included a subset of 3802 individuals extracted from the original cohort of 19,342 hypertensives. Visit-to-visit BP variability measurements were recorded prospectively for within visit and between visit BP variability over 5.5 years. Blood samples for DNA isolation were collected of which 3,802 individuals of European ancestry from UK and Ireland were genotyped allowing a genome wide analysis to be conducted on the risk variants of BP variability.

3.2.1.2 Ischemic Stroke

The stroke population included 8,624 cases and 12,722 controls from 7 different cohorts: Australian Stroke Genetics Collaborative (ASGC) (McEvoy et al., 2010, Holliday et al., 2012), Bio-Repository of DNA in Stroke (BRAINS) (Yadav et al., 2011, Cotlarciuc I, 2012), Genetics of Early Onset Stroke (GEOS) (MacClellan et al., 2006, Kittner et al., 1998), Ischemic Stroke Genetics Study and Siblings with Ischemic Stroke Study (ISGS)

(Meschia et al., 2003)/SWISS (Meschia et al., 2006)), Welcome Trust Case Control Consortium 2 United Kingdom (WTCCC2-UK) (International Stroke Genetics et al., 2012), Welcome Trust Case Control Consortium 2 Germany (WTCCC2-Germany) (International Stroke Genetics et al., 2012) and Vitamin Intervention for Stroke Prevention trial (VISP) (Spence et al., 2001).

All participating cohorts received institutional ethical clearance and signed consent from each participating study subject. ISGS/SWISS, GEOS and VISP used gender and age matched stroke-free controls recruited from the local population. BRAINS and WTCCC2-UK used the WTCCC 1958 British Birth cohort and National Blood Service (NBS) controls. WTCCC2-Germany derived controls of German Caucasian origin from the KORAgen study (www.gsf.de/kora).

TOAST classification (Adams et al., 1993) was performed by an in-house neurologist and all stroke cases were classified into 3 categories: cardio embolic stroke, large artery disease and small vessel disease. All cohorts except VISP provided stroke subtype data.

Details of stroke cohort study populations are as follows:

ASGC: ASGC stroke _cases_ comprised stroke patients of European ancestry who were admitted to four clinical centers across Australia (The Neurosciences Department at Gosford Hospital, Gosford; the Neurology Department at John Hunter Hospital, Newcastle; The Queen Elizabeth Hospital, Adelaide; and the Royal Perth Hospital, Perth) between 2003 and 2008(Holliday et al., 2012b). Stroke was defined by World Health Organization criteria as a sudden focal neurological deficit of vascular origin, lasting more than 24 h and confirmed by imaging, such as CT and/or MRI brain scan. Other investigative tests such as ECG, carotid Doppler and trans-esophageal echocardiogram were conducted to define ischemic stroke mechanism as clinically appropriate. Cases were excluded from participation if they were aged <18 years were diagnosed with hemorrhagic stroke or had transient ischemic attack rather than ischemic stroke or if they were unable to undergo baseline brain imaging. On the basis of these criteria, a total of 1,230 ischemic stroke cases were included in the current study. Ischemic stroke subtypes were assigned using TOAST criteria on the basis of clinical, imaging and risk factor data. _ASGC controls_ were participants in the Hunter Community Study (HCS), a population-based cohort of individuals aged 55–85 years, predominantly of European

ancestry and residing in the Hunter Region in New South Wales, Australia. Detailed recruitment methods for the HCS have been previously described. Briefly, participants were randomly selected from the New South Wales State electoral roll and were contacted by mail between 2004 and 2007. Consenting participants completed five detailed self-report questionnaires and attended the HCS data collection center, at which time a series of clinical measures were obtained. A total of 1,280 HCS participants were genotyped for the current study. All study participants gave informed consent for participation in genetic studies. Approval for the individual studies was obtained from the relevant institutional ethics committees.

BRAINS (http://www.brainsgenetics.com)is an ongoing, multicenter, in-hospital study, which recruits consenting acute stroke patients into a highly characterized bio bank (Yadav et al., 2011, Cotlarciuc I, 2012). All adult (>18 years of age) *stroke cases* were recruited with either ischemic or hemorrhagic pathology MRI confirmed lesions. Ischemic stroke subtypes were further sub-classified according to TOAST criteria (Adams et al., 1993). Known monogenic causes of stroke were excluded. BRAINS has two principal arms. The first arm recruits UK European stroke patients while the second arm recruits South Asian stroke patients from multiple sites in the UK and also from sites in India. *Control data* for the European arm was provided by the WTCCC 2 while control subjects for the South Asian arm are recruited simultaneously as the affected stroke patient and usually is the proband's spouse.

GEOS is a population-based case-control study designed to identify genes associated with early-onset ischemic stroke and to characterize interactions of identified stroke genes and/or SNPs with environmental risk factors (Cheng et al., 2011). Participants were recruited from the greater Baltimore-Washington area in 4 different time periods: Stroke Prevention in Young Women-1 (SPYW-1) conducted from 1992-1996, Stroke Prevention in Young Women-2 (SPYW-2) conducted from 2001-2003, Stroke Prevention in Young Men (SPYM) conducted from 2003-2007, and Stroke Prevention in Young Adults (SPYA) conducted in 2008. *Case participants* were hospitalized with a first cerebral infarction identified by discharge surveillance from one of the 59 hospitals in the greater Baltimore-Washington area and direct referral from regional neurologists. The abstracted hospital records of cases were reviewed and adjudicated for ischemic stroke subtype by a pair of neurologists according to previously published

procedures with disagreements resolved by a third neurologist. The ischemic stroke subtype classification system retains information on all probable and possible causes, and is reducible to the more widely used TOAST system that assigns each case to a single category. *Control participants* without a history of stroke were identified by random-digit dialing and were balanced to cases by age and region of residence in each recruitment period. Genomic DNA was isolated from a variety of sample types, including cell line, whole blood, mouthwash and buccal swab. Samples were genotyped at the Johns Hopkins Centre for Inherited Disease Research (CIDR) using the Illumina HumanOmni1-Quad_v1-0_B BeadChip (Illumina, San Diego, CA, USA). Individuals were excluded if they were unexpected duplicates, gender discrepancy and unexpected relatedness.

ISGS/SWISS: ISGS is a multicenter inception cohort study of first-ever ischemic stroke in adult men and women (Meschia et al., 2003). *Cases* were recruited from inpatient stroke services at five academic medical centers in Florida, Georgia, Virginia and Minnesota. A study neurologist on the basis of medical history, physical examination and CT or MR imaging of the brain confirmed the diagnosis of ischemic stroke. Cases had to be enrolled within 30 days of onset of stroke symptoms. Cases were excluded if they had a mechanical aortic or mitral valve, central nervous system vasculitis, or bacterial endocarditis at the time of the stroke. They were also excluded if they were known to have: CADASIL, Fabry disease, homocystinuria, MELAS, or sickle cell anemia. Stroke severity at enrollment was assessed using the NIH Stroke Scale (NIHSS) and outcomes at 90-days were assessed by telephone using the Barthel Index, Glasgow Outcome Scale, and the modified Rankin scale (Kasner, 2006). Diagnostic evaluation included: head CT (95% of individuals enrolled) or MRI (83%), electrocardiography (92%), cervical arterial imaging (86%), and echocardiography (74%). A vascular neurology committee reviewed the medical records of every case and assigned ischemic stroke subtype diagnoses according to criteria from the Trial of ORG10172 (TOAST) (Adams et al., 1993), the OCSP (Bamford et al., 1991), and the Baltimore-Washington Young Stroke Study (Johnson et al., 1995). DNA was donated to the NINDS DNA Repository (Coriell Institute, Camden, NJ) for eligible samples with appropriate written informed consent. A separate certified neurologist adjudicator additionally assigned a

subtype diagnosis using the standardized Causative Classification of Stroke web-based algorithm (Ay et al., 2007).

SWISS is a multicenter affected sibling pair study(Meschia et al., 2011b). *Probands* with ischemic stroke were enrolled at 66 US medical centers and 4 Canadian medical centers. Probands are adult men and women over the age of 18 years diagnosed with ischemic stroke confirmed by a study neurologist on the basis of history, physical examination and CT or MR imaging of the brain. Probands were required to have a history of at least one living sibling with a history of stroke. Probands were excluded if they had a mechanical aortic or mitral valve, central nervous system vasculitis, or bacterial endocarditis at the time of the index ischemic stroke. Probands were also excluded if they were known to have CADASIL, Fabry disease, homocystinuria, MELAS, or sickle cell anemia. Siblings were enrolled using proband-initiated contact (Worrall et al., 2001) or direct contact when permitted by Institutional Review Boards. Concordant (affected) siblings had their diagnosis of ischemic stroke confirmed by review of medical records by a vascular neurology committee. Concordant siblings had the same eligibility criteria as probands. Subtype diagnoses were assigned to the index strokes of probands and concordant siblings according to TOAST criteria (Adams et al., 1993). Discordant siblings of the proband were confirmed to be stroke-free using the Questionnaire for Verifying Stroke-free Status (Meschia et al., 2000). Lymphoblastoid cell lines were created on all subjects. A certified neurologist adjudicator additionally assigned a subtype diagnosis using the standardized CCS web-based algorithm to all concordant siblings and a subset of probands for whom medical records were available (Ay et al., 2007).

VISP: The VISP trial (P.I. James Toole, MD, Wake Forest University School of Medicine (WFU); R01 NS34447) was a multicenter, double-blind, randomized, controlled clinical trial that enrolled patients aged 35 or older with Hcys levels above the 25th percentile at screening and a non-disabling cerebral infarction (NDCI) within 120 days of randomization. NDCI was defined as an ischemic brain infarction not due to embolism from a cardiac source, characterized by the sudden onset of a neurological deficit. The deficit must have persisted for at least 24 hours, or if not, an infarction in the part of the brain corresponding to the symptoms must have been demonstrated by CT or MRI imaging. The trial was designed to determine if daily intake of a multivitamin tablet with

high doses folic acid, vitamin B6 and vitamin B12 reduced recurrent cerebral infarction (1° endpoint), and nonfatal MI or mortality (2° endpoints). Subjects were randomly assigned to receive daily doses of the high-dose formulation (n=1,827), containing 25mg pyridoxine (B6), 0.4mg cobalamin (B12), and 2.5mg folic acid; or the low-dose formulation (n=1,853), containing 200µg pyridoxine, 6µg cobalamin and 20µg folic acid. Enrolment in VISP began in August 1997, and was completed in December 2001, with 3,680 participants enrolled, from 55 clinic sites across the US and Canada and one site in Scotland.

Subsets of VISP participants gave consent and were included in the GWAS component of VISP, supported by the National Human Genome Research Institute (NHGRI), Grant U01 HG005160, as part of the Genomics and Randomized Trials Network (GARNET). Samples were genotyped at the Johns Hopkins Center for Inherited Disease Research (CIDR), and genotyping was performed using the Illumina HumanOmni1-Quad_v1-0_B BeadChip (Illumina, San Diego, CA, USA). Individuals were excluded if they were unexpected duplicates or had gender discrepancies. All VISP participants are stroke cases, therefore we obtained GWAS data (dbGAP) for 1047 external controls from the High Density SNP Association Analysis of Melanoma: Case-Control and Outcomes Investigation (Study Accession: phs000187.v1.p1). These samples were also genotyped on the Illumina HumanOmni1-Quad.

WTCCC2-United Kingdom and WTCCC2-Germany: The WTCCC2 samples were genotyped as part of the WTCCC 2 ischemic stroke study (International Stroke Genetics et al., 2012). *Stroke cases* included samples recruited by investigators at St. George's University London (SGUL), University of Oxford and Edinburgh Stroke Study in the UK and the Department of Neurology, KlinikumGroßhadern, Ludwig-Maximilians-University, Munich. The SGUL collection comprised 1224 ischemic stroke samples from a hospital based setting. All cases were of self-reported Caucasian ancestry. Ischemic stroke subtypes were determined according to TOAST criteria based on relevant clinical imaging and available information on cardiovascular risk factors. The University of Oxford collection comprised 896 ischemic stroke cases, consecutively collected as part of the Oxford vascular study (OXVASC). Cases were of self-reported Caucasian ancestry, and ischemic stroke subtypes were determined according to TOAST criteria based on relevant clinical imaging. For the Edinburgh Stroke Study, consecutive consenting

patients with stroke who were admitted to or seen as outpatients at the Western General Hospital, Edinburgh were prospectively recruited between 2002 and 2005. Cases in this study were those with a clinically evident stroke, demonstrated by brain imaging (CT or MRI) to be ischemic. An experienced stroke physician assessed each patient as soon as possible after the stroke, prospectively recording demographic and clinical details, including vascular risk factors and results of brain imaging and other investigations. The Munich samples included 1383 ischemic stroke cases. Cases were consecutive European Caucasians recruited from a single dedicated Stroke Unit at the Department of Neurology, KlinikumGroßhadern, Ludwig-Maximilians-University, Munich. Ischemic stroke subtypes were determined according to TOAST criteria based on relevant clinical and imaging data. _Controls_ for the UK samples were drawn from shared WTCCC controls obtained from the 1958 Birth Cohort. This is a prospectively collected cohort of individuals born in 1958 (www.b58cgene.sgul.ac.uk/) and ascertained as part of the national child development study (www.cls.ioe.ac.uk/studies.asp). Data from this cohort are available as a common control set for a number of genetic and epidemiological studies. For the German samples controls were Caucasians of German origin participating into the population KORAgen study (www.gsf.de/kora). This survey represents a gender- and age stratified random sample of all German residents of the Augsburg area and consists of individuals 25 to 74 years of age, with about 300 subjects for each 10-year increment. All controls were free of a history of stroke or transient ischemic attack.

3.2.2 Genotyping, Imputation and Quality control

ASCOT

Genotyping for the ASCOT samples was performed using the Illumina Human CNV370 Bead Array. For the SNPs that were not directly genotyped, genotypes were obtained through imputation. Quality control and imputation of the ASCOT data have been described previously (Deshmukh et al., 2012). After stringent quality control and genotype imputation, a total of ~2.5 million SNPs and 3,802 individuals were tested for association.

Ischemic Stroke

For the stroke meta-analysis site-specific quality control was performed in PLINK to remove individuals failing the following filters: (1) Call rate ≤ 95%, (2) Non-European ancestry, (3) Outlying autosomal heterozygosity, and (4) Cryptic relatedness (pi-hat ≥ 0.2). Quality control also removed SNPs failing the following filters: (1) Call frequency ≤ 95%, (2) MAF ≤ 0.01 and (3) HWE ≥ 10^{-6}. Post imputation, SNPs with imputation $r^2<0.3$ or MAF ≤ 0.01 were removed.

A detailed description of genotyping, imputation and quality control methods for each participating study in the ischemic stroke analysis is given in **Table 3.1**.

3.2.3 Data Analysis

ASCOT

In the ASCOT study BP was measured in all participants over an average of all 13 visits across 5.5 years. Measurements for the first 6 months after starting therapy were excluded because this was a period of forced medication titration and any differential medication effects could have acted as a confounder. Data simulations demonstrated that the combination of within visit BP variability and visit-to-visit BP variability allowed the use of more BP measurements. Within-individual visit-to-visit BP variability phenotype was expressed as mean (±SD) and coefficient of variation (SD/mean) using the 2nd and 3rd readings for every visit for ASCOT-BPLA cohort. The Variance Independent of Mean (VIM) transformation was applied if there was a correlation between the mean SBP and coefficient of variation (Rothwell et al., 2010). The SBP-VIM was derived for all on-treatment SBP values, analyzing total variability (within visit and between visit variability) using a coefficient of variation (SD/mean^k) where k was determined from curve fitting (Rothwell et al., 2010). Analysis also included use of Residual Standard Deviation (RSD) for effect size estimates which is the square root of the total squared deviation of data points from a linear regression of blood pressure values against time, divided by (n–2), where n is the number of readings (Rothwell et al., 2010). All analyses were adjusted for age, gender, gender*age, SBP mean, and the first 10 principal components (from decomposition of the genotype matrix).

Ischemic Stroke

For the stroke meta-analysis, the candidate SNPs were extracted from the genome wide data and site-specific logistic regression analysis was performed to test association of top SNP with overall ischemic stroke and its major sub-types (large artery disease, cardio embolic stroke and small vessel disease) under an additive genetic model. Age and gender were used as covariates. Log (OR), standard errors and p values from different studies were pooled via inverse variance meta-analysis using a fixed effects model. Meta-analysis was carried out for over-all ischemic stroke and its sub-types based on the TOAST criteria (Adams et al., 1993) using METAL software (Willer et al., 2010). Pooled ORs were calculated using estimated effect size of the SNP and standard error of the effect size estimate. 95% confidence intervals were calculated using odd ratios and standard error. A detailed description of the statistical analysis methods for each participating study is given in **Table 3.2**.

Power for the stroke meta-analysis was calculated using the CATS genetic power calculator (Skol et al., 2006). The following parameters were used to calculate the power for the replication of SNPs rs976683 in the ischemic stroke population using an additive model: N (cases): 8624, N (controls): 12,722, stroke prevalence: 7.2% (Lee et al., 2011), rs976683 MAF: 0.25 and significance level: 0.05. The sample size provided sufficient power to detect modest effect sizes ranging from 1.1-1.4 for overall ischemic stroke but had reduced power for subtypes.

3.3 Results

3.3.1 ASCOT GWAS

The ASCOT study population consisting of 3802 subjects was primarily male (82.3%) with a mean age of 63.7 (±8.1) years. Mean SBP and mean VIM was 138.2 (±9.8) mm Hg and 0.004 (±0.001) mm Hg respectively (**Table 3.3**).

GWAS for BP variability identified a cluster of 17 correlated SNPs within the *Neuroligin-1* (*NLGN1*) gene on 3q26.31) (ENCODE ID: ENSG00000169760.13) (**Figure 3.1** and **Table 3.4**). Within the cluster, 12 SNPs were directly genotyped and 5 were imputed. Seven SNPs (3 imputed and 4 genotyped) reached genome wide significance (p≤5 x 10^{-8}) with the strongest association at the imputed SNP rs976683 (p=1.4 x 10^{-8}) (**Figure 3.2A and 3.2B**). The effect size for SNP rs976683 association was extremely small (β=0.000179) corresponding to a 0.01% mm Hg change in BP variability per copy of the risk allele. Conditional analysis, conditioned on rs976683, provided no evidence of another independent signal at this locus.

The top genotyped SNP to reach genome wide significance (p=1.72x10^{-8}) was rs9830510 (**Figure 3.2C and 3.2D**). The direction of effect was in concordance with rs976683 however the SNPs were not highly correlated (LD r^2 ~0.5).

3.3.2 Ischemic Stroke Analysis

3.3.2.1 Population Demographics

8624 cases and 12722 controls of European descent from 7 studies spread across Europe, America and Australia: ASGC, BRAINS (European arm), GEOS, ISGS/SWISS, VISP, WTCCC2-UK and WTCCC2-Germany were available. The mean age of study participants ranged from 41.0 ± 7.0 years to 72.87 ± 13.16 years for stroke cases and 39.5 ± 6.7 to 66.28 ± 7.54 for controls. The male: female ratio was approximately 50:50. The three main ischemic stroke subtypes; cardio embolic, large vessel disease and small vessel disease accounted for 1523, 1639 and 1254 cases, respectively. The demographic data such as age, gender distribution and stroke sub-type frequencies for each population are summarized in **Table 3.5**.

3.3.2.2 Association with Overall Ischemic Stroke and Sub-Types

SNP rs976683 was directly genotyped in all 7 cohorts with an average MAF of 0.26 (**Table 3.6**) and was not significantly associated (at $p \leq 0.05$) with the increased risk of ischemic stroke or its subtypes. Pooled odds ratios were as follows; overall-ischemic stroke (OR 1.02; 95% CI 0.97-1.07; p=0.52), cardio embolic (OR 1.07; 95% CI 0.97-1.16; p=0.17), large vessel disease (OR 0.98; 95% 0.89-1.07; p=0.60) and small vessel disease (OR 1.07; 95% CI 0.97-1.17; p=0.19). There was no significant heterogeneity between studies (**Table 3.7**).

Genotyped SNP rs9830510 was also tested for association in the ischemic stroke cohort to ensure that the association result of imputed SNP rs976683 was not an imputation artifact. SNP rs9830510 was directly genotyped in all 7 cohorts with an average MAF of 0.15 (**Table 3.6**). Association with increased risk of ischemic stroke or its subtypes was not significant (at $p \leq 0.05$) with pooled odds ratios as follows; overall ischemic stroke (OR 0.96; 95% CI 0.90-1.02; p=0.54), cardio embolic (OR 1.03; 95% CI 0.91-1.15; p=0.83), LVD (OR 0.76; 95% CI 0.66-0.80; p=0.03) and small vessel disease (OR 1.01; 95% CI 0.89-1.14; p=0.92). There was no significant heterogeneity between studies (**Table 3.7**).

3.4 Discussion

The current study provides evidence supporting the role of the *Neuroligin-1* (*NLGN1*) gene in BP variability but was unable to demonstrate any association between this locus and ischemic stroke or any of its subtypes. A GWAS for BP variability in the Anglo IR-UK discovery cohort identified a cluster of 17 correlated SNPs within the *NLGN1* gene, which encode a neuronal cell surface protein implicated in the growth and remodeling of the vascular system (Bottos et al., 2009). The strongest association reaching genome wide significance was at imputed SNP rs976683 ($p=1.4x10^{-8}$) and correlated genotyped SNP rs9830510 ($p=1.7x10^{-8}$), which represents a novel locus for BP variability in hypertensives and has not been detected in any of the previously published BP GWA studies. The effect size for the sentinel association was extremely small ($\beta=0.000179$) corresponding to a 0.01% unit change in BP variability per copy of the risk allele. Similar observations have been made in genome wide analysis of other measures of BP where effect sizes were also very small (1mmHg SBP and 0.5mmHg DBP) but could have the potential to significantly alter the outcomes at a population level. This evidence leads us to believe that the observed effect (albeit small) may be part of a battery of unrelated and common gene loci that exert independent but small effects, which compound to cause the disease. However this hypothesis can only be confirmed via large prospective GWAS studies.

In the initial phase of the project, there was a paucity of available cohorts with long term BP variability data and The ASCOT group lacked an opportunity to replicate the top hits with BP variability in an independent population. Since BP variability is strongly associated with ischemic stroke, we instead tested the top SNPs for association with ischemic stroke in an independent population comprising 8,624 cases and 12,722 controls from 7 cohorts. This is a common exploratory approach used to study candidate genes that maybe associated with different vascular disorders such as MI and stroke through their effect on shared risk factors such as HTN, diabetes and smoking (Cheng et al., 2012). Our sample size provided sufficient power to detect modest effect sizes ranging from 1.1-1.4 for overall ischemic stroke but as with other studies, had reduced power for subtypes due to small sample size. SNPs rs976683 and rs9830510 were not significantly associated ($p \leq 0.05$) with the risk of overall stroke or its subtypes with the estimated pooled ORs ranging from 1.02-0.96 for overall ischemic stroke, 1.07-

1.03 for cardio embolic, 0.98-0.76 for large vessel disease and 1.07-1.01 for small vessel disease.

Finally, a conditional analysis on the ASCOT Scandinavian (DK-FI-NO-SE) arm consisting of 3900 individuals provided no further evidence of an independent signal at the locus ($p = 0.18$). Failure to replicate this association could be attributed to population stratification induced by Anglo-Scandinavian differences such as admixture of Finnish and central European ancestry (Lao et al., 2008) and recruitment of the ASCOT-SE samples in Sweden. Further, the genetic effect could be confined to specific sub-populations of smokers, alcohol consumers and furosemide-exposed individuals.

The power to detect the effect size of a genetic risk variant is dependent on its minor allele frequency (Tabangin et al., 2009). It is interesting that the minor allele frequency of SNPs rs976683 and rs9830510 in both study populations were similar (0.25 and 0.15 respectively). However, even though the point estimates of the effect sizes observed for stroke were larger than BP variability, no comparative conclusions could be drawn from this, as neither SNP was significantly associated with the increased risk of stroke. *NLGN1* gene may play a role in BP variability via processes involving the growth and remodeling of the vascular system (Bottos et al., 2009).

Neuroligins are postsynaptic cell adhesion proteins that interact with neurexin at the synapse (Ichtchenko et al., 1995, Baudouin and Scheiffele, 2010). NLGN1 is one of 5 different isoforms (NLGN1 at 3q26, NLGN2 at 17p13, NLGN3 at Xq13, NLGN4 at Xp22, and NLGN4Y at Yq11). In humans, NLGN3 and NLGN4 have been associated with autism and mutations in NLGN4 have also been associated with intellectual disability, seizures, and EEG abnormalities. Together they form the neurexin/neuroligin adhesion system of synapses, which is highly conserved through evolution (Biswas et al., 2008). NLGN's are essential for important neurological processes and mutations in the gene leads to breakdown of critical interactions between brain cells and severe behavioral problems. The wide spread impact of the misfiring *NLGN1* gene is demonstrated in its association with various cellular processes. Studies in Caenorhabditiselegans mutants deficient in nlgn-1, an orthologue of human neuroligin genes, have demonstrated recovery of gentle touch response when transgenic NGLN1 mutants were expressed (Calahorro and Ruiz-Rubio, 2012). SNP rs6445141 in the NLGN1 gene was shown to be associated with

granular corneal dystrophy, type II in a GWAS study conducted on Korean CDG2 patients (Eun-Ju Lee, July 2011). In an Autism genome-wide copy number variation study on European Caucasian children, CNV region 174754378–174771975 was found to be associated with autism (p value=0.001)(Glessner et al., 2009). The NLGN1 protein in ubiquitously produced outside the central nervous system and expression of its α and β protein isoforms in the blood vessel walls and pancreatic beta-cells (Suckow et al., 2008) suggests roles in atherosclerosis and insulin regulation respectively. So far SNP rs976683 has only been implicated in Parkinson's disease (Edwards et al., 2010).

Study Strengths

The ASCOT GWA study is the first ever attempt to decipher the underlying genetic burden of long-term blood pressure variability and the only cohort in the world with long term BP variability data from 13 visits spanning over 5 years. A well phenotyped data set enabled the correlation of clinical markers with genotype data providing a first ever insight into the genetics of a novel measure of blood pressure.

The ASCOT study also addressed the possible confounding effect of anti-hypertensive medication on the BP variability and stroke phenotypes and their influence on the results. The ASCOT-Blood Pressure Lowering Arm cohort tested the impact of two treatment regimes, beta-blocker (Atenolol) and calcium channel blocker (Amplodipine), on hypertensives with a moderate risk of vascular disease. The study showed that the calcium channel blocker was more effective in lowering the risk of stroke by lowering both mean SBP and BP variability (Dahlof et al., 2005). Further support was provided in a meta-analysis on the effects of β-blocker selectivity on BP variability and stroke which concluded that nonselective β-blockers increase variability in BP which may explain their non-effectiveness in preventing stroke and MI (Webb et al., 2011).

Due to the opposite effects observed for the calcium channel and beta-blockers on BP variability, both treatment arms were analyzed separately. In an effort to rule out the pharmacogenetic effects, measurements for the first 6 months after starting both treatments were excluded from the BP variability phenotype calculation because this was a period of forced medication titration and any differential medication effects could have acted as confounders.

In the absence of a similar replication population, the ischemic stroke cohort provided a different but relevant phenotype, which is associated with elevated blood pressure. The replication study consisted of 8624 cases and 12722 controls of European descent from 7 studies spread across Europe, America and Australia providing sufficient power to detect small to modest effect sizes. Detailed demographics and stroke subtype classification enabled the sub-group analysis of LVD, SVD and CE stroke, which are known to have distinct genetic etiologies (Holliday et al., 2012). The study population was age and gender balanced in an effort to reduce confounding effects of these variables.

Study Limitations

The failure to detect an association with overall stroke could be due to several reasons. Genes affecting multi-factorial diseases such as stroke usually have small effect sizes and are difficult to identify in modestly sized study populations. Insufficient statistical power, given the small-observed effect size for rs96683 on BP variability, is the most likely cause of an undetectable association with ischemic stroke. Another reason could be the heterogeneous etiology of ischemic stroke, which makes it difficult to differentiate true signals from noise. Large studies such as the recent METASTROKE (Traylor et al., 2012) meta-analysis, which included 15 stroke cohorts comprising of 12,000 cases and 60,000 controls, also failed to identify any new genetic risk variants and only validated previous findings of variants within genes *PITX2*, *ZFHX3*, and *HDAC9*. Despite the large study population, the observed effect sizes were small (OR 1.39 - 0.96) suggesting that a combined burden of risk alleles carried by an individual is the likely cause, as shown in hemorrhagic stroke (Falcone et al., 2012). These studies have also highlighted the subtype-specific nature of the risk, which lends support to the fact that true association's maybe hidden under the multi-factorial pathogenesis of stroke.

Other limitations include possible inaccuracy of the TOAST classification into stroke sub-types. The case-control study design of our meta-analysis may be another limitation as some studies can induce survival bias by including recurrent stroke, thus allowing selection of milder forms of strokes. Identified SNPs from the ASCOT GWAS could also be artifactual, since they were not successfully replicated in an independent BP variability cohort.

Another limitation was the inability to perform an independent replication in a suitable population, which is a standard strategy to confirm, identified positive results in the discovery stage of a GWAS. This was due to a paucity of cohorts with long-term blood pressure variability data in which to replicate. This is not unprecedented as other GWAS groups who are also dealing with unusual phenotypes or phenotypes not often measured, such as cervical dissection in stroke and venous thrombosis, have found it difficult to bring together large study populations for replication. To overcome this issue we employed an alternate strategy to replicate in a downstream phenotype such as stroke.

The replication population used to follow up further the original association of rs976638 with BP variability was the ASCOT Scandinavian (DK-FI-NO-SE) population consisting of 3900 individuals, in whom BP was measured and at the same time points as the UK and Irish ASCOT individuals. This was the only other potential replicable population we were able to identify. The cohort consisted of hypertensive individuals from different Scandinavian countries i.e. Denmark, Finland, Norway and Sweden. The Finnish and Swedish populations are known to have admixture (Lao et al., 2008, Humphreys et al., 2011), and be genetically heterogeneous compared to other European descent populations. This could be further compounded by the fact that the ASCOT-SE samples were recruited in Sweden. In such a mixed population, BP variability may occur disproportionally in a sub-population with higher prevalence and alleles that are more common in this population may become falsely associated with the disease leading to spurious associations. Further, it is also possible that the genetic effect is confined to specific sub-populations of smokers, alcohol consumers and furosemide-exposed individuals within the ASCOT-UK-IR cohort. Identified SNPs from the ASCOT-UK-IR GWAS could also be artifactual.

Therefore, although not an ideal resource for follow-up of the original observation with BP variability the ASCOT Scandinavian population was the only available resource. Its strength was that these individuals were selected using identical recruitment criteria as the UK and Irish ASCOT individuals and BP measurements were taken at the same time points allowing identical analysis of BP variability.

Conclusion

BP variability is associated with adverse outcome independent of treatment suggesting that the genetic burden is a genuine driver of BP variability in the general hypertensive population. In order to understand the true relationship between visit-to-visit BP variability and risk of stroke, large prospective longitudinal studies following healthy cohorts for stroke occurrence are required. There is a need for international guidelines for clinical monitoring of BP variability that advocate diagnosis and assessment of treatment response in HTN to be based upon the average of a series of blood pressure measures. Calibration of measuring devices is also needed to avoid phenotypic bias.

The ASCOT GWA study lends support to a genetic contribution to BP variability and suggests a cluster of genetic variants that maybe associated with it. Negative replication results for the BP variability locus (albeit in stroke and another in a population with admixture) conclude that the lead identified SNP was not associated with ischemic stroke or its subtypes. This does not negate the initial results but simply means that the identified SNPs do not play a major role in the genetic etiology of ischemic stroke. Nor does it mean that BP variability does not influence ischemic stroke in a physiological fashion nor does it exclude other genetic associations on it. If an association does exist, it is likely to be modest for overall stroke or restricted to a single stroke subtype. It is likely that the causal variant differs from the identified SNPs and the functional region may also lie away from the implicated candidate gene (Kleinjan and van Heyningen, 2005).

The current study is the first to report any attempt at dissecting the genetics of BP variability and its effect on ischemic stroke. The findings argue strongly that based on the initial findings future BP studies should include long-term visit-to-visit BP variability, which is an unusual yet potentially important phenotype.

3.5 Figures and Tables

Figure 3.1: Genome-wide QQ and Manhattan plots for the ASCOT Anglo IR-UK GWAS showing a cluster of 17 *NLGN1*SNPs associated with BP variability ($p<5 \times 10^{-7}$). Individual $-\log_{10}$ p values are plotted against their genomic position by chromosome. The dotted line at 10^{-6} marks the threshold for promising SNPs and the solid line at 10^{-8} marks the genome-wide significance threshold. Figures reproduced with permission from the ASCOT group (Dr. Patricia Munroe, p.b.munroe@qmul.ac.uk).
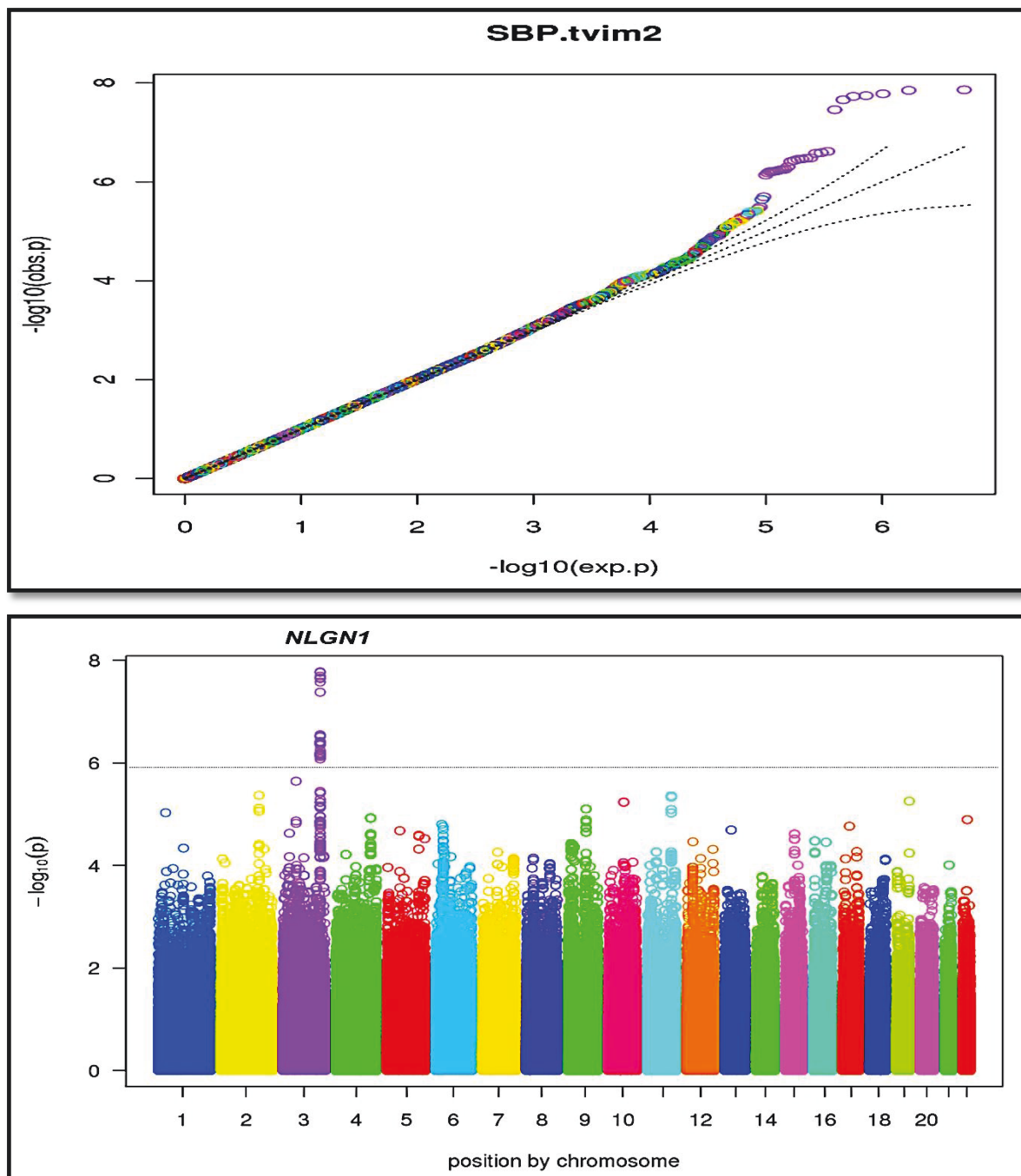
Figure 3.2: Regional association and LD plots for the 17 correlated SNPs within the *Neuroligin-1* (*NLGN1*) gene (3q26.31). The plots A & B are conditioned on the imputed sentinel SNP rs976683 and C & D are conditioned on the top genotyped SNP rs9830510. In plot 2A & C, each colored square represents a SNP p value, with the color scale correlating the r² values for that SNP to the target SNP (red diamond) taken from the HapMap phase 2 CEU panel. In plot 2B & C, the target SNP (orange diamond) is represented in linkage disequilibrium with the cluster of 16 SNPs and other SNPs in the HapMapphase 2 CEU panel.
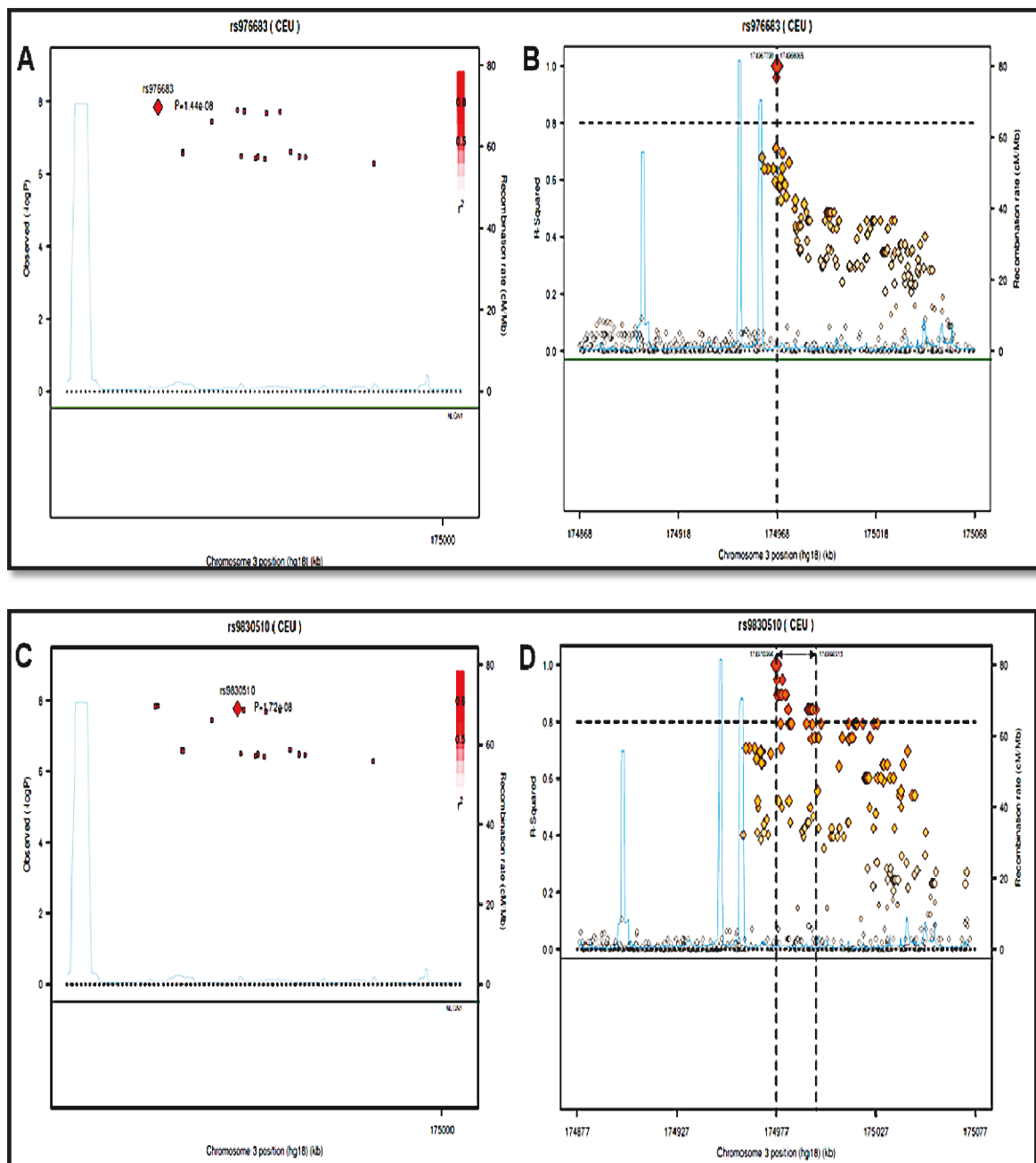
Table 3.1: Details of genotyping, imputation and quality control for the ischemic stroke cohorts

	ASGS	BRAINS	GEOS	ISGS-SWISS	VISP	WTCCC2-UK	WTCCC2-Ger
Genotyping Platform	Illumina Human 610 Quad	Illumina Human 610 Quad	HumanOmni1-Quad_v1-0_B BeadChip	IlluminaHuman Hap 550k	Illumina HumanOmni1-Quad v1-0 B	Illumina 660	Illumina Human660W-Quad (cases) &Human 550k (controls)
Genotyping Calling algorithm	Genome studio	Genome studio V2010.1 Genotyping module	IlluminaBeadStudio version3.3.7	IlluminaBeadStudio	GenomeStudio V2010.2 Genotyping Module V1.7.4 GenTrain version 1.0	Gencall	Illuminus
Call rate threshold (Individuals)	≥ 0.95	≥ 0.95	>0.98	≥ 0.95	≥ 0.95	0.95	Bayesian clustering
Call frequency threshold (SNPs)	≥ 0.95	≥ 0.95	>0.95	≥ 0.95	≥ 0.95	0.95	0.95
Imputation software	MACH 1.0.16	MACH 1.0	BEAGLE V3.3	MACH 1.0	MACH 1.0	MACH	MACH
Imputation build	HapMap build 36 release 24	HapMap Build 36 Release 22	HapMap Build 36 Release 22	1000 genomes (06_2010)	HapMap build 36 release 22	HapMap II	HapMap 2
LD threshold (r²) for surrogate markers	0.8	0.8	0.8	0.8	0.8	0.8	0.8

Imputed Quality score threshold for imputed SNP	0.3	0.3	0.3	0.3	0.3	0.3	0.3

Table 3.2: Details for statistical analysis forthe ischemic stroke cohorts

	ASGS	BRAINS	GEOS	ISGS-SWISS	VISP	WTCCC2-UK	WTCCC2-Ger
Model	Logistic regression	Logistic regression	Logistic Regression	Logistic Regression	Logistic regression	Logistic Regression	Additive model, Bayesian hierarchical model.
Adjustment covariates	Sex and age	Sex and age	age, study recruitment stages and MDS (component 1)	Sex, age, principal components 1 & 2	Sex, age, PC1, PC2	None	none
Statistical software	Plink, mach2dat, SAS	Plink v1.07 , STATA v11, SPSS v20, METAL	PLINK v1.07	PLINK v1.07 for data cleaning, MACH for imputation, R and MACH2DAT for generation of summary statistics	Plink v1.07	Plink & METAL	SNPTEST, own software

Table 3.3: ASCOT-UK-IR stroke cohort population demographics

Clinical Phenotype	
N	3802
Age (mean ± SD)	63.7 ± 8.1
Males, N (%)	3131 (82%)
SBP baseline (Mean ± SD)	161.6 ± 17.6
DBP baseline (Mean ± SD)	92.4 ± 9.9
VIM (Mean ± SD)	0.004 (±0.001)

Table 3.4: Association results from the ASCOT GWAS identifying 17 correlated *NLGN1* SNPs (p≤5x10^{-7})

SNP	Position	A1/A2	RAF	r^2	β	SE	p
rs976683†	174968065	C/T	0.24	0.95	0.0001786	3.15x10^{-5}	1.44 x10^{-8}
rs12635897†	174967790	C/G	0.24	0.95	0.0001784	3.15x10^{-5}	1.49 x10^{-8}
rs9830510	174976996	A/G	0.86	1.00	-0.000215	3.81 x10^{-5}	1.72 x10^{-8}
rs9882520	174977714	A/G	0.87	0.99	-0.000217	3.86 x10^{-5}	1.88 x10^{-8}
rs12495045	174981764	A/C	0.13	0.99	0.0002175	3.87 x10^{-5}	1.91 x10^{-8}
rs6776924	174980201	A/T	0.87	0.99	-0.000216	3.85 x10^{-5}	2.12 x10^{-8}
rs1948161†	174974090	C/T	0.81	0.96	-0.000189	3.43 x10^{-5}	3.55 x10^{-8}
rs4377507	174982953	A/G	0.89	0.99	-0.000215	4.16 x10^{-5}	2.49 x10^{-7}
rs6779230†	174970831	A/C	0.72	0.96	-0.000153	2.96 x10^{-5}	2.55 x10^{-7}
rs6779246†	174970869	C/G	0.29	0.96	0.0001521	2.96 x10^{-5}	2.77 x10^{-7}
rs9868353	174977376	A/G	0.12	0.99	0.0002028	3.97 x10^{-5}	3.27 x10^{-7}
rs7428277	174979295	A/G	0.12	0.99	0.0002035	3.99 x10^{-5}	3.37 x10^{-7}
rs9876713	174983921	A/G	0.11	0.99	0.0002117	4.15 x10^{-5}	3.38 x10^{-7}
rs1488549	174984586	C/T	0.11	0.99	0.0002116	4.15 x10^{-5}	3.43 x10^{-7}
rs4568169	174978999	A/T	0.88	0.99	-0.000202	3.97 x10^{-5}	3.66 x10^{-7}
rs6774109	174980026	A/G	0.12	0.99	0.0002015	3.97 x10^{-5}	3.85 x10^{-7}
rs7629797	174992286	C/T	0.89	1.00	-0.000208	4.14 x10^{-5}	5.10 x10^{-7}

* Effect sizes are shown as a unit or percentage change in BP variability per copy of the risk allele.

**† represents imputed SNPs. The sentinel SNP rs976683 and top genotyped SNP rs9830510 are in bold.

Table 3.5: Details of population demographics for ischemic stroke cohorts

	ASGS		BRAINS		GEOS		ISGS-SWISS	
	Case	**Control**	**Case**	**Control**	**Case**	**Control**	**Case**	**Control**
N	1162	1244	342	2473	448	498	1070	1488
Age in years (mean±SD)	72.87 ± 13.16	66.28 ± 7.54	71.43 ± 14.02	45 ± 0	41.0 (7.0)	39.5 (6.7)	66.62 ± 13.63	64.12 ± 17.29
Male n (%)	688 (59.21)	625 (50.24)	191 (56)	1292 (52)	275 (61.4)	282 (56.6)	607 (57%)	715 (48%)
IS stroke subtype,n (%)								
-Cardio embolic	240	---	79	---	90	---	247	---
-Large Artery	421	---	42	---	37	---	229	---
-Small Vessel	310	---	30	---	54	---	201	---
HTN, n (%)	732 (63.99)	809 (65.08)	240 (71)	---	137 (30.6)	79 (15.9)	691 (65)	518 (35)
Diabetes, n (%)	249 (21.75)	126 (10.52)	46 (14)	---	52 (11.6)	12 (2.4)	220 (20)	163 (11)
Hypercholestrimemia, n (%)	435 (42.48)	513 (41.24)	145 (44)	---	126 (28.1)	117 (23.5)	NA	NA
Smoking, n (%)	207 (18.45)	80 (6.67)	69(21)	---	187 (41.7)	117 (23.5)	196 (18)	716 (48)

Table 3.5: Population demographics continued

	VISP		WTCCC2-UK		WTCCC2-Ger	
	Case	**Control**	**Case**	**Control**	**Case**	**Control**
N	1726	1047	2702	5175	1174	797
Age in years (mean±SD)	67.99 ± 10.66	51.22 ± 12.57	72.1 ± 12.5	---	66.7 ± 12.9	62.7 ± 10.9
Male n (%)	1121 (65)	622 (59)	1468 (54.3)	---	727 (62)	410 (51)
IS stroke subtype, n (%)						
-Cardio embolic	---	---	537	---	330	---
-Large Artery	---	---	564	---	346	---
-Small Vessel	---	---	553	---	106	---
HTN, n (%)	1203 (70)	---	1936 (71.1)	---	751 (64)	---
Diabetes, n (%)	429 (25)	---	403 (14.0)	---	270 (23)	---
Hypercholestrimemia, n (%)	140 (8)	---	1280 (47.4)	---	479 (41)	---
Smoking, n (%)	860 (53)	---	1785 (66.1)	---	366 (31)	---

Table 3.6: SNPs rs976683 and rs9830510 characteristics in the ischemic stroke meta-analysis cohorts

Cohorts	rs976683			rs9830510			Imputed/Genotyped
	Minor Allele	Major Allele	MAF	Minor Allele	Major Allele	MAF	
ASGC	C	T	0.28	G	A	0.15	Genotyped
BRAINS	C	T	0.24	G	A	0.16	Genotyped
GEOS	C	T	0.25	G	A	0.15	Genotyped
ISGS-SWISS	C	T	0.28	G	A	0.17	Genotyped
VISP	C	T	0.28	G	A	0.17	Genotyped
WTCCC-UK	C	T	0.25	G	A	0.15	Genotyped
WTCCC-Ger	C	T	0.25	G	A	0.16	Genotyped

Table 3.7: Association results for SNP rs976683 with overall ischemic stroke and its subtypes

Stroke	Cohorts	N	A1/A2	Association		Heterogeneity	
				OR (95% CI)	p	Q (p)	I²
All stroke	7	8624	t/c	1.02 (0.97-1.07)	0.52	4.85 (0.56)	0
CE	6	1523	t/c	1.07 (0.97-1.16)	0.17	3.31 (0.65)	0
LVD	6	1639	t/c	0.98 (0.89-1.07)	0.60	5.41 (0.37)	7.6
SVD	6	1254	t/c	1.07 (0.97-1.17)	0.19	2.92 (0.71)	0

Table 3.8: Association results for SNP rs9830510 with overall ischemic stroke and its subtypes

Stroke	Cohorts	N	A1/A2	Association		Heterogeneity	
				OR (95% CI)	p	Q (p)	I²
All stroke	7	8624	a/g	0.96 (0.90-1.02)	0.54	2.37 (0.88)	0
CE	6	1523	a/g	1.03 (0.91-1.15)	0.83	4.43 (0.49)	0
LVD	6	1639	a/g	0.76 (0.66-0.87)	0.03	1.43 (0.92)	0
SVD	6	1254	a/g	1.01 (0.89-1.14)	0.92	5.28 (0.38)	5.4

* Effect sizes are shown as odds ratios for the % increase or decrease per copy of the risk allele.

CHAPTER 4: META-ANALYSIS OF GENOME WIDE ASSOCIATION STUDIES ON CAROTID STENOSIS IN ISCHEMIC STROKE

Abstract

Carotid stenosis of >50% is associated with high risk of ischemic stroke accounting for nearly 20% of all ischemic strokes and TIA's caused by emboli originating from unstable plaque (>50% stenosis) in large arteries. The study hypothesized that there is a shared genetic burden for ischemic stroke and carotid disease, and individuals with superimposed carotid stenosis on ischemic stroke may carry a greater genetic risk burden than those with either standalone disease.

A meta-analysis of 7 genome wide association studies was conducted to increase the statistical power of identifying novel gene variants associated with the risk of carotid disease in ischemic stroke. A total of 14,867 individuals of European ancestry (1164 cases and 13,703 healthy controls) and approximately 2 million genotyped and imputed SNPs were tested for association using an additive genetic model. Evidence of a shared genetic basis of carotid disease and stroke was also investigated by analyzing previously published genome wide significance GWAS-derived SNPs associated with stroke and carotid disease.

The study identified three genomic regions in chromosomes 3 to be associated at genome wide significance (p value $\leq$ 5 x 10^{-8}) with > 50% carotid stenosis in ischemic stroke. The associated SNPs mapped in or near genes *LRIG1* (OR 0.39, 95% CI 0.13-0.64, p=9.69x10^{-13}), *ROBO1* (OR 0.66, 95% CI 0.53-0.78, p=3.10x10^{-10}) and *CAPN7* (OR 0.58, 95% CI 0.39-0.76, p=5.89x10^{-9}).

Three genetic risk variants associated with > 50% carotid stenosis and ischemic stroke were identified. Findings of the study may help in a better understanding of the genetic risk burden of symptomatic carotid stenosis.

4.1 Introduction

Carotid disease is a major risk factor for ischemic stroke and > 50% carotid stenosis accounts for nearly 20% of all occurring ischemic stroke in the middle and anterior cerebral artery territories (Chaturvedi et al., 2005, Fairhead et al., 2005). The prevalence of moderate to high grade extra cranial carotid artery stenosis (50-99%) in adults ranges from 1–3 % and doubles to 6-9% in individuals >65 years of age (Eckstein et al., 2012). Embolising plaques or acute carotid occlusions cause cerebral ischemia in 1-5% of all patients with an asymptomatic 50-99% stenosis of the internal carotid artery (Eckstein et al., 2012). Stroke risk in patients with > 50% carotid stenosis is directly related to the extent of stenosis; patients with 60-69% stenosis have 11% risk of stroke within 3 years, compared with 32% for patients with >90% stenosis (Jashari et al., 2012).

Carotid plaque, defined as focal thickening of > 50% is a heritable trait (h^2=23%-28%) and is under appreciable additive genetic influences (Hunt et al., 2002). It is strongly related to early parental CHD death as compared to IMT (Zureik et al., 1999) and with the risk of vascular events (Inaba et al., 2012). Evidence supporting the overlap of genetic risk factors between carotid disease and stroke exists and has revealed a shared genetic basis of these complex traits which is most likely through the effect of common risk factors such as blood pressure (Zakopoulos et al., 2005). For example, genetic variants at the 9p21 locus have been found to be associated with the risk of carotid disease and stroke (Congrains et al., 2012, Gschwendtner et al., 2009, Holdt and Teupser, 2012) and are known to exert their effect through modulation of ANRIL and CDKN2A/B. A recent study has identified a risk association of a microRNA SNP (rs3735590) near the PON1 gene with both ischemic stroke and IMT (Liu et al., 2013). Another example is the *MTHFR* gene, which is strongly associated with stroke via modulation of its biochemical marker Hcys, is also associated with carotid disease (Bova et al., 1999). Increasing numbers of epsilon4 alleles of the well-studied stroke associated APOE gene were also found to be associated with the risk of developing carotid plaque in men (Beilby et al., 2003). Evidence showing shared causality further lends support to mechanistic functioning of genetic risk variants (Kamstrup et al., 2012).

Although there are common genetic influences for carotid stenosis and stroke, it is likely that each has its own set of unique genes that are independently associated with both traits, as shown in other carotid disease phenotypes (Spence and Hegele, 2004). Studies reporting weak associations for genetic determinants of cardiovascular risk with carotid disease also support this view (Manolio et al., 2004). Additionally phenotypes of early stage carotid disease (IMT) are not associated with genetic risk factors of coronary artery disease and MI, suggesting that these genetic risk factors may not exert their effect via early vascular remodeling or early arthrosclerosis and may play a role in the late stage disease (Conde et al., 2011).

Despite a shared genetic etiology there is a genetic distinction between ischemic stroke and carotid stenosis (Kostulas et al., 1999) and stroke patients with superimposed carotid stenosis are likely to carry a greater genetic burden as compared to individuals with only stroke or healthy controls. Genetic investigations of complex disorders such as stroke are often confounded by heterogeneity of disease pathogenesis. Genetic variants associated with ultrasound-confirmed carotid disease may differ from variants associated with clinically relevant carotid stenosis and requires the study of carotid stenosis in stroke as a distinct phenotype. So far several large GWA studies (Bis et al., 2011) and smaller candidate gene studies (Lanktree et al., 2009) have been published on carotid disease phenotypes in the healthy general population; however no studies have studied clinically relevant carotid disease in ischemic stroke.

Based on published evidence, the study hypothesized that ischemic stroke and carotid disease carry a shared genetic burden and individuals with superimposed carotid stenosis on ischemic stroke may carry a greater genetic risk burden than those with either standalone disease or no disease. Previous studies have examined carotid stenosis and ischemic stroke as distinct phenotypes but so far no study has pursued both traits in the same population i.e. symptomatic carotid stenosis. An *a priori* genome wide association study to identify gene variants associated with > 50% carotid stenosis in patients with ischemic stroke was conducted. In order to correlate the findings with published studies, evidence of a shared genetic basis of carotid disease and stroke was investigated by analyzing previously published genome wide significant GWAS-derived SNPs associated with stroke and carotid disease.

4.2 Methods

4.2.1 Study Populations

14,867 individuals of European ancestry were included in the discovery meta-analysis of 7 GWA studies on carotid stenosis in ischemic stroke. The studies included individuals with genotype and phenotype data from the following cohorts: BRAINS (n=2527)(Cotlarciuc I, 2012, Yadav et al., 2011), GAROS 1 (n=3060)(Rost et al., 2010), GAROS 2 (n=152)(Rost et al., 2010), GEOS (n=545)(MacClellan et al., 2006, Kittner et al., 1998), ISGS (n=1433)(Meschia et al., 2006), WTCCC-Ger (n=1162)(International Stroke Genetics et al., 2012) and WTCCC-UK (n=5988)(International Stroke Genetics et al., 2012) **(Figure 4.1)**. All participating cohorts received institutional ethical clearance and signed consent from each participating study subject. Detailed description of the study cohorts are as follows:

BRAINS: BRAINS cohort (http://www.brainsgenetics.com) is an on-going, multicenter, in-hospital study which recruits consenting acute stroke _cases_ into a highly characterized bio bank (Yadav et al., 2011, Cotlarciuc I, 2012). All adult (>18 years of age) stroke patients are recruited with either ischemic or hemorrhagic pathology MRI confirmed lesions. Ischemic stroke subtypes are further sub-classified according to TOAST criteria (Adams et al., 1993). Known monogenic causes of stroke are excluded. BRAINS has two principal arms; the first arm recruits UK European stroke patients (BRAINS-UK) while the second arm recruits South Asian stroke patients from multiple sites in the UK and also from sites in India (BRAINS-South Asia). _Control_ data for the European arm is provided by the British 1958 Birth Cohort (http://www.b58cgene.sgul.ac.uk) while control subjects for the South Asian arm are recruited simultaneously as the affected stroke patient and usually is the proband's spouse. A total of 2527 (54 cases and 2473 controls) individuals were included from the BRAINS cohort.

GAROS 1 and GAROS 2: _Cases_ included consecutive patient's ≥ 18 years of age presenting with ischemic stroke and admitted to the Massachusetts General Hospital (MGH) Stroke Unit through the Emergency Department, or evaluated in the MGH Neurology outpatient clinics, or on the inpatient Medical and Vascular Surgical services from 2003 to 2010. A stroke was defined as either (1) a radiographically proven (head

CT or MRI) infarct associated with the appropriate clinical stroke syndrome or (2) a fixed neurological deficit persisting more than 24 hours, consistent with a vascular pattern of involvement and without radiographic evidence of demyelinating disease, or other non-vascular structural disease. _Controls_ are 55 years or older recruited from the stroke-free adults presenting to the MGH outpatient clinics, matched with the cases on the basis of age, sex and ancestry information obtained from principal component analysis of GWAS data. Potential control subjects with a known history of carotid stenosis, transient monocular blindness, transient ischemic attack, ischemic or hemorrhagic stroke, systemic embolization, other thrombotic episodes, or surgery within the last 3 months were excluded. The Institutional Review Board approved all aspects of this study, and informed consent for collection of data was obtained for all subjects or their next of kin (case subjects only). Exclusion criteria for both case and control subjects were: (1) Stroke due to vasculitis, (2) Stroke due to endocarditis, (3) Active cancer or cancer treatment within the past two years, (4) Brain tumor, (5) Subdural hematoma, (6) Active lupus, (7) Recent cocaine use, (8) Meningitis, (9) Active myeloproliferative disorder, (10) Known genetic cause for stroke (i.e. MELAS, CADASIL), (11) Traumatic carotid dissections with no other stroke risk factors (i.e. neck injuries/manipulations), (12) End-stage liver disease and, (13) End-stage renal disease. A total of 3060 (32 cases and 3028 controls) and 152 (15 cases and 137 controls) individuals were included from GAROS 1 and GAROS 2 cohorts respectively.

GEOS: GEOS is a population-based case-control study designed to identify genes associated with early-onset ischemic stroke and to characterize interactions of identified stroke genes and/or SNPs with environmental risk factors (Cheng et al., 2011). Participants were recruited from the greater Baltimore-Washington area in 4 different time periods: Stroke Prevention in Young Women-1 (SPYW-1) conducted from 1992-1996, Stroke Prevention in Young Women-2 (SPYW-2) conducted from 2001-2003, Stroke Prevention in Young Men (SPYM) conducted from 2003-2007, and Stroke Prevention in Young Adults (SPYA) conducted in 2008. _Case participants_ were hospitalized with a first cerebral infarction identified by discharge surveillance from one of the 59 hospitals in the greater Baltimore-Washington area and direct referral from regional neurologists. The abstracted hospital records of cases were reviewed and adjudicated for ischemic stroke subtype by a pair of neurologists according to

previously published procedures with disagreements resolved by a third neurologist. The ischemic stroke subtype classification system retains information on all probable and possible causes, and is reducible to the more widely used TOAST system that assigns each case to a single category. *Control participants* without a history of stroke were identified by random-digit dialing and were balanced to cases by age and region of residence in each recruitment periods. A total of 545 (47 cases and 498 controls) individuals were included from the GEOS cohort.

ISGS: ISGS is a multicenter inception cohort study of first-ever ischemic stroke in adult men and women (Meschia et al., 2003). *Cases* were recruited from inpatient stroke services at five academic medical centers in Florida, Georgia, Virginia and Minnesota. A study neurologist on the basis of medical history, physical examination and CT or MR imaging of the brain confirmed the diagnosis of ischemic stroke. Cases had to be enrolled within 30 days of onset of stroke symptoms. Cases were excluded if they had a mechanical aortic or mitral valve, central nervous system vasculitis, or bacterial endocarditis at the time of the stroke. They were also excluded if they were known to have: CADASIL, Fabry disease, homocystinuria, MELAS, or sickle cell anemia. Stroke severity at enrollment was assessed using the NIHSS and outcomes at 90-days were assessed by telephone using the Barthel Index, Glasgow Outcome Scale, and the modified Rankin scale (Kasner, 2006). Diagnostic evaluation included: head CT (95% of individuals enrolled) or MRI (83%), electrocardiography (92%), cervical arterial imaging (86%), and echocardiography (74%). A vascular neurology committee reviewed the medical records of every case and assigned ischemic stroke subtype diagnoses according to criteria from the Trial of ORG10172 (TOAST) (Adams et al., 1993), the OCSP (Bamford et al., 1991), and the Baltimore-Washington Young Stroke Study (Johnson et al., 1995). DNA was donated to the NINDS DNA Repository (Coriell Institute, Camden, NJ) for eligible samples with appropriate written informed consent. A separate certified neurologist adjudicator additionally assigned a subtype diagnosis using the standardized CCS web-based algorithm (Ay et al., 2007). A total of 1433 (76 cases and 1357 controls) individuals were included from the ISGS cohort.

WTCCC-Germany and WTCCC-UK: The WTCCC2 samples were genotyped as part of the WTCCC 2 ischemic stroke studies (International Stroke Genetics et al., 2012). Stroke *cases* included samples recruited by investigators at St. George's University London (SGUL), University of Oxford and Edinburgh Stroke Study in the UK and the Department of Neurology, KlinikumGroßhadern, Ludwig-Maximilians-University, Munich. The SGUL collection comprised 1224 ischemic stroke samples from a hospital based setting. All cases were of self-reported Caucasian ancestry. Ischemic stroke subtypes were determined according to TOAST criteria based on relevant clinical imaging and available information on cardiovascular risk factors. The University of Oxford collection comprised 896 ischemic stroke cases, consecutively collected as part of the Oxford vascular study (OXVASC). Cases were of self-reported Caucasian ancestry, and ischemic stroke subtypes were determined according to TOAST criteria based on relevant clinical imaging. For the Edinburgh Stroke Study, consecutive consenting patients with stroke who were admitted to or seen as outpatients at the Western General Hospital, Edinburgh were prospectively recruited between 2002 and 2005. Cases in this study were those with a clinically evident stroke, demonstrated by brain imaging (CT or MRI) to be ischemic. An experienced stroke physician assessed each patient as soon as possible after the stroke, prospectively recording demographic and clinical details, including vascular risk factors and results of brain imaging and other investigations. The Munich samples included 1383 ischemic stroke cases. Cases were consecutive European Caucasians recruited from a single dedicated Stroke Unit at the Department of Neurology, KlinikumGroßhadern, Ludwig-Maximilians-University, Munich. Ischemic stroke subtypes were determined according to TOAST criteria based on relevant clinical and imaging data.

Controls for the UK samples were drawn from shared WTCCC controls obtained from the 1958 Birth Cohort. This is a prospectively collected cohort of individuals born in 1958 (http://www.b58cgene.sgul.ac.uk/), and ascertained as part of the national child development study (http://www.cls.ioe.ac.uk/studies.asp). Data from this cohort are available as a common control set for a number of genetic and epidemiological studies. For the German samples controls were Caucasians of German origin participating into the population KORAgen study (www.gsf.de/kora). This survey represents a gender- and age stratified random sample of all German residents of the Augsburg area and

consists of individuals 25 to 74 years of age, with about 300 subjects for each 10-year increment. All controls were free of a history of stroke or transient ischemic attack. A total of 5988 (596 cases and 5392 controls) and 1162 (344 cases and 818 controls) individuals were included from WTCCC-Ger and WTCCC-UK cohorts respectively.

4.2.2 Carotid Stenosis Phenotype

Each study evaluated the carotid arteries using site-specific high-ultrasound B-mode ultrasonography. Carotid plaques are known to increase the risk of stroke irrespective of their location and are not consistently related to higher risks of infarction in the ipsilateral versus collateral cerebral hemispheres (Hollander et al., 2002). Therefore we used a single maximum measurement for stenosis, either bilateral or unilateral, in the common carotid arteries (CCA), external carotid arteries (ECA) and internal carotid arteries (ICA) to calculate the phenotype. For example, if the degree of carotid stenosis in a patient was as following; ICA (20-50%), ECA (50-70%) and CCA (70-90%), the maximal stenosis value of 90% was used as the carotid phenotype. This allowed us to take into account the severity of stenosis irrespective of the location. Patients with > 50% carotid stenosis were chosen as cases based on the cut off for clinically relevant carotid stenosis (Thapar et al., 2013).

4.2.3 Genotyping and Quality Control

BRAINS: Whole genome DNA was isolated from EDTA-blood samples. BRAINS samples were genotyped on the Illumina HumanHap 610-quad Beadchip according to manufacturer's standard protocol. The genotyping was performed with the help of Ms. Kerra Pearce at the University College London (k.pearce@ucl.ac.uk). Details of the procedure are listed in **Appendix 1**.

Genome studio V2010.1 Genotyping module was used for the initial quality control of samples and all samples below the genotyping call rate of <95% were removed. The GenCall algorithm clustered the SNPs and plots were manually examined for improper clustering. Further quality control was performed in PLINK to remove individuals failing the following filters: (1) Call rate ≤ 95%, (2) Non-European ancestry, (3) Outlying autosomal heterozygosity, and (4) Cryptic relatedness (pi-hat ≥ 0.2). Quality control

also removed SNPs failing the following filters: (1) Call frequency ≤ 95%, (2) MAF ≤ 0.01 and (3) HWE ≤ 10^{-6}.

GAROS 1 and GAROS 2: Genomic DNA was extracted from blood for all subjects using standard operating protocols. Samples were genotyped at the Broad Institute, Cambridge using the Affymetrix 6.0 and Illumina Human 610 Quad bead chips for GAROS 1 and GAROS 2 respectedly.

GEOS: Genomic DNA was isolated from a variety of sample types, including cell line, whole blood, mouthwash and buccal swab. Samples were genotyped at the Johns Hopkins Center for Inherited Disease Research (CIDR) using the Illumina HumanOmni1-Quad_v1-0_B BeadChip (Illumina, San Diego, CA, USA). Individuals were excluded if they were unexpected duplicates, gender discrepancy and unexpected relatedness.

ISGS: The ISGS/SWISS samples were genotyped on Illumina HumanHap550K chips and the Illumina Bead Studio software was used for the initial quality control of samples. Further quality control was performed in PLINK to remove individuals failing the following filters: (1) Call rate ≤ 95%, (2) Non-European ancestry, (3) Outlying autosomal heterozygosity, and (4) Cryptic relatedness (pi-hat ≥ 0.2). Quality control also removed SNPs failing the following filters: (1) Call frequency ≤ 95%, (2) MAF ≤ 0.01 and (3) HWE ≤ 10^{-6}.

WTCCC-Ger and WTCCC-UK: All WTCCC2 cases were genotyped as part of the WTCCC2 Ischemic Stroke study using the Illumina Human610W-Quad array. British controls were genotyped using the Illumina Human1.2M-Duo. German controls were genotyped on the Illumina Human 550k platform. Quality control procedures in the WTCCC2 excluded SNPs not genotyped on all case and control collections and SNPs with Fisher information measure <0.98, genotype call rate <0.95, MAF <0.01 or HWE p value ≤ 1×10^{-20} in either the case or control collections. Samples were excluded if identified as outliers on call rate, heterozygosity, ancestry and average probe intensity based on a Bayesian clustering algorithm. Samples were also removed if they exhibited discrepancies between inferred and recorded gender and cryptic relatedness with other WTCCC2 samples (pair wise identity-by-descent >0.05). Autosomal genotype imputation was performed using MACH based on HapMap Phase 2 European (CEU) reference data. Details of the genotyping are provided in **Table 4.1**.

4.2.4 Genotype Imputation

Single Nucleotide polymorphisms analyzed in GWA studies are mostly derived experimentally from gene arrays that produce high throughput data in the hundreds of thousands. Genotypes are called from their allele signal fluorescent intensities and improper separations of signals are registered as missing genotypes. There is a huge loss of genetic information due to these missing genotypes that can introduce bias in the analysis and an inability to combine GWA studies that have used varied genotyping platforms that differ in their coverage of the genome. Re-genotyping missing data is an expensive and time-consuming option. Imputation provides a good alternative to filling in these missing genotype data gaps in the dataset and reducing the impact of missing data on association analysis.

Imputation methods are based on the assumption that when SNPs are in LD, the unobserved genotypes can be predicted based on genotypes observed at other SNPs. The Markov Chain based haplotyper (MACH) software (Li et al., 2006) is a popular choice for inferring missing genotypes in samples of unrelated individuals using observed genotypes together with haplotypes from a reference dataset such as the HapMap or 1000Genomes.

BRAINS samples were imputed using the MACH 1.0 software and the HapMap Build 36 Release 22 imputation build. SNPs with a squared correlation between imputed and true genotypes (r^2) <0.3, MAF $\leq$ 0.01, HWE $\leq 10^{-6}$ and pre-imputation call frequency $\leq$ 0.95 were removed from further analysis. Dr. Ioana Cotlarciuc performed the imputation for the BRAINS samples.

GAROS 1 and GAROS 2 samples were imputed using the IMPUTE v2.2.2 software and the 1000 Genomes Phase I integrated variant set and Haplotype release set (March 2012, updated Aug 24 2012). SNPs with MAF $\leq$ 0.01, HWE $\leq 10^{-6}$ and pre-imputation call frequency $\leq$ 0.95 were removed from further analysis.

GEOS samples were imputed using the IMPUTE v2.2.2 software and the 1000 Genomes phase 1 (June 2011) reference data set. SNPs with a squared correlation between imputed and true genotypes (r^2) <0.3, monomorphic SNPs, HWE $\leq 10^{-4}$ and pre-imputation call frequency $\leq$ 0.95 were removed from further analysis.

ISGS samples were imputed using the MACH 1.0 (miniMACH) software and the 1000 Genomes.2012-03-14 imputation build. SNPs with a squared correlation between imputed and true genotypes (r^2) <0.3 and MAF ≤ 0.01 were removed from further analysis.

WTCCC-Ger and WTCCC-UK samples were imputed using the IMPUTE software and the HapMap II imputation build 36. SNPs with a squared correlation between imputed and true genotypes (r^2) <0.3 and MAF ≤ 0.01 were removed from further analysis.

Details of the genotype imputation are provided in **Table 4.1**.

4.2.5 Genome Wide Association Analysis

Each participating study cohort performed its own site-specific logistic regression analysis using PLINK software (Purcell et al., 2007) to test association of genotyped and imputed SNPs, which passed the quality control criteria under an additive genetic model. Age and gender were used as covariates except for the BRAINS and WTCCC-Ger and WTCCC-UK cohorts because these cohorts utilized the British 1958 Birth Cohort where all individuals were the same age (45 years) at the time of blood collection.

Each participating cohort provided the following summary stats: SNP, risk allele, non-risk allele, minor allele frequency, log (OR), standard errors and p values. Data received from participating cohorts was subjected to further quality control and the following exclusion criteria were used to remove SNPs failing the following filters: (1) Call frequency $\leq$ 95%, (2) MAF $\leq$ 0.01 and (3) HWE $\leq 10^{-6}$, (4) $r^2 < 0.3$, (5) p value = 1 and (6) negative p values.

Details of the statistical analysis are provided in **Table4.2.**

4.2.6 Meta-Analysis of GWAS Studies

A meta-analysis was carried out using a fixed effects model and inverse variance method in METAL software (Willer et al., 2010). A fixed effects model is based on the mathematical assumption that a there is only one true or 'fixed' effect underlying each study and all differences in observed effect is due to sampling error. The model provides greater power but does not take inter-study heterogeneity into account. Evidence of association from the studies was pooled using appropriate weights, which are calculated as the inverse of the square of the standard errors ($1/SE^2$) for every effect estimate (β). The following equation was used to calculate the weighted pooled effects size:

$$\beta = \frac{\sum_i \beta_i \omega_i}{\sum_i \omega_i}$$

Where β = pooled effect size

β_i = effect size estimate for study

ω_i = weight given to study i, equal to inverse of variance ($1/SE_i^2$)

Pooled ORs were calculated as the exponential of the estimated effect size (Exp (β)) for each SNP. 95% confidence intervals were calculated using the following formula:

$$95\% \text{ CI} = \varkappa \pm 1.96 \text{ (SE)}$$

Where $\varkappa$ = point estimate i.e. odd ratios

1.96 = Z score for a 95% CI for a Z distribution

SE = standard error

We also applied the genomic control method (Devlin and Roeder, 1999, Marchini et al., 2004, Bacanu et al., 2000) to control for population heterogeneity. METAL calculates the inflation (λ) of the test statistic by comparing the median test statistic to that expected by chance, and then applies the genomic control correction to the standard error weighted meta-analysis.

Statistical heterogeneity, which is the variation in study outcomes between studies, was also estimated using the Cochran's Q test, which is incorporated in the METAL software (Cochran, 1954). Cochran's Q test is a non-parametric test used to determine whether variables have identical effects or not. Cochran's Q test is calculated as the weighted sum of squared differences between individual study effects and the pooled effect across studies.

$$Q = \sum \omega (E - E_c)^2$$

Where $E_c = \frac{\sum \omega E}{\sum \omega}$

W = weight of the individual study E

E = Effect of the individual study

The Q statistic follows a $\chi 2$ distribution with k-1 degrees of freedom (df) where k is the number of studies included in the meta-analysis. A high value for Q (> k-1) suggests statistical heterogeneity. The significance level of the Q statistic is estimated from the $\chi 2$ distribution with $k-1$ degrees of freedom.

The I^2 index test compliments the chi-squared test and quantifies the extent of variation across studies in a meta-analysis that is due to heterogeneity rather than chance (Higgins and Thompson, 2002). Unlike the Q statistic, the I^2 statistic does not depend on the number of studies included in the meta-analysis. The I^2 statistic was calculated using the following formula:

$$I^2 = \frac{(Q - df)}{Q} \times 100$$

Where $Q = \chi 2$ statistic

df = degrees of freedom equal to number of studies minus 1 (k-1)

A total of 2 million genotyped and imputed SNPs, which passed the quality control criteria, were included in the final meta-analysis from the 7 study cohorts. Bonferroni correction was applied to adjust for multiple testing and the significance threshold was

set at $p \leq 5 \times 10^{-8}$ which represents the Bonferroni corrected conventionally used p value of 0.05.

4.2.7 Visualization of GWAS Data

Probability value plots are widely used visual presentation aids for GWAS data. Two such plots are the $-\log 10$ (p) genome-wide association Manhattan plots and Q-Q plots. Manhattan plots represent the $-\log 10$ (p) values for the entire GWAS dataset in an easily readable format. The $-\log 10$ of the p values are plotted on the Y-axis by chromosome and base pair position on the X-axis. Because of local correlation of the genetic variants, arising from infrequent genetic recombination, groups of significant P values tend to rise up high on the Manhattan plot, making the graph look like a Manhattan skyline (Ehret et al., 2010).

Q-Q plots were also used for visual inspection of the GWAS data. Modest levels of bias within a large GWA study can skew the null distribution and inflate test statistics. A Q-Q plot summarizes the quality of the GWAS data and can reveal underlying population sub-structure and experimental artifacts, which may warrant further analysis. Q-Q plots provide evidence of the extent of deviation (if any) of the observed p values from the null hypothesis (central line). The plots are produced by plotting the observed p values for each SNP (largest to smallest) against expected values from a theoretical $\chi 2$-distribution. There is concordance between the observed and expected p values if all data points lie on or close to the central line. If there is an early upward separation of the expected from the observed, this means that many moderately significant p values are more significant than expected under the null hypothesis.

Regional association and linkage plots allow fine mapping (up to 25 Kbp) of chromosomal regions of interest and are widely used to visualize GWA significant loci. Regional plots for genome wide significant loci are usually centered on the sentinel SNP for each locus. SNPs are plotted with their p values against their genomic position. The most significant SNP can be represented in linkage disequilibrium with other SNPs in the various HapMap panels or original dataset and are shaded according to their pair wise correlation (r^2) with the target SNP.

4.2.8 Power Calculation

The power of a genome wide association study is dependent on many factors such as the magnitude of the effect of a gene variant, its prevalence and sample size. The power of GWAS studies is also dependent not only on the number of cases but importantly on the number of controls recruited. A larger number of controls may compensate for a smaller number of cases and result in similar overall obtained power as if more cases had been recruited.

Power for our meta-analysis was calculated using the CATS genetic power calculator (Skol et al., 2006). The following parameters were used to calculate the power using an additive genetic model: N (cases): 1164, N (controls): 13703, carotid disease in stroke prevalence in Europeans: 0.005 % (White et al., 2005), MAF: 0.1-0.5 and significance level: $\leq 5 \times 10^{-8}$.

4.3 Results

4.3.1 Population Characteristics

The discovery genome-wide analysis of >50% carotid stenosis in ischemic stroke included 14,867 individuals of European ancestry (1164 cases and 13,703 healthy controls) from 7 population-based studies. The studies included individuals with genotype and carotid stenosis data from the following cohorts: BRAINS (n=2527), GAROS 1 (n=3060), GAROS 2 (n=152), GEOS (n=545), ISGS (n=1433), WTCCC-Ger (n=1162) and WTCCC-UK (n=5988) where 'n' is the total number of individuals in each cohort.

BRAINS cohort included 54 genotyped cases with information on >50% carotid stenosis and a mean age of 73.30 ± 12.24 years. Males made up 73 % of the cases. The Mean age for the 2473 British 1958 Births Cohort controls was 45 years.

GAROS 1 cohort included 32 genotyped cases with information on >50% carotid stenosis and a mean age of 68.23 ± 11.37 years. Males made up 69 % of the cases. The mean age for the 3028 healthy controls was 42.99 ± 7.76 of which 76 % were males.

GAROS 2cohort included 15 genotyped cases with information on >50% carotid stenosis and a mean age of 63.39 ± 12.19 years. Males made up 67 % of the cases. The mean age for the 137 healthy controls was 73.02 ± 7.47 of which 50 % were males.

GEOS cohort included 47 cases with >50% carotid stenosis data with a mean age of 44.7 ± 3.7 years and 66% males. Control population included 498 healthy individuals with a mean age of 39.5 ± 6.7 years and 57% males.

ISGS cohort included 76 genotyped cases with >50% carotid stenosis data with a mean age of 70.87 ± 13.41 years and 66% males. 1357 healthy controls consisted of 47% males and a mean age of 64.75 ± 12.63 years.

WTCCC-Germany cohort included 344 genotyped cases with >50% carotid stenosis data with a mean age of 66.00 ± 10.90 years and 70% males. The mean age for the 818 British 1958 Births Cohort controls was 45 years.

WTCCC-UK cohort included 596 genotyped cases with a mean age of 70.0±10.90 years and 63% males. The mean age for the 5392 British 1958 Births Cohort controls was 45 years.

The population demographics for all 7 cohorts are detailed in the **Table 4.3.**

4.3.2 Genotyping, Quality Control and Imputation

The participating cohorts were genotyped on various commercially available genotyping platforms as follows: BRAINS (Illumina HumanHap 610-quad), GAROS 1 (Affymetrix 6.0), GAROS 2 (Illumina HumanHap 610-quad), GEOS (Illumina HumanOmni1-Quad_v1-0_B), ISGS (Illumina HumanHap550K), WTCCC-Ger (Illumina HumanHap 660-quad) and WTCCC-UK (Illumina HumanHap 660-quad).

In order to facilitate a meta-analysis of all 7 cohorts, imputation of the genotyped SNPs was performed for each dataset. Genotypes were inferred for the BRAINS dataset using the HapMap Build 36 Release 22 reference dataset and consisted of 2,259,726SNPs. Genotypes were inferred for the GAROS 1and GAROS 2 datasets using the 1000 Genomes Phase 1 variant reference dataset and consisted of ~37 million SNPs. Genotypes were inferred for the GEOS dataset using the 1000 Genomes Phase 1 variant reference dataset and consisted of 16,219,283SNPs. The ISGS cohort was imputed using the 1000 Genome.2012-03-14 reference dataset and consisted of about 18,639,697 million SNPs. A total of approximately 2 million overlapping SNPs were included in the final meta-analysis. Details of the genotyping and imputation metrics for all 7 cohorts are described in the **Table 4.1.**

Q-Q plots were plotted to assess the quality of data for each GWA study (**Figure 4.2**). The GWAS association plots largely followed the null distribution with results for BRAINS, WTCCC-Ger and WTCCC-UK showing strong deviation of the observed p values from the null hypothesis indicating genome wide significant associations. Q-Q plots for GAROS 1 and GAROS 2 showed evidence of high inflation with skewed distributions indicating a marked difference between the distributions of the observed p values as compared to the null distribution. METAL software has the ability to adjust studies for genomic control in order to control for small amounts of population stratification and unaccounted relatedness (Willer et al., 2010). Prior to meta-analysis METAL calculates

the inflation of the test statistic and applies the genomic control correction to the SE of each study. The genomic inflation factor (λ) for the meta-analysis was low (0.98) suggesting low population stratification and hence no additional adjustment was required.

4.3.3 Meta-Analysis

Meta-analysis of the 7 participating cohorts was performed in METAL (Willer et al., 2010) using a fixed effects models and inverse variance method of weighted beta coefficients. The Q-Q plot of the meta-analysis association results showed strong deviation of the observed p values from the null hypothesis indicating genome wide significant associations of gene polymorphisms with carotid stenosis in ischemic stroke (**Figure 4.3**). We obtained a λ of 0.98 for our meta-analysis suggesting low population stratification and hence no additional adjustment was required.

The meta-analysis results were reported as log (OR) and standard errors, which were converted to odds ratios and 95% confidence intervals. Inter study heterogeneity was assessed for each significant association. All identified associations had significant heterogeneity (Q ≥ degrees of freedom and $p \le 5 \times 10^{-8}$) **(Table 4.4)**.

The *a priori* threshold for genome-wide significance was set at $p \le 5 \times 10^{-8}$, which corresponds to the Bonferroni corrected conventional p value of 0.05 for approximately 2.0 million multiple tests conducted.

4.3.3.1 Genetic loci associated with>50% carotid stenosis in ischemic stroke

Meta-analysis of 7 GWA studies identified three genomic regions in chromosomes 3 to be associated at GWA significance (p value≤5x10^{-8}) with >50% carotid stenosis in ischemic stroke (**Figure 4.4 and Table 4.4**).

LRIG 1 locus

Two correlated inter-genic SNPs (r^2 ≥ 0.60, D=1), rs6788586 and rs6785590, spanning 74 Kb were identified on chromosome 3 (3p14.1) near the human leucine-rich repeats and immunoglobulin-like domain 1 (*LRIG 1*) gene (GENCODE ID: ENSG00000144749)(**Figure 4.5A**). Both SNPslay~207 Kb 5' of *LRIG 1*. SNP rs6788586 was highly correlated with two other functionally non-annotated SNPs, rs61523447 (r^2=1, D=1) and rs79841184 (r^2 =0.91, D=0.97) that lie 207Kb and 202 Kb 5' of LRIG1 respectively. SNP rs6788586 was associated with an OR 0.39 (95% CI 0.13-0.64, p=9.69x10^{-13}). SNP rs6785590 was highly correlated with the functionally non-annotated SNP rs1379184 (r^2=0.94, D=1) that lay 209Kb 5' of *LRIG1*.rs6785590 was associated with an OR 2.20 (95% CI 1.98-2.42, p=2.05x10^{-12}). There were no coding SNPs within any gene in high LD with both SNPs.

ROBO1 locus

Two highly correlated (r^2=1, D=1) intronic SNPs, rs7644521 and rs3923526, spanning 406 Kb were identified on chromosome 3 (3p12.2) within the roundabout, axon guidance receptor, homolog 1 gene (*ROBO 1*)(GENCODE ID: ENSG00000169855) (**Figure 4.5B**). The sentinel SNPs, rs7644521 and rs3923526 were associated with OR 0.66 (95% CI 0.53-0.78, p=3.1x10^{-10}) and OR 1.49 (95% CI 1.36-1.62, p=5.43x10^{-10}) respectively. Both SNPs were in high LD (r^2≥0.8) with a cluster of 9 intronic SNPs spanning 28 Kb. There were no coding SNPs within any gene in high LD with both SNPs.

CAPN7 locus

SNP rs4685240 was identified on chromosome 3 (3p25.1), 5.7 Kb 3' of LOC344875 (GeneID: 344875), 27 Kb 5' of AC090954.1 and 66 Kb 5' of Calpain Like Protease 7 gene (*CAPN7*) (GENCODE ID: ENSG00000131375) **(Figure 4.5C)**. rs4685240 was in high LD (r^2≥ 0.8) with two cluster of SNPs: 14 intronic SNPs 35 Kb-25 Kb 5' of the *CAPN7* gene and 11 SNPs 19 Kb-35 Kb 5' of AC090954.1.SNP rs4685240 was associated with an OR of 0.58 (95% CI 0.39-0.76, p=5.9x10^{-9}).

4.3.3.2 Association with previously published GWAS-derived stroke and carotid disease SNPs

In a cross phenotype comparison, we replicated previously published genome wide significant and suggestive SNPs associated with stroke and carotid disease phenotypes (IMT and plaque) in our dataset (**Table 4.6**). The direction of effect and effect size were consistent for rs17398575 (*PIK3CG*) from a plaque analysis (OR 1.18, 95% CI 1.12–1.23, p=2.3 × 10^{-12}) (Bis et al., 2011) and our study (OR 1.26, 95% CI 1.15-1.37, p=4.97×10^{-5}), although the association was nominal. Stroke associated SNP rs2107595 near gene *HDAC9* (OR 1.39 (95% CI 1.27–1.53, p=2.03 × 10^{-16})(Traylor et al., 2012) was also suggestively associated in our study (OR 1.35 (95% 1.23-1.47, p=7.16 × 10^{-7}). No other GWAS derived SNP showed association in our study. Conversely none of the top SNPs from our analysis were found to be associated with any other vascular phenotypes.

4.4 Discussion

Carotid stenosis is a major risk factor for stroke and the commonly held belief that carotid disease is largely caused by environmental risk factors has recently been challenged by large population based GWA studies that have identified genetic variants to be associated with the risk of carotid stenosis and plaque (Bis et al., 2011). A genome wide association meta-analysis in 14,867 individuals of European ancestry from 7 studies on >50% carotid stenosis in ischemic stroke was performed and identified genome wide significant associations (p≤5x10^{-8}) for three gene loci.

The study is the first to attempt a better understanding of the genetics of superimposed and clinically relevant carotid stenosis in ischemic stroke. Findings from the study extend our current knowledge of advanced carotid disease.

Genetic loci associated with > 50% carotid stenosis in ischemic stroke

LRIG1

The most notable association was for the *LRIG1*gene, which was also the sentinel association in our study. The top SNPs were located 207 Kb 5' of the LRIG1 gene locus which is a member of the human leucine-rich repeats and immunoglobulin-like domains (LRIG) gene family and includes LRIG1, LRIG2 and LRIG3 proteins. LRIG1is a type 1 transmembrane protein with 15 leucine-rich repeats in its extracellular domain and 3 immunoglobulin (Ig)-like domains. LRIG1 negatively regulates EGFR signaling by reducing the phosphorylation of ERBB and resultant decreased activation of downstream receptor tyrosine kinase signaling. The role of LRIG1 in human cancers is well known but complex. For some human cancers such as nasopharyngeal, renal, and breast cancers, the LRIG1 locus (3p14.3) was deleted whereas for others (colorectal, lung, glioblastoma and ovarian) it is neither deleted nor mutated (Wang et al., 2013). It is likely that the occurrence and severity of disease depends on the differential expression of the LRIG1 protein confounded by the tissue type, cancer stage and cancer subtype (Wang et al., 2013). LRIG 1 has also been found to be associated with blood urea nitrogen (Okada et al., 2012) and QRS duration and cardiac ventricular conduction (Sotoodehnia et al., 2010) suggesting a possible role in vascular disease.

Although the role of *LRIG1* in carotid disease is unclear it has been associated with carotid phenotypes in a CAD population (Bis et al., 2011). SNP rs17045031, located 217 Kb 5' of *LRIG1,* showed a suggestive association with carotid artery plaque, with the G allele being associated with both lower odds of plaque (>50% stenosis) and lower risk of coronary artery disease (p=0.04)(Bis et al., 2011). Although the sentinel SNP (rs6788586) was not in high LD with rs17045031, the direction of our results were concordant with their analyses, with the G allele being associated with lower odds of > 50% stenosis (OR 0.39, p=9.69x10^{-13}). The concordance of results is not surprising given the systemic nature of arthrosclerosis and shared risk factors such as HTN, diabetes and old age, between coronary and carotid arthrosclerosis (Jashari et al., 2012). However, SNP rs17045031 was not associated with carotid stenosis in our study population.

The HaploReg V2 software (Ward and Kellis, 2012) was used to explore mechanistic functioning of both SNPs. The software uses LD information from the 1000 Genomes phase 1 project characterize of non-coding SNPs by studying their predicted chromatin state, sequence conservation and their effect on regulatory motifs. SNP rs6788586 was found to effect 17 regulatory protein binding motifs: AIRE_1, AIRE_2, Barhl1, FAC1, Foxa_known1, Foxa_known2, Foxi1, Foxj1_1, Foxj2_1, HDAC2_disc2, Isl2, Nkx2_3, Nkx3_3, RREB-1_1, RREB-1_2, Sox_6 and p300_disc3 and SNP rs6785590 was found to effect two protein binding motifs: SRF_known5 and p300_disc5. An *in silico* regulatory SNP detection analysis using the is-rSNP software (Macintyre et al., 2010) revealed that both SNPs also cause a transcription factor binding sight altering mutation and affects the binding sights of 10 cellular transcription factors at p value ≤ 0.05.

ROBO1

The second genome-wide significant region was at the axon guidance receptor gene *ROBO1* (3p12.2). Two intronic SNPs, rs7644521 and rs3923526 emerged as genome-wide significant with the strongest association at rs7644521. The C allele of rs7644521 was associated with lower odds of >50% stenosis in ischemic stroke (OR 0.66 (95% CI 0.53-0.78, p=3.10x10^{-10}). ROBO1 plays a major role in axonal guidance in neurogenesis and is ubiquitously expressed in various human tissues of the nervous, vascular and immune systems. ROBO1 and DUTT1 are alternative splice variants with different initial exons and initiation codons. Although it's function in vascular diseases

remains unknown, the varied roles ROBO1 protein may play in the body are supported by its associating with risk of many disorders such as developmental dyslexia (Hannula-Jouppi et al., 2005), tumor suppression (Dallol et al., 2002), cervical cancer (Narayan et al., 2006), chronic schizophrenia (Potkin et al., 2009) and autism (Anitha et al., 2008). A recent study examined the association of 144 ROBO1 SNPs with theoretically motivated measures of language impairment, reading and spelling ability (Dyslexia), and short-term verbal information storage and manipulation in a family based study and found the T allele of SNP rs7644521 to be nominally associated with the risk of digits-forward memory span (p=0.0002) and working memory (p=0.046) (Bates et al., 2011). ROBO1 has also been implicated in forebrain development and stress response, a measure of schizophrenic dysfunction (Potkin et al., 2009). A GWAS study on siblings with ischemic stroke identified two ROBO1 intronic SNPs, rs1383407 and rs328049, to be nominally associated with stroke affected siblings (Meschia et al., 2011). However, neither SNPs were significant in our study (p values 0.82 and 0.53 respectively) nor were the SNPs in LD with our top SNP.

Mechanistic characterization using the HaploReg V2 software (Ward and Kellis, 2012) revealed that SNP rs7644521 affects the SP1_known3 regulatory protein binding motifs. *In silico* regulatory SNP detection analysis using the is-rSNP software (Macintyre et al., 2010) revealed that rs7644521 also causes transcription factor binding sight-altering mutation and affects the binding sights of over 138 cellular transcription factors at p value ≤ 0.05.

CAPN7

The third genome-wide significant region was at *CAPN7* (3p24). The G allele of SNP rs4685240 was associated with lower odds of >50% stenosis in ischemic stroke (OR 0.58 (95% CI 0.39-0.76, p=5.89×10^{-9}). CAPN7 is a 92.65 KDa protein comprising of 813 amino acids. The ubiquitous, well-conserved family of calcium-dependent, cysteine proteases and have been implicated in neurodegenerative processes, as their activation can be triggered by calcium influx and oxidative stress. Studies have found CAP7 intronic SNPs rs1318937 and rs10510438 to be associated with alcoholism (Zuo et al., 2012, Zuo et al., 2012) and macular degeneration (dbGaP Study Accession: phs000001.v3.p1) respectively. However, neither SNPs were significant in

our study (p values 0.91 and 0.52 respectively) nor were the SNPs in LD with our sentinel SNP.

How SNPs operate at physiological level is clearly important. Our top SNPs map in or near genes in pathways with established roles in signal transduction (*LRIG1*), cell adhesion and guidance (*ROBO1*) and cysteine proteases (*CAPN7*). Besides their predicted conventional roles, the proteins are implicated in various disorders across the body implicating the involvement of previously unidentified biological processes. We demonstrate that LRIG1, a known risk factor for sub-clinical carotid disease (Bis et al., 2011) and candidate gene at the most significantly associated locus in our study, is also involved in late stage carotid disease. *ROBO1*, a known risk factor for ischemic stroke (Meschia et al., 2011) was also associated with carotid disease supporting the role of a shared genetic risk burden. CAPN7 was a novel find which has not been implicated in any vascular disease so far.

Study Strengths

The strengths of the study include a well-designed analysis plan, population based case-control design, high quality genotyping and in depth phenotyping of study subjects. The analysis was restricted to individuals of European ancestry and corrected for underlying population substructure, therefore ruling out artefactual associations caused by differences in allelic frequencies amongst ethnic groups. The study was also able to replicate results from other carotid disease and stroke GWAS's in the current dataset and replicate findings from the CHARGE consortium GWAS that analyzed carotid disease phenotypes such as IMT and plaque. This replication also allowed us to relate the findings to carotid disease in a large independent healthy population thereby providing important additional context to the results. The current study is the first to attempt a better understanding of superimposed carotid stenosis in ischemic stroke by dissecting genetic risk variants associated with carotid stenosis.

Study Limitations

Lack of Independent Replication and Functional Studies

The current study lacks an independent replication, which is a standard strategy in the current GWAS climate, to confirm identified positive results in the discovery stage of a GWAS. We highlight the difficulty in identifying cohorts with > 50% carotid stenosis measurements in an ischemic stroke population with genotyped data, in which to replicate the discovery results. This is not unprecedented as other GWAS groups who are also dealing with unusual phenotypes or phenotypes not often measured, such as cervical dissection in stroke and venous thrombosis, have found it difficult to bring together large study populations for replication. We are currently in the process of recruiting more cohorts for our discovery and replication stages.

The understanding of how SNPs behave at a physiological level to cause disease is ultimately the goal of any GWA study. Differential expression of top SNPs from a GWAS need to be followed up with expression and translational studies in cell lines and animal models to validate the findings. The current study lacked functional studies to validate the top hits.

Small Sample Size and Low Power

The study consisted of 14,867 individuals (1164 ischemic stroke cases and 13,703 healthy controls), which provided very low statistical power to detect modest effect sizes exerted by genetic variants (OR 1.1-1.4). For MAF's ranging from 0.01-0.5 a sample size of >10,000 cases and controls were required (**Table 4.7**). Complex disorders such as carotid stenosis and stroke are likely to have a polygenic inheritance with small individual genetic effect sizes coming together to produce a larger effect. Low statistical power can also result in inflation of effect size. Our study cohorts were small and much fewer in numbers leading to less robust conclusions compared to other GWA studies on carotid phenotypes (Bis et al., 2011). Our discovery result may indicate a founder's effect, which is confined to the population isolate within the Europeans (Peltonen et al., 2000). A replication analysis in a large well-powered study is required to validate our findings. There was limited funding for genome-wide genotyping, therefore only 500 of the BRAINS resource have been genotyped so far. Individuals recruited in the UK and genotyped in the year 2010 were included in the final analysis. Additional genotyping of

~ 1000 BRAINS samples will be performed in the future, giving the potential for inclusion in the validation study.

Case-Control Design

Our GWA study is a case-control design, which compares cases (ischemic stroke patients with > 50% stenosis) to healthy controls, with the assumption that cases have a higher prevalence of risk variants for the phenotypic trait as compared to the controls. The recruitment of study subjects from a hospital based patient population, though common practice is indicative of biased sampling and introduces selection bias via the unconscious inclusion of subjects that are not representative of the general population.

Lack of Stroke Sub-Type Data

Large GWAS studies have highlighted the stroke subtype specific nature of genetic risk variants (Holliday et al., 2012, Traylor et al., 2012). Our study lacked information on stroke TOAST classification that hampered sub-group analysis, an essential requirement for any stroke study. Due to the heterogeneous etiology of ischemic stroke sub-types it is possible that a small effect lies within a sub-group and is undetectable in a pooled analysis.

Experimental Heterogeneity and Lack of Data on Carotid Imaging Site

Our study showed evidence of heterogeneity, which may be due to disparity in measurement techniques. Ultrasound protocols varied across the participating cohorts and most did not have detailed phenotyping information on the site of imaging for the carotid artery. Studies have reported differences in relationships of risk factors and disease incidence with IMT measured at different carotid sites (Manolio et al., 2004). This suggests site-specific differences in their cause and possible genetic differences as well. For example, the turbulent flow of the internal carotid artery, which is associated with lipid accumulation and plaque hemorrhage, burdens a greater risk of disease as compared to IMT in the common carotid, which has a laminar flow (O'Leary et al., 1999). However, internal cIMT measurements are more challenging to obtain and most of the participating cohorts in our study did not have detailed phenotyping information on the site of imaging for the carotid artery. Low study participant numbers also prevented us from conducting a detailed analysis by site-specific stenosis. Such experimental heterogeneity introduced due to disparity in measurement techniques

may have compromised our ability to detect small associations. It is clear that future studies must use a standardized protocol for measurement of stenosis and possibly a higher resolution technique, such as MRI.

Poor Coverage by Genotyping Arrays and Lack of Causality

All our cohorts used commercially available Illumina genotyping platforms, which typically have poor coverage of rare variants, which may play an important causative role in carotid disease. Although imputation based on LD pattern using the HapMap and 1000 Genomes reference datasets helped populate the dataset to several million SNPs, these may have missed rare functional variants. Additionally a mix of imputation reference data sets used by the participating cohorts allowed us to meta-analyze approximately 2 million SNPs, which cover only a small proportion of SNPs available through 1000 Genomes (Abecasis et al., 2012). Future projects should ideally meta-analyze cohorts which have imputed data from 1000 Genomes reference datasets only.

The three genetic loci identified in our study were either intergenic or intronic SNPs, which were not in high LD with any non-synonymous SNPs. It is likely that the causal variant differs from the identified SNPs and the functional region may lie away from the implicated candidate gene (Kleinjan and van Heyningen, 2005).

Conclusion

Results from the current study identified three genetic loci associated at genome wide significance with >50% carotid stenosis in ischemic stroke. The most strongly associated SNP in the analysis lies close to LRIG1, which has also been identified as a risk variant for carotid disease in a CAD population. Findings lend support to a common genetic risk burden for vascular diseases.

Discovery of genetic risk variants associated with complex disorders such as carotid disease and ischemic stroke may lead to an in-depth understanding of molecular pathways underlining them and possibly reveal new drug targets for therapy. Innovations in genotyping technology, imaging modalities and statistical computation will eventually lead to well powered studies that are strongly positioned to provide specific pathophysiological gene targets. Findings from the study warrant further validation in large well-powered replication studies and need to be followed up with functional analysis in cell lines and animal models.

4.5 Figures and Tables

Figure 4.1: Flow diagram of carotid stenosis-ischemic stroke GWAS analyses
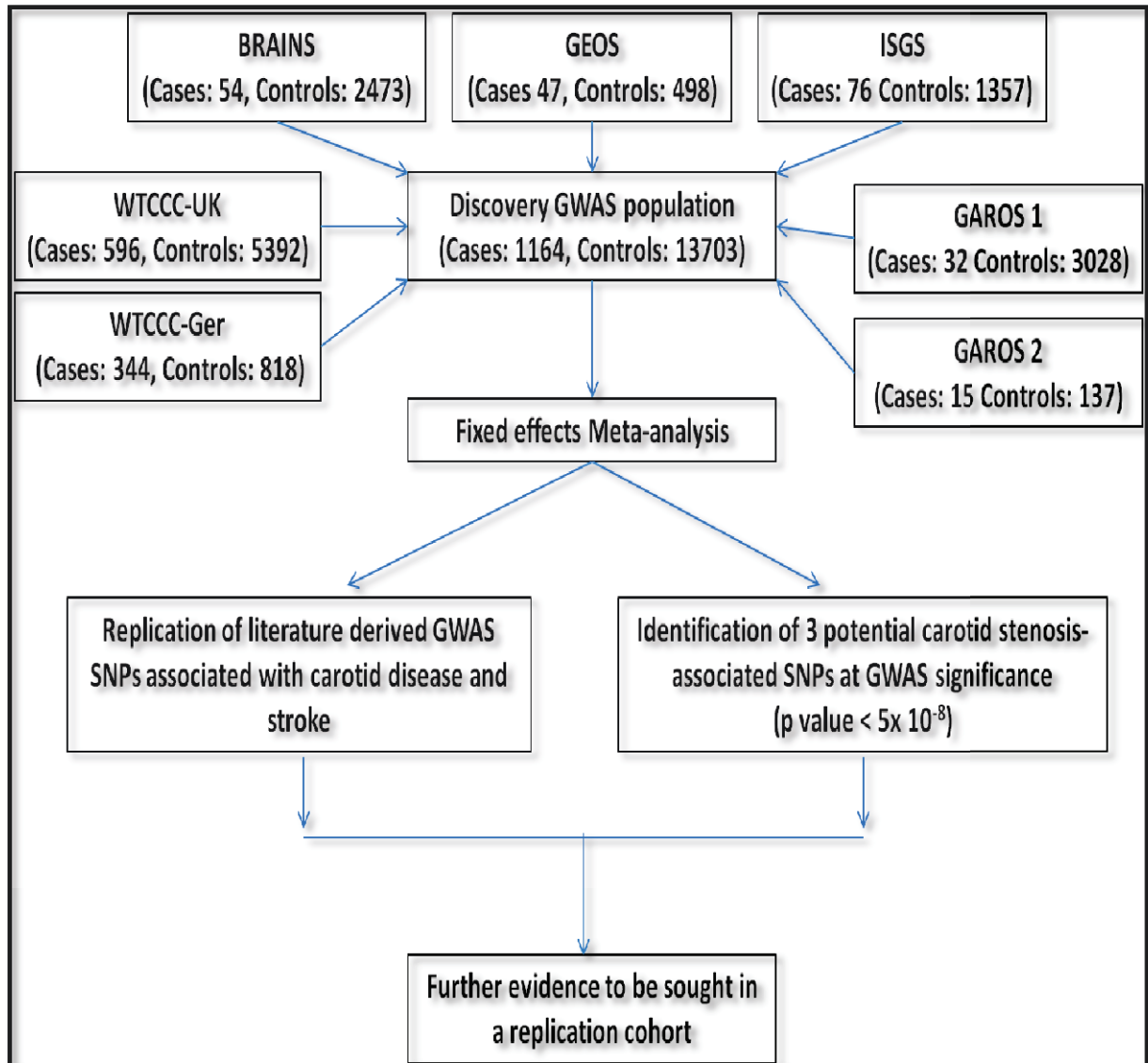
* Cases (total number of ischemic stroke cases with > 50% carotid stenosis), Controls (total number of healthy individuals)

Figure 4.2: Q-Q plot for association results from the (A) BRAINS, (B) GAROS 1, (C) GAROS 2, (D) GEOS, (E) ISGS, (F) WTCCC-Ger and (G) WTCCC-UK. −log10 (p) observed p values are plotted against the −log10 (p) expected p values. Upward deviations of the observed p values (black circles) towards the Y-axis, at the extreme right end of the plot indicate that they are more significant than the expected p values (red line) under the null hypothesis. Strong deviations from the null hypothesis suggest significantly associated genetic loci.

A. BRAINS (λ=0.98)

B. GAROS 1 (λ=1.1)

C. GAROS 2 (λ=0.87)

D. GEOS (λ=0.95)

E. ISGS (λ=0.78)
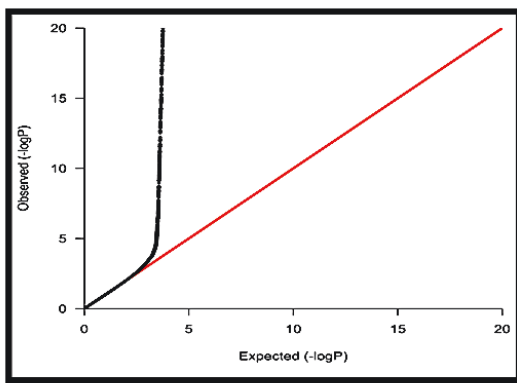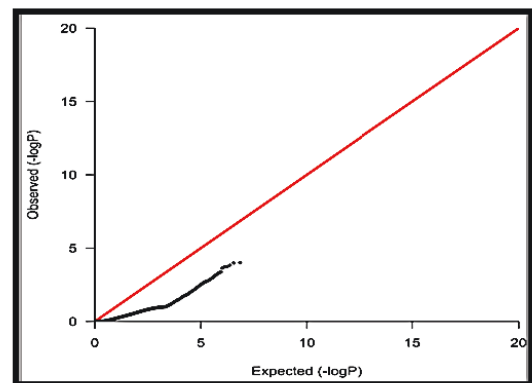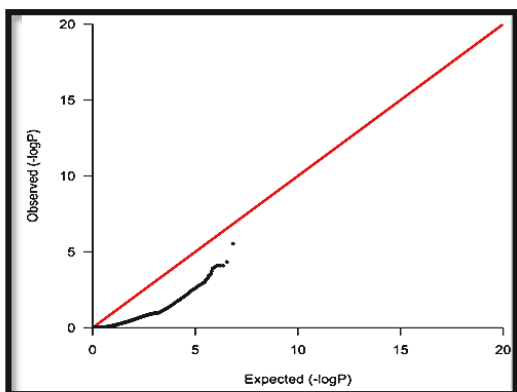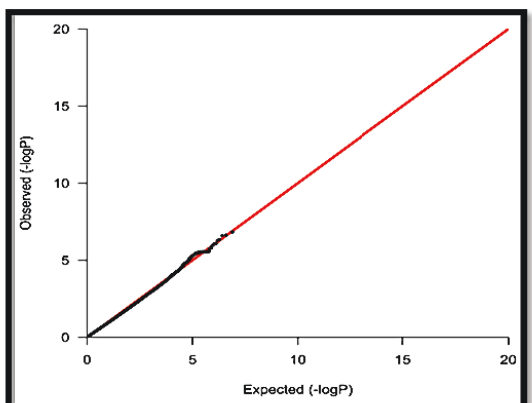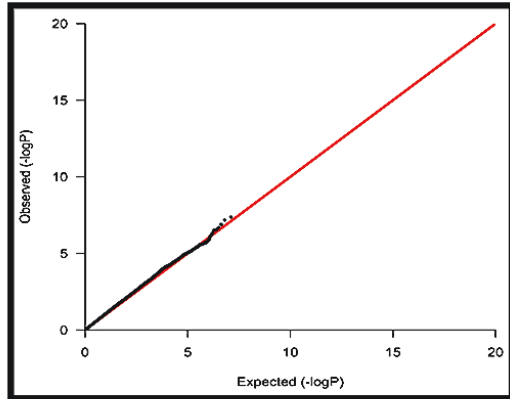
F. WTCCC-Ger (λ=1.005)
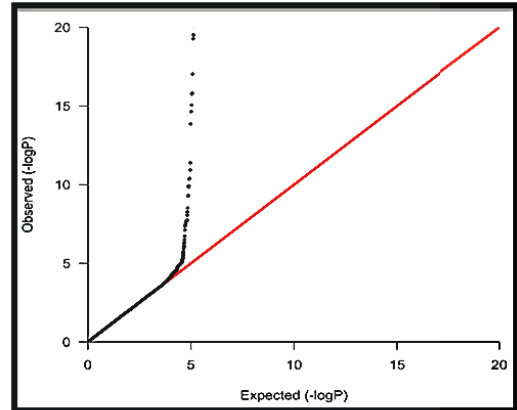
G. WTCCC-UK (λ=1.0007)
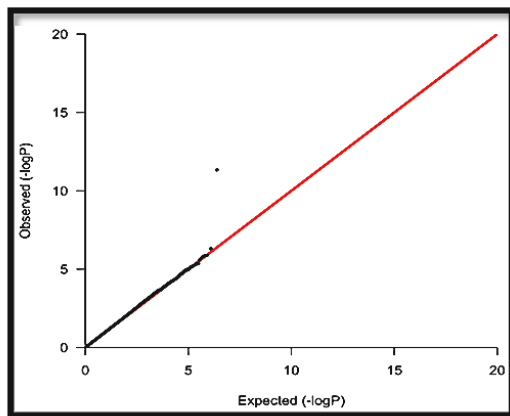
Figure 4.3: Q-Q plot for the meta-analysis association results showing genome wide significant deviations from the null hypothesis suggesting strongly associated genetic polymorphisms with carotid stenosis in ischemic stroke. Genomic inflation factor (λ) for the meta-analysis was 0.98.
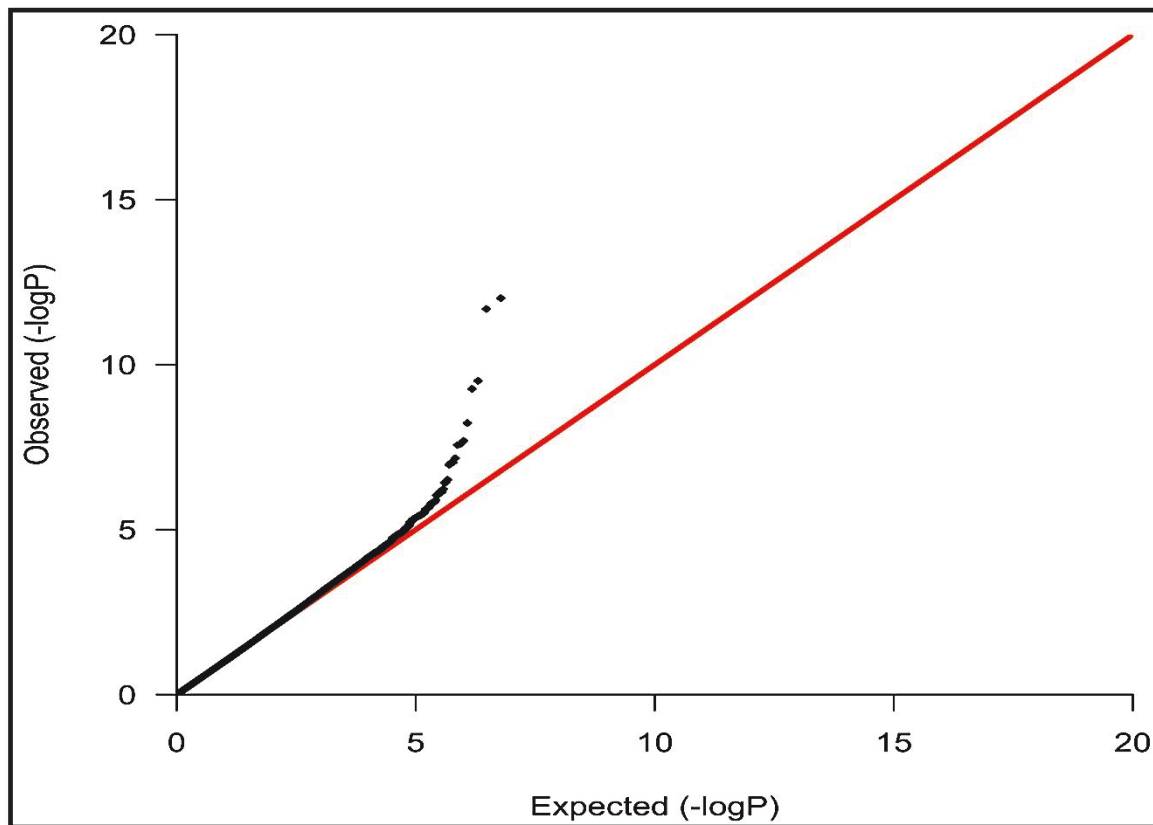
Figure 4.4: Genome-wide Manhattan plot showing three genetic loci in or near genes LRIG1, ROBO1 and CAPN7 associated at GWA significance ($p < 5 \times 10^{-8}$) with >50% carotid stenosis in ischemic stroke. Individual $-\log_{10}$ p values for SNPs are plotted against their genomic position by chromosome. The dotted line at 10^{-6} marks the threshold for promising SNPs and the solid line at 10^{-8} marks the genome-wide significance threshold.
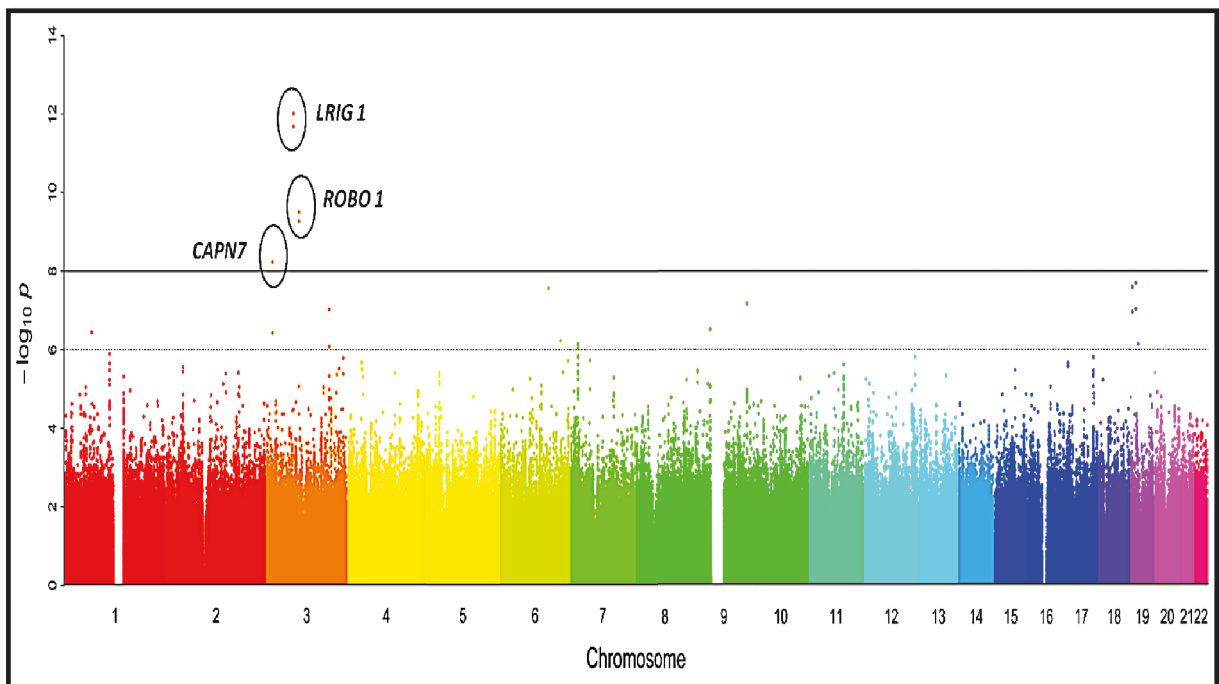
Figure 4.5: Regional LD plots for the three genome wide significant loci associated with >50% carotid stenosis in ischemic stroke. The plots are centered on the sentinel SNP for each locus. SNPs are plotted with their meta-analysis p values against their genomic position. The most significant SNP (orange diamond) is represented in linkage disequilibrium with other SNPs in the HapMap phase 2 CEU panel and are shaded according to their pair wise correlation (r^2) with the target SNP. The blue lines represent the estimated recombination rates. Genes are annotated in green.
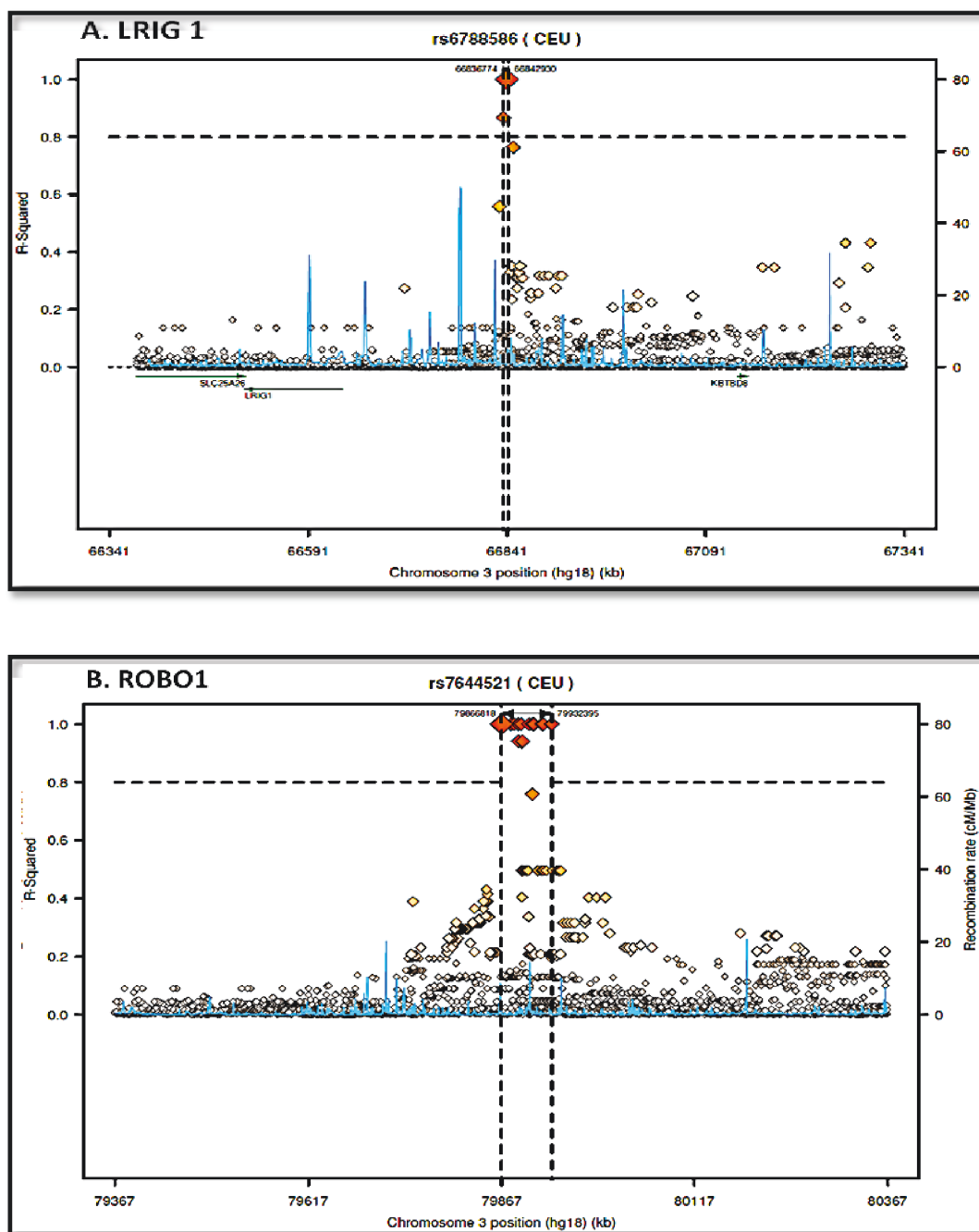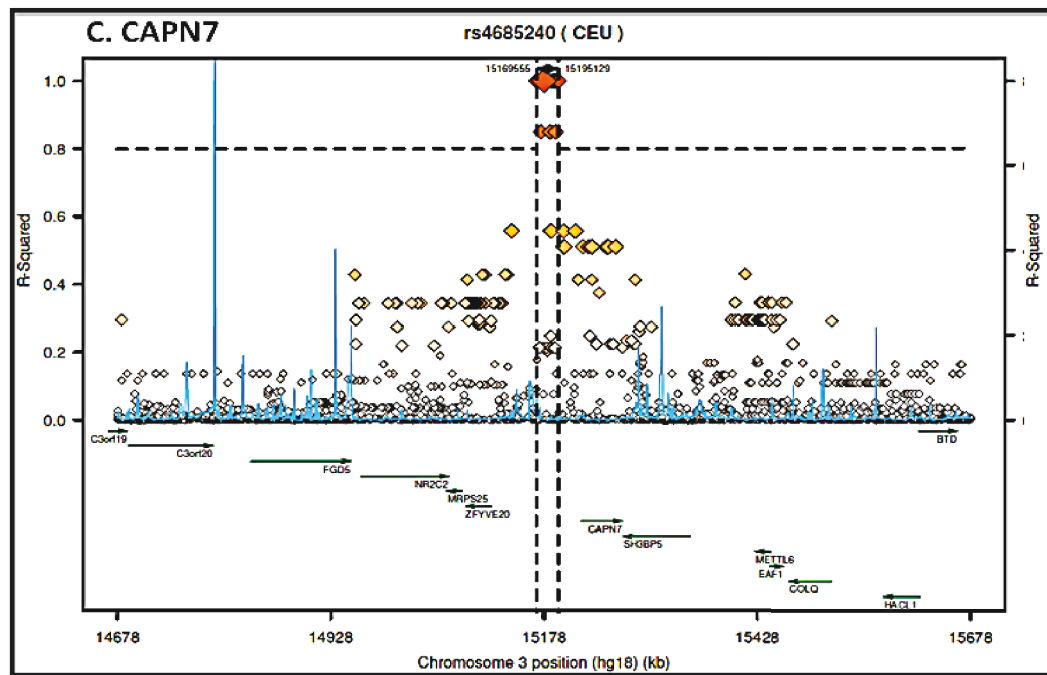
Table 4.1: Details of genotyping and imputation for the carotid stenosis-ischemic stroke cohorts

	BRAINS	GAROS-1	GAROS-2	GEOS	ISGS	WTCCC-Ger	WTCCC-UK
Genotyping Platform	Illumina Human 610 Quad	Affymetrix 6.0	Illumina Human 610 Quad	Illumina HumanOmni1-Quad_v1-0_B	Illumina 550k Human 610 Quad	Illumina Human 610 Quad	Illumina Human 610 Quad
Genotyping Centre	Institute of Child Health, UCL	Broad Institute, Cambridge MA	Broad Institute, Cambridge MA	Centre of Inherited Disease Research, Johns Hopkins Univ	NIA Laboratory of Neurogenetics, UCL Institute of Neurology	Sanger	Sanger
Genotyping calling algorithm	Genomestudio V2010.1 Genotyping module	Birdsuite	GenCall	IlluminaBeadstudio v3.3.7	Genome studio V2010.1 Genotyping module	Gencall	Gencall
Call rate threshold (individuals)	≥ 0.95	≥ 0.90	≥ 0.90	≥ 0.98	≥ 0.95	0.95	0.95
Call frequency threshold (SNPs)	≥ 0.95	≥ 0.95	≥ 0.95	≥ 0.95	≥ 0.95	0.95	0.95
Imputation software	MACH 1.0	IMPUTE v2.2.2	IMPUTE v2.2.2	IMPUTE2	MACH 1.0/miniMac	Impute	Impute

Imputation reference panel	HapMap build 36 release 22	1000 Genomes Phase I integrated variant set NCBI b37, Haplotype release 2012	1000 Genomes Phase I integrated variant set NCBI b37, Haplotype release 2012	June 2011 (phase I, 1094 samples); https://mathgen.stats.ox.ac.uk/impute/data_download_1000G_phase1_interim.html	1000 genomes .2012-03-14	HapMap II 36	HapMap II 36
Pre-imputation MAF	< 0.01	< 0.01	< 0.01	Monomorhpic	< 0.01	< 0.01	< 0.01
Pre-imputation HWE filter	$< 10^{-6}$	$< 10^{-6}$	$< 10^{-6}$	$< 10^{-4}$	$< 10^{-6}$	$< 10^{-6}$	$< 10^{-6}$
Pre-imputation call frequency	< 0.95	< 0.95	< 0.95	< 0.95	< 0.95	< 0.97	< 0.97
SNPs used for imputation (N)	556,847	580,445	522,488	861131	416385	--	--
Quality threshold for imputed SNP	0.3	--	--	0.3	0.3	0.3	0.3
Imputed SNPs for analysis (N)	2259726	37,250,499	37,124,082	16219283	18639697	--	--

Table 4.2: Details of statistical analysis for the carotid stenosis-ischemic stroke cohorts

	BRAINS	GAROS 1	GAROS 2	GEOS	ISGS	WTCCC-Ger	WTCCC-UK
Model	Logistic Regression	Logistic Regression	Logistic Regression	Logistic Regression	Logistic Regression	Logistic Regression	Logistic Regression
Adjustment covariates	Sex	Sex, Age, PC1 & PC2	Sex, Age, PC1 & PC2	Sex, Age, PC1 & PC2	Sex, Age, PC1 & PC2	Sex	Sex
Statistical software	Plink v1.07 , STATA11 & 12, MATLAB, METAL	Plink v1.07 & SNPTEST v2.4.1 64bit static2 version	Plink v1.07 & SNPTEST v2.4.1 64bit static2 version	PLINK v1.07	PLINK v1.07 , MACH , R & MACH2DAT	Plink & METAL	Plink & METAL

Table 4.3: Population demographics for 7 participating cohorts

	BRAINS		GAROS-1		GAROS-2		GEOS		ISGS		WTCCC-UK		WTCCC-Ger	
	Case	Control	Case	Control	Case	Control	Case	Control	Case	Control	Case	Control	Case	Control
N	54	2473	32	3028	15	137	47	498	76	1357	596	5392	344	818
Age in years (mean ± SD)	73.30 (± 12.24)	45	68.23 (± 11.37)	42.99 (± 7.76)	63.39 (± 12.19)	73.02 (± 7.47)	44.7 (± 3.7)	39.5 (± 6.7)	70.87 (± 13.41)	64.75 (± 12.63)	70.00 (± 10.9)	45	66.00 (± 10.9)	45
Male n (%)	42 (73)		68.75	75.83	66.67	49.63	31 (66.0)	282 (56.6)	50 (66)	363 (47)	375 (63)	--	241 (70)	--
IS stroke subtype n (%)														
-Cardioembolic	2 (3.5)	--	9.37	--	13.33	--	0 (0.0)	--	9 (11.84)	--	22 (3.7)	--	0 (0)	
-Large Artery	28 (50)	--	78.12	--	66.67	--	37 (78.7)	--	31 (40.7)	--	478 (80.2)	--	344 (100.0)	--
-Small Vessel	2 (3.5)	--	0	--	0	--	0 (0.0)	--	7 (9.2)	--	6 (1.6)	--	0 (0)	--
HTN n (%)	28 (49)	--	78.12	--	93.33	60.58	23 (48.9)	79 (15.9)	53 (70%)	465 (34.2)	441 (74)	--	251 (73)	--
Diabetes n (%)	7 (12)	--	28.12	--	20	9.49	10 (21.3)	12 (2.4)	63 (82.9)	147 (10.8)	95 (16)	--	79 (23)	--
Hypercholestrimemia n (%)	16 (28)	--	56.25	--	66.67	49.63	19 (41.3)	117 (23.6)	33 (43.42)	120 (8.8)	411 (69)	--	169 (49)	--
Smoking n(%)	13 (23)	--	71.87	--	73.33	54.41	27 (57.5)	117 (23.5)	64 (84.2)	650 (47.9)	459 (77)	--	131 (38)	--

Stenosis													
-50%-70% n (%)	--	--	--	--	--	--	4 (8.5)	--	--	--	--	--	--	--
-70%- occlusion n (%)	--	--	--	--	--	--	43 (91.5)	--	--	--	--	--	--	--

Table 4.4: Genome wide association results identifying three genetic loci associated at GWA significance ($p \leq 5 \times 10^{-8}$) with > 50% carotid stenosis in ischemic stroke, (A) by participating cohorts and (B) meta-analysis results.

Cohort (N)	LRIG (rs6788586)		ROBO 1 (rs7644521)		CAPN7(rs4685240)	
	Log (OR)	SE	Log (OR)	SE	Log (OR)	SE
BRAINS (2527)	0.38	0.39	0.02	0.25	-0.36	0.39
GAROS 1 (3060)	1.23	0.79	0.08	0.42	0.26	0.51
GAROS 2 (153)	-1.42	0.96	-0.44	0.63	0.46	0.75
GEOS (545)	-0.20	0.55	-0.36	0.27	0.76	0.54
ISGS (1433)	0.98	0.69	0.28	0.24	0.12	0.34
WTCCC Ger (1062)	-4.91	0.32	-2.82	0.17	-1.60	0.19
WTCCC UK (5988)	-0.26	0.16	-0.02	0.08	-0.21	0.11

* p values did not reach significance (p<0.05) for any individual cohort

SNP	Chr	BP	Gene	A1/A2	Association		Heterogeneity	
					OR (95% CI)	p	Q (p)	I^2 (%)
rs6788586	3	66758659	LRIG	g/t	0.39 (0.13-0.64)	9.69^{-13}	174.04 (1.0^{-35})	96
rs7644521	3	79784534	ROBO1	c/t	0.66 (0.53-0.78)	3.10^{-10}	224.81 (1.73^{-47})	97.3
rs4685240	3	15203172	CAPN7	g/t	0.58 (0.39-0.76)	5.89^{-09}	39.04 (6.84^{-08})	84.6

*Effect sizes are shown as odds ratios for the % increase or decrease per copy of the risk allele.

**Where more than one SNP at a locus surpassed our *P* value threshold, the SNP with the lowest *P* value is shown.

Table 4.5: Characteristics of SNPs rs6788586, rs7644521 and rs4685240 in the meta-analysis cohorts

	rs6788586			rs7644521			rs4685240		
Cohorts	A1/A2	MAF	I/G	A1/A2	MAF	I/G	A1/A2	MAF	I/G
BRAINS	G/T	0.02	I	C/T	0.17	I	G/T	0.08	I
GAROS 1	G/T	0.05	G	C/T	0.16	G	G/T	0.09	G
GAROS 2	G/T	0.05	I	C/T	0.18	G	G/T	0.09	G
GEOS	G/T	0.04	I	C/T	0.18	I	G/T	0.08	I
ISGS	G/T	0.04	I	C/T	0.18	G	G/T	0.09	I
WTCCC-Ger	G/T	0.16	I	C/T	0.32	I	G/T	0.14	I
WTCCC-UK	G/T	0.04	G	C/T	0.17	G	G/T	0.08	G

Table 4.6: Association of carotid-ischemic stroke cohort with GWAS-derived cIMT (A), plaque (B) and stroke (C) associated SNPs

A. Carotid IMT

SNP	Gene	Chr	N	a1/a2	beta	se	p	a1/a2	beta	se	p
				CHARGE GWAS (Bis et al., 2011)				**Current study**			
rs11781551	ZHX2	8	41,295	a/g	−0.0078	0.0012	2.4×10^{-11}	a/g	0.027	0.0464	0.56
rs445925	APOC1	19	17,185	a/g	−0.0156	0.0028	1.7×10^{-8}	a/g	-0.0836	0.0797	0.29
rs6601530	PINX1	8	32,631	g/a	0.0078	0.0014	1.7×10^{-8}	a/g	0.0017	0.0489	0.97
rs4712972	SLC17A4	6	40,959	g/a	0.0099	0.0018	7.8×10^{-8}	a/g	0.0309	0.0649	0.63

B.Plaque

SNP	Gene	Chr	N	a1/a2	OR (95% CI)	p	a1/a2	OR (95% CI)	p
				CHARGE GWAS (Bis et al., 2011)			**Current study**		
rs17398575	PIK3CG	7	29,255	a/g	1.18 (1.12–1.23)	2.3×10^{-12}	a/g	1.26 (1.15-1.37)	4.97×10^{-05}
rs1878406	EDNRA	4	29,827	t/c	1.22 (1.15–1.29)	6.9×10^{-12}	c/t	0.87 (0.75-1.00)	0.039
rs17045031	LRIG 1	3	27,896	a/g	0.74 (0.66 - 0.83)	4×10^{-7}	a/g	1.19 (0.93- 1.46)	0.18
rs6511720	LDLR	19	22,952	t/g	0.84 (0.78 - 0.89)	1×10^{-7}	g/t	1.16 (1.02-1.31)	0.04

C.Stroke

SNP	Gene	Chr	a1/a2	OR (95% CI)	p	Study	a1/a2	OR (95% CI)	p
			GWAS derived SNPs				**Current study**		
rs2107595	HDAC9	7	a/g	1.39 (1.27–1.53)	2.03×10^{-16}	(Traylor et al., 2012)	a/g	1.35 (1.23-1.47)	7.16×10^{-07}
rs6843082	PITX2	4	g/a	1.36 (1.27–1.47)	2.80×10^{-16}	(Traylor et al., 2012)	a/g	0.92 (0.81-1.03)	0.17
rs879324	ZFHX3	16	a/g	1.25 (1.15–1.35)	2.28×10^{-8}	(Traylor et al., 2012)	a/g	1.06 (0.94-1.19)	0.27
rs556621	6p21.1	6	a/c	1.21 (1.13–1.30)	4.70×10^{-8}	(Holliday et al., 2012)	g/t	0.87 (0.77-0.96)	0.004
rs11833579	NINJ2	12	a/g	1.41 (1.27–1.56)	2.30×10^{-10}	(Ikram et al., 2009)	a/g	0.98 (0.87-1.10)	0.85
rs2200733	PITX2	4	t/c	1.52 (1.35-1.71)	5.82×10^{-12}	(Gretarsdottir et al., 2008)	c/t	0.85 (0.71-0.99)	0.027

Table 4.7: Effect of MAF and cases: control recruitment on achieving 80% power

MAF	Effect size	Cases	Controls
0.01	1.8	8026	>10000
0.02	1.6	6241	>10000
0.05	1.4	5050	>10000
0.1	1.3	4517	>10000
0.2	1.2	6084	>10000
0.3	1.2	4423	>10000
0.4	1.2	3859	>10000
0.5	1.2	3891	>10000

* Results set for achieving 80% power, genome wide significance (p value=5×10^{-8}), prevalence of carotid disease in ischemic stroke=0.005% (White et al., 2005) and an additive genetic model.

CHAPTER 5: DISCUSSION AND FUTURE DIRECTIONS

5.1 Contribution of the Current Study to the Genetics of Ischemic Stroke

Stroke is a devastating disease of the nervous system, which is symptomatic of a diffuse vascular dysfunction. Due to its complex etiology and multi-layered pathology, stroke is considered to be more of a syndrome than a disease. It is likely that stroke has a polygenic component with a battery of individual genes (with small effects) coming together to exert an effect that manifests into disease. Recent large-scale genetic association studies support this hypothesis by demonstrating the sub-type specific nature of risk association and small effect sizes (OR 1.15-1.85) (Traylor et al., 2012, Holliday et al., 2012). The genetic risk associated with stroke can also arise from the genetic risk burden of a risk factor.

This thesis makes several original contributions to the knowledge of ischemic stroke genetics and summarizes the state of knowledge in ischemic stroke genetics particularly in the context of 3 unique confounding risk factors, South Asian ethnicity, long term blood pressure variability and clinically relevant carotid disease (>50% stenosis). In doing so, 3 widely used strategies have been utilized; literature based meta-analysis, candidate gene study and genome wide association study to decipher the genetic risk variants that may be associated with stroke and its risk factors. The study shows that, irrespective of the methodology used, the observed effect sizes for various genetic risk associations were small and broadly similar, further supporting previous findings.

5.1.1 Chapter 1: Introduction

The Introduction section is an exhaustive review of all relevant literature related to the genetic epidemiology of ischemic stroke and has (in-part) been published as a review article (Sharma et al., 2013) with a special focus on the utility of genetic information to clinicians in their everyday clinical practice. We begin with the global impact of stroke on the lives of individuals and economies around the world, which strongly brings home the point that stroke is a devastating disease with huge socio-economic repercussions and needs to be tackled on a global stage. We further delve into the modern day definitions of ischemic stroke, its sub-types, common classification systems and imaging modalities. This is information that is critical in the correct phenotypic characterization

of study participants and determines their eligibility to be included in genetic studies. Next we present the current state of knowledge on the heritability of stroke. Observational and experimental studies are reviewed extensively to establish a strong evidence base for heritability and genetic etiology of stroke. Section 4 presents a comprehensive review of literature on the genetic epidemiology of stroke via its intermediate phenotypes such as ethnicity, blood pressure, carotid disease, diabetes, smoking and obesity. Finally, we examine the various strategies that have been successfully used to gain an insight into the genetic underpinnings of ischemic stroke. Some of these strategies have also been implemented in this thesis.

5.1.2 Chapter 2: Gene polymorphisms associated with ischemic stroke in South Asians: A literature based meta-analysis

The literature-based meta-analysis on the gene polymorphisms associated with ischemic stroke in South Asians was the first and most comprehensive study undertaken to pool underpowered stand-alone candidate gene based genetic studies into a more statistically powered model (Yadav et al., 2013). The study emphasized the paucity of research being conducted in South Asians and the resultant lack of genetic studies on stroke. This is despite the fact that South Asians constitute ~20% of the world's population and are projected to shoulder 80% of the world's burden of stroke by the year 2050.

Although this study points towards a similar genetic risk burden of stroke for all examined ethnic groups, the results remain open to debate since the analyzed dataset was very small and the results were based on pooled summary statistics from different studies rather than individual patient level data. Similar limitations further plague the analysis of homocystiene levels in South Asians that demonstrated a significant difference in the levels of the intermediate phenotype with healthy South Asians carrying the *MTHFR 677CC* genotype having almost double the levels of plasma Hcys as compared to Europeans. Not withstanding the lack of power, the result from the current study show that the South Asian ethnicity places individuals at a greater risk of stroke from birth, partly due to their ethnic make-up and possibly because of gene-environment interaction. Study results tie in well with previously published studies (Holmes et al., 2011). As a future directive large scale comparative analysis of patient

level Hcys data in healthy South Asians and Europeans with *MTHFR 677CC* genotype may provide an explanation for the higher prevalence of stroke in South Asia and opens up avenues for further research. Clinicians may utilize information provided by such studies to test patients for the recessive *MTHFR C677T* genotype, if patients present with high Hcys levels in the absence of risk factors such as low folic acid levels.

Little is known about the molecular mechanisms underlining the role of *MTHFR C677T* polymorphisms and elevated plasma homocysteine levels with the risk of stroke, although both markers have been widely reported in small hospital based genetic studies. The results from such small-scale genetic studies are in sharp contrast to large randomized control trials testing the use of B vitamins and folic acid in reducing homocysteine levels and stroke risk. A recent meta-analysis of 18 large RCT's reporting data on 57,143 individuals and 2,555 stroke events showed that B-vitamin supplementation was not associated with a lower risk of stroke (Zhang et al., 2013). Although results from large studies are considered more reliable, we can argue that since clinical trials generally last between 3-5 years, the long term effects of elevated plasma homocysteine on cerebrovascular disease may not be fully captured by these trials. Stroke is an age related disorder and most stroke patients manifest the disease in their 60's and 70's. The long-term effects of *MTHFR C677T*polymorphism and elevated plasma homocysteine levels in stroke can only be understood by long-term prospective studies, with genetic data, spanning several decades.

5.1.3 Chapter 3: Genome wide analysis of blood pressure variability in ischemic stroke

The candidate gene study examining long term BP variability in ischemic stroke is the first to report any attempt at dissecting the genetics of BP variability and ischemic stroke (Yadav et al., 2013).

The ASCOT UK-IR GWAS was a 'case' only cohort with limited funding for genome-wide genotyping, therefore only 50% (2000 individuals) of the ASCOT resource were genotyped in the initial discovery GWAS using the Illumina CAReiSelect array. Individuals recruited in the UK and Ireland were chosen for genotyping as many of these individuals were also participating in ASCOT sub-studies, and additional measurements were being taken, giving the potential in the future for inclusion in other

genetic studies. The recruitment criteria and population demographics of the UK and Irish ASCOT participants did not differ from the rest of the ASCOT study (Dahlof et al., 2005). An analysis of blood pressure phenotypes (systolic BP, diastolic BP, mean arterial pressure, pulse pressure and hypertension) was performed including controls from the national blood donors and the Allied Irish bank normotensive controls. Unfortunately the *NLGN1* SNP or a close proxy was not genotyped in this panel and the association could not be tested. *NLGN1*SNP rs976683 was looked up in the International Consortium for Blood Pressure (ICBP et al., 2011) GWAS result file but was not significantly associated with both SBP (P value (GC) SBP = 0.194) and DBP (P value (GC) DBP = 0.065).

In the absence of a suitable replication population with long-term blood pressure variability data, the initial replication of the ASCOT UK-IR discovery GWAS findings was conducted in an independent ischemic stroke population comprising of 8,624 stroke cases and 12,722 controls from 7 different cohorts. This is not unprecedented as other GWAS groups who are also dealing with unusual phenotypes or phenotypes not often measured, such as cervical dissection in stroke and venous thrombosis, have found it difficult to bring together large study populations for replication. An alternative strategy is to replicate in a downstream phenotype, which is a common exploratory approach, used to study candidate genes that maybe associated with different vascular disorders such as MI and stroke through their effect on shared risk factors such as hypertension, diabetes and smoking (Cheng et al., 2012). The study had sufficient power to detect a modest effect size of between 1.1-1.4. However, given the small effect size of the top SNP on blood pressure variability, a smaller effect size on stroke may well have been missed despite using nearly 15,000 subjects. Although we were unable to demonstrate an association between the BP variability SNPs and ischemic stroke, we have argued strongly that based on our initial findings, future BP studies should include long-term visit-to-visit BP variability. It neither means that BP variability has no influence on ischemic stroke in a physiological manner nor does it exclude other genetic associations on it. It simply means that the identified SNPs do not play a major role in the genetic etiology of ischemic stroke in our population. There is a possibility that the form of blood pressure variability that the *NLGN1* is presumed to affect, may have no effect whatsoever on stroke risk. In order to decipher the true relationship between blood

pressure variability and ischemic stroke, prospective studies in healthy cohorts with long term BP variability data are required.

To follow up further the original association of SNP rs976638 with BP variability the SNP was genotyped in the ASCOT Scandinavian (DK-FI-NO-SE) population consisting of 3900 individuals, in whom vBP was measured at the same time points as the UK and Irish ASCOT individuals. This was the only other potential replicable population that was available. However the analysis did not provide support for association (p=0.18). Failure to replicate this association could be due to population stratification induced by Anglo-Scandinavian differences such as admixture of Finnish and central European ancestry (Lao et al., 2008) and recruitment of the 'ASCOT-SE' samples in Sweden. Further, the genetic effect could be confined to specific sub-populations of smokers, alcohol consumers and furosemide-exposed individuals. The negative replication results observed in the ASCOT GWAS, possibly due to low power and phenotypic heterogeneity, do not negate the initial findings of the ASCOT UK-IR GWAS. It is possible that the lack of association may reflect environmental heterogeneity between Scandinavian and the UK-IR cohorts, and/or ascertainment differences, with UK-IR cohort primarily being a hospital based recruitment and Scandinavian cohort being more focused on GP based recruitment. Therefore, although not an ideal resource for follow up of the original observation with BP variability it was the only available resource. Its strength was that these individuals were selected using identical recruitment criteria as the UK and Irish ASCOT individuals and BP measurements were taken at the same time points allowing identical analysis of BP variability.

The work conducted in this chapter was the first to report any attempt at dissecting the genetics of BP variability and ischemic stroke. The study puts a flag in the ground for this unusual yet potentially important phenotype and strongly argues that based on the initial findings future BP studies should include long-term visit-to-visit BP variability.

5.1.4 Chapter 4: Meta-analysis of genome wide association studies on carotid stenosis in ischemic stroke

Carotid stenosis is associated with high risk of ischemic stroke accounting for nearly 20% of all ischemic strokes and TIA's caused by emboli originating from unstable plaque (>50% stenosis) in large arteries. Several cellular and molecular pathways underpin carotid disease and stroke, pointing towards a polygenic etiology. Studies have examined both phenotypes in independent populations identifying risk variants unique to each; however no study has examined both phenotypes together. A meta-analysis of 7 genome wide association studies was conducted to increase the statistical power of identifying gene variants associated with the risk of carotid disease in ischemic stroke. A total of 14,867 individuals of European ancestry (1164 cases and 13,703 healthy controls) were tested for association using logistic regression and an additive genetic model. The study identified three genomic regions in chromosomes 3 to be associated at genome wide significance ($p \leq 5 \times 10^{-8}$) with >50% carotid stenosis in ischemic stroke: *LRIG1* ($p=9.69 \times 10^{-13}$), *ROBO1* ($p=3.10 \times 10^{-10}$) and *CAPN7* ($p=5.89 \times 10^{-9}$). Evidence was also sought for a shared genetic basis of carotid disease and stroke by analyzing previously published GWAS-derived SNPs associated with stroke and carotid disease. Interestingly, we were able to replicate the association of SNP rs17398575 (*PIK3CG*) from a plaque analysis (OR 1.18, 95% CI 1.12–1.23, $p=2.3 \times 10^{-12}$) (Bis et al., 2011) and stroke associated SNP rs2107595 (*HDAC9)* (OR 1.39, 95% CI 1.27–1.53, $p=2.03 \times 10^{-16}$)(Traylor et al., 2012).

One of the major difficulties of this study was deriving a uniform definition of the carotid stenosis phenotype. Study specific diagnosis was conducted using high-ultrasound B-mode ultrasonography. All participating cohorts collected the data on carotid stenosis retrospectively and the information on the grading method (NASCET vs. ECST) was not available for all the studies. Since both methodologies have significant differences in grading stenosis (NASCET 50% stenosis = ECST 70% stenosis) (Thapar et al., 2013), a quantitative phenotype could not be calculated for a linear regression analysis.

Although the diagnostic accuracy of carotid ultrasound in predicting significant stenosis is high (86% inter sonographer agreement for 70-99% stenosis) (Thapar et al., 2013),

as with any imaging modality it can be prone to measurement error. The carotid stenosis phenotypes is highly quantitative and instrumental heterogeneity can be introduced due to different sonographers handling different machines at various time points. Carotid ultrasound is a technician dependent imaging method which is greatly influenced by the expertise and training of the sonographers. The participating studies in the carotid GWAS were retrospective and conducted many years ago, which may have lead to inaccurate measurements by the sonographer that do not meet the current international guidelines. The diagnosis of carotid stenosis may not have been blinded and conducted by multiple individuals leading to bias and interobserver differences. As a future goal, genetic studies on carotid phenotypes must device a diagnostic standard operating protocol, which is uniformly followed by all participating cohorts in an effort to reduce measurement inaccuracies.

The presence of superimposed carotid disease in ischemic stroke is indicative of a severe disease burden that may not only account for an increased genetic risk but could also alter the functioning of the risk variants. Findings from the current study suggest that individuals with stroke and superimposed carotid disease may carry a greater combined genetic burden than individuals with either stand-alone disease. Conversely none of the top SNPs from the analysis were found to be associated with any other vascular phenotypes suggesting that these are novel candidate genes for stroke and advanced carotid disease. Study results need to be validated in a suitable replication cohort and refined by fine mapping of the suggested genomic region. Functional studies are also required to establish causality.

5.1.5 Summary of Main Results

❖ Genetic risk associations for common variants of genes *PDE4D, ACE I/D* and *IL10* with ischemic stroke in South Asians were established. Although the study data set was small in comparison to European studies, the observed effect sizes were similar to those in other ethnicities.

❖ High Hcys levels in healthy South Asians with *MTHFR TT* genotype as compared to healthy Europeans carrying the same genotype were observed. This finding suggests that South Asians may carry an inherent genetic liability to stroke risk via elevated plasma Hcys levels. To confirm this observation, future studies need to examine individual patient level data taking into consideration folic acid, Vitamin B6 and Vitamin 12 levels.

❖ Mendelian randomization established a causal relationship between *MTHFR C677T* and ischemic stroke with similar results as Europeans (Bentley et al., 2010). This lends support to the increasing demand for evidence supporting causality of risk variants via their intermediate biomarkers. Mendelian randomization using individual patient level data and *MTHFR* as an instrument variable is required to validate this finding.

❖ A novel gene candidate (*NLGN1*) was associated with the increased risk of blood pressure variability. The inability to replicate the association with ischemic stroke may be due to heterogeneity and low power, and therefore future studies require a larger replication cohort.

❖ GWAS for carotid stenosis in ischemic stroke identified genes previously associated with CAD-cIMT (*LRIG1*) (Bis et al., 2011) and ischemic stroke (*ROBO1*)(Meschia et al., 2011) lending support to our finding being true associations. We also identified a novel gene candidate (*CAPN7*) for carotid stenosis-ischemic stroke, which has not been implicated as disease causing in any vascular disease.

5.2 Future Directions

Considerable amounts of treasure and time has been spent on studying the genetics of stroke and most studies point towards common genetic risk variants carrying small to modest risk and accounting for only a small percentage of the heritability. Clinically, the results from the current stroke studies may still be in their infancy, in terms of inclusion value in patient care. It is clear that for a true understanding of the genetics of stroke, researchers will need to extend their efforts beyond the current strategies.

5.2.1 Studies in Prospective Healthy Cohorts

Most studies on stroke include individuals at two extremes of the trait distribution i.e. stroke patients and healthy individuals. Although this is an acceptable approach and has successfully identified risk variants, a true understanding of the genetic burden of stroke lies in studying large prospective longitudinal studies following healthy cohorts at moderate risk for stroke occurrence. Studies have questioned the generalizability of stroke patients recruited into studies and recommended careful consideration of various prognostic factors before grouping patients into a single study (Busija et al., 2013). International guidelines for clinical monitoring of stroke, standardized sub-typing based on pathophysiology and calibration of biochemical tests are required to be followed uniformly by all participating cohorts in future genetic studies.

5.2.2 Detailed Phenotyping of Stroke Sub-Types

Ample evidence has been presented to establish the sub-type specific nature of genetic risk of stroke. This is not surprising as stroke sub-types have very different etiologies and genetic variants underlining these pathophysiological mechanisms are bound to differ. Most genetic studies in stroke have reported the lack of proper sub-type classification as a major limitation and cause of low power in the analysis of stroke sub-types. Future studies must ensure detailed phenotyping of stroke sub-types, using a uniform classification system, in order to minimize heterogeneity and provide sufficient statistical power for the analysis.

5.2.3 Stroke GWAS in South Asians

Although studies have shown the existence of genetic differences i.e. varying allele frequencies, between different ethnic groups, the magnitude of effect broadly remains the same (Ioannidis et al., 2004). However, this may be far from the truth since an overwhelming number of studies have been conducted on European populations (Traylor et al., 2012, Holliday et al., 2012) and so far no GWAS studies of comparable size or power have been conducted specifically in South Asian stroke patients. Just as with individuals of African ancestry (Frazer et al., 2007), genetic studies in South Asians may reveal new variations, which are absent in European populations. Future goals should include replication in South Asians of findings of published GWA studies and conducting a GWAS in this population. Established and highly phenotyped data banks for South Asian stroke patients (and controls) are well on their way to begin the first stage of genetic association studies (Yadav et al., 2011, Cotlarciuc I, 2012).

5.2.4 Missing Heritability of Stroke

Most genetic studies conducted on stroke to date have been on common variants with a prevalence of >5% of the general population and relatively small to modest effect sizes (<1.5). Although the best heritability estimates come from family studies, estimates for stroke have been calculated from population based GWA studies using experimentally determined common variants. The heritability varies amongst stroke subtypes; 37.9% for all ischemic stroke, 40.3% for large-vessel disease, 32.6% for cardio embolic and 16.1% for small-vessel disease (Bevan et al., 2012), partly due to differences in the underlying pathology of each sub type. These estimates are most likely inflated since they excluded the heritability that could be attributed to rare low frequency variants (MAF < 0.1%) and non-SNP variations of the genome such as CNVs (Zhang et al., 2010). Further evidence towards disproving these heritability estimates lie in the fact that almost 90% of stroke heritability has been assigned to ten modifiable risk factors (O'Donnell et al., 2010a). This raises the question as to whether stroke risk can be greatly reduced just by effectively controlling these risk factors.

Given the high prevalence of stroke, natural selection may have played a role in selecting small gene effects over large ones, which in evolutionary terms, would be beneficial for the survival of the human race (Pritchard and Cox, 2002). Different selection pressures such as changes in environmental factors, infections and modulations of the immune system may have also resulted in the disappearance of large effect producing common variants of stroke (Manolio et al., 2009).

Current heritability estimates on stroke don't take gene-environment and gene-gene effects into consideration and highlight our limited understanding of this complex interplay of factors. For a true measure of heritability, whole genome next generation sequencing of stroke patients may be able to provide an answer as it would take into account all genomic variations. Deep sequencing of the human genome may be able to identify rare variants that are hidden within already know genes associated with stroke risk (Manolio et al., 2009) and validate or disprove the actual risk reported for these genes. The 1000 Genomes project is a human genome reference dataset that is in the process of deep sequencing (up to 4X coverage) variants that occur in at least 1% of the population. Once all phases of the project are complete, the reference data set will allow researchers to populate their genotyped data sets to several million rare and common markers allowing for the detection of variants with low allele frequencies.

5.2.5 Identification of Novel Stroke Biomarkers

Although the knowledge of underlying genetic mechanism gives researchers in-depth understanding of the mechanistic functioning's of stroke, there is a need to translate the results to a biological level. Blood biomarkers are biomolecules that are released into the blood stream in response to neural tissue damage. Alternately they may be produced prior to a stroke and could play a role in causing stroke. Knowledge of reliable blood biomarkers for stroke would make it possible to develop blood tests for stroke and TIA's to guide treatment and ultimately improve outcomes. They may also help to improve inaccurate prognostic models for stroke.

Several studies have examined blood biomarkers for stroke (Whiteley et al., 2009, Montaner, 2009, Jensen et al., 2009) with little potential to be successfully used in everyday clinical practice. A large literature based meta-analysis by Hasan et al identified 3 known biomarkers (C-reactive protein, P-selectin and Hcys) to be

marginally differentiated between ischemic stroke and healthy control subjects (Hasan et al., 2012). The study demonstrated the wide spread utility of biomarkers and its importance in diagnosis and prognosis of stroke. With the identification of new stroke genetic risk variants, there is a possibility of identifying new biomarkers, which may help in the improved prognosis of stroke.

Another recently published study examined GluN1-S2 serum antibodies produced in response to neuronal tissue damage after stroke. Interestingly the study showed that the antibodies were generated after cerebral injury and the levels of antibodies produced were proportional to the infarct size. As biomarkers, these antibodies could reveal important information about the presence and severity of the stroke (Kalev-Zylinska et al., 2013).

Research on stroke blood biomarkers promises an exciting opportunity to decipher the molecular pathophysiology of stroke at a biological level. Individuals may be tested for stroke specific blood biomarkers, which could help in improving the prognosis and diagnosis of stroke. Although, currently there are no specific biomarkers for stroke, successful identification of biomarkers for other vascular disorders suggests that this may be possible for stroke as well (Musialek et al., 2013).

5.3 Conclusion

Developments in genotyping technology and statistical analysis methods have provided substantial evidence in support of the genetic burden of ischemic stroke and its risk factors. This thesis utilized new methodologies to lend further support to the current evidence base and the findings broadly reflect results similar to previous estimates of the genetic liability of stroke. Novel insights into the genetics of South Asian ethnicity, long term blood pressure variability and symptomatic carotid disease have been gained through this study.

REFERENCES

ABECASIS, G. R., et al. 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature,* 491**,** 56-65.

ADAMS, H. P., JR., et al. 1993. Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. *Stroke,* 24**,** 35-41.

ADIB-SAMII, P., et al. 2013. 17q25 Locus is associated with white matter hyperintensity volume in ischemic stroke, but not with lacunar stroke status. *Stroke,* 44**,** 1609-15.

ALBIN, B., et al. 2013. Comparison of Stroke Mortality in Finnish-Born Migrants Living in Sweden 1970-1999 and in Swedish-Born Individuals. *J Immigr Minor Health.* (Epub ahead of publication)

ALLURI, R.V., et al. 2005. MTHFR C677T gene mutation as a risk factor for arterial stroke: a hospital based study. *Eur J of Neurology,* 12**,** 40-44.

AMARENCO, P., et al. 2009. New approach to stroke subtyping: the A-S-C-O (phenotypic) classification of stroke. *Cerebrovasc Dis,* 27**,** 502-8.

ANDERSON, C. D., et al. 2013. Common variants within oxidative phosphorylation genes influence risk of ischemic stroke &intracerebral hemorrhage. *Stroke,* 44**,** 612-9.

ANDERSON, C. D., et al. 2011. Common mitochondrial sequence variants in ischemic stroke. *Ann Neurol,* 69**,** 471-80.

ANITHA, A., et al. 2008. Genetic analyses of roundabout (ROBO) axon guidance receptors in autism. *Am J Med Genet B Neuropsychiatr Genet,* 147B**,** 1019-27.

ARIMA, K., et al. 2003. Cerebral arterial pathology of CADASIL and CARASIL (Maeda syndrome). *Neuropathology,* 23**,** 327-34.

ARIYARATNAM, R., et al. 2007. Genetics of ischaemic stroke among persons of non-European descent: a meta-analysis of eight genes involving approximately 32,500 individuals. *PloS Med,* 4**,** e131.

ARREGUI, M., et al. 2012. Significant associations of the rs2943634 (2q36.3) genetic polymorphism with adiponectin, high density lipoprotein cholesterol and ischemic stroke. *Gene,* 494**,** 190-5.

ARTAVANIS-TSAKONAS, S., et al. 1999. Notch signaling: cell fate control and signal integration in development. *Science,* 284**,** 770-6.

ASDAGHI, N., et al. 2011. Oxfordshire community stroke project classification poorly differentiates small cortical and subcortical infarcts. *Stroke,* 42**,** 2143-8.

ASTRUP, J., et al. 1981. Thresholds in cerebral ischemia - the ischemic penumbra. *Stroke,* 12**,** 723-5.

AY, H., et al. 2007. A computerized algorithm for etiologic classification of ischemic stroke: the Causative Classification of Stroke System. *Stroke,* 38**,** 2979-84.

BABU, M. S., et al. 2012. Association of genetic variants of fibrinolytic system with stroke and stroke subtypes. *Gene.* 495(1), 76-80.

BACANU, S. A., et al. 2000. The power of genomic control. *Am J Hum Genet,* 66**,** 1933-44.

BAK, S., et al. 2002. Genetic liability in stroke: a long-term follow-up study of Danish twins. *Stroke,* 33**,** 769-74.

BAMFORD, J., et al. 1991. Classification and natural history of clinically identifiable subtypes of cerebral infarction. *Lancet,* 337**,** 1521-6.

BANERJEE, T.K., et al. 2001. Stroke in the urban population of Calcutta: an epidemiological study. *Neuroepidemiology,* 20**,** 201–207.

BANERJEE, I., et al. 2008. Inflammatory system gene polymorphism and the risk of stroke: a case-control study in an Indian population. *Brain Res Bull,* 75**,** 158-165.

BANERJEE, A., et al. 2009. Sex-specific familial clustering of MI in patients with acute coronary syndromes. *Circ Cardiovasc Genet,* 2**,** 98-105.

BANERJEE, T. K. 2006. Epidemiology of stroke in India. *Neurlogy Asia,* 11**,** 1– 4.

BATES, T. C., et al. 2011. Genetic variance in a component of the language acquisition device: ROBO1 polymorphisms associated with phonological buffer deficits. *Behav Genet,* 41**,** 50-7.

BAUDOUIN, S. & SCHEIFFELE, P. 2010. SnapShot: Neuroligin-neurexin complexes. *Cell,* 141**,** 908, 908 e1.

BAWEJA, G., et al. 2004. Prevalence of stroke and associated risk factors in Asian Indians living in the state of Georgia, United States of America. *Am J Cardio,* 93**,** 267-9.

BEILBY, J. P., et al. 2003. Apolipoprotein E gene polymorphisms are associated with carotid plaque formation but not with intima-media wall thickening: results from the Perth Carotid Ultrasound Disease Assessment Study (CUDAS). *Stroke,* 34**,** 869-74.

REFERENCES

BENTLEY, P., et al. 2010. Causal relationship of susceptibility genes to ischemic stroke: comparison to ischemic heart disease and biochemical determinants. *PLoS One,* 5**,** e9136.

BEVAN, S., et al. 2008. Variation in the PDE4D Gene and Ischemic Stroke Risk. A Systematic Review and Metaanalysis on 5200 Cases and 6600 Controls. *Stroke,* 39**,** 1966–1971.

BEVAN, S., et al. 2005. Phosphodiesterase 4D gene, ischemic stroke, and asymptomatic carotid atherosclerosis. *Stroke,* 36**,** 949-53.

BEVAN, S., et al. 2012. Genetic heritability of ischemic stroke and the contribution of previously reported candidate gene and genomewide associations. *Stroke,* 43**,** 3161-7.

BHAUMIK, S. 2013. India launches stroke registry to combat "epidemic". *BMJ*, 346, f223.

BHARGAVA, S., et al. 2012. Lowering Hcys and modifying nutritional status with folic acid and vitamin B(12) in Indian patients of vascular disease. *J Clin Biochem Nutr,* 50**,** 222-6.

BHATTACHARYA, S. 2005. A 5 years prospective study of incidence, morbidity and mortality profile of stroke in a rural community of eastern India. *J Indian Med Assoc,* 103**,** 655– 659.

BIS, J. C, et al. 2011. Meta-analysis of genome-wide association studies from the CHARGE consortium identifies common variants associated with carotid intima media thickness and plaque. *Nat Genet,* 43**,** 940-7.

BISSOLI, L., et al. 2002. Effect of vegetarian diet on Hcys levels. *Ann Nutr Metab,* 46**,** 73-9.

BISWAS, A., et al. 2009. Homocystine levels, polymorphisms and the risk of ischemic stroke in young Asian Indians. *J Stroke Cerebrovasc Dis,* 18**,** 103-10.

BISWAS, A., et al. 2009b. Prothrombotic polymorphisms, mutations, and their association with pediatric non-cardioembolic stroke in Asian-Indian patients. *Ann Hematol,* 88**,** 473-8.

BISWAS, M. 2009. Etiology and risk factors of ischemic stroke in Indian-American patients from a hospital-based registry in New Jersey, USA. *Neurology Asia,* 14**,** 81 – 86.

BISWAS, S., et al. 2008. Bridging the synaptic gap: neuroligins and neurexin I in Apis mellifera. *PLoS One,* 3**,** e3542.

REFERENCES

BISWAS, S. 2013. India unveils plans to improve stroke research and care.*Lancet*, 381, 190.

BLAND, J. M. & ALTMAN, D. G. 2000. Statistics notes. The odds ratio. *BMJ,* 320**,** 1468.

BOTTOS, A., et al. 2009. The synaptic proteins neurexins and neuroligins are widely expressed in the vascular system and contribute to its functions. *Proc Natl Acad Sci U S A,* 106**,** 20782-7.

BOVA, I., et al. 1999. The A677V methylenetetrahydrofolate reductase gene polymorphism and carotid atherosclerosis. *Stroke,* 30**,** 2180-2.

BRASS, L. M., et al. 1992. A study of twins and stroke. *Stroke,* 23**,** 221-3.

BRION, M.A., SHAKHBAZOV, K. & VISSCHER, P.M. 2013. Calculating statistical power in Mendelian randomization studies. *International Journal of Epidemiology*, 42:1497–1501.

BURGESS, S. & THOMPSON, S. G. 2013. Use of allele scores as instrumental variables for Mendelian randomization. *International Journal of Epidemiology*, 42, 1134–1144.

BURGESS, S. & THOMPSON, S. G. 2011. Avoiding bias from weak instruments in Mendelian randomization studies. *Int J Epidemiol,* 40**,** 755-64.

BUSHNELL, C. D. 2008. Stroke and the female brain. *Nat Clin Pract Neurol,* 4**,** 22-33.

BUSIJA, L., et al. 2013. Do patients who take part in stroke research differ from non-participants? Implications for generalizability of results. *Cerebrovasc Dis,* 35**,** 483-91.

CAICOYA, M., CORRALES, C. & RODRIGUEZ, T. 1999. Family history and stroke: a community case-control study in Asturias, Spain. *J Epidemiol Biostat,* 4**,** 313-20.

CALAHORRO, F. & RUIZ-RUBIO, M. 2012. Functional phenotypic rescue of Caenorhabditis elegans neuroligin-deficient mutants by the human and rat NLGN1 genes. *PLoS One,* 7**,** e39277.

CAPLAN, L. R. 2011. Stroke classification: a personal view. *Stroke,* 42**,** S3-6.

CAPLAN, L. R., et al. 2004. New England Medical Center Posterior Circulation registry. *Ann Neurol,* 56**,** 389-98.

CAPPUCCIO, F. P. 1997. Ethnicity and cardiovascular risk: variations in people of African ancestry and South Asian origin. *J Hum Hypertens,* 11**,** 571-6.

CASAS, J.P., et al. 2005. Homocysteine and stroke: evidence on a causal link from mendelian randomisation. *Lancet,* 365**,** 224-232.

REFERENCES

CASAS, J. P., et al. 2004. Meta-analysis of genetic studies in ischemic stroke: thirty-two genes involving approximately 18,000 cases and 58,000 controls. *Arch Neurol,* 61**,** 1652-61.

CHALELA, J. A., et al. 2007. Magnetic resonance imaging and computed tomography in emergency assessment of patients with suspected acute stroke: a prospective comparison. *Lancet,* 369**,** 293-8.

CHANDALIA, M., et al. 2003. HyperHcysmia in Asian Indians living in the United States. *J Clin Endocrinol Metab,* 88**,** 1089-95.

CHANDRA, G., et al. 2006. Cystathionine beta-synthase T833C/844ins68 polymorphism and stroke. *Neurology India,* 54**,** 446.

CHANG, L. T., et al. 2010. Link between interleukin-10 level and outcome after ischemic stroke. *Neuroimmunomodulation,* 17**,** 223-8.

CHATTERJEE, K., FALL, S. & BARER, D. 2010. Mood after stroke: a case control study of biochemical, neuro-imaging and socio-economic risk factors for major depression in stroke survivors. *BMC Neurol,* 10**,** 125.

CHATURVEDI, S., et al. 2005. Carotid endarterectomy--an evidence-based review: report of the Therapeutics and Technology Assessment Subcommittee of the American Academy of Neurology. *Neurology,* 65**,** 794-801.

CHEN, P. H., et al. 2012. Classifying Ischemic Stroke, from TOAST to CISS. *CNS Neurosci Ther,* 18**,** 452-6.

CHENG, Y. C., et al. 2012. Are MI--associated single-nucleotide polymorphisms associated with ischemic stroke? *Stroke,* 43**,** 980-6.

CHENG, Y. C., et al. 2011. Genome-wide association analysis of ischemic stroke in young adults. *G3 (Bethesda),* 1**,** 505-14.

CHEONG, M. Y., et al. 2011. Association of the adiponectin gene variations with risk of ischemic stroke in a Korean population. *Yonsei Med J,* 52**,** 20-5.

CHOWDHURY, A. H., et al. 2001. Apolipoprotein E genetic polymorphism and stroke subtypes in a Bangladeshi hospital-based study. *J Epidemiol,* 11**,** 131-8.

COCHRAN, W. G. 1954. The combination of estimates from different experiments. *Biometrics,* 10**,** 101-129.

COCHRAN, IMS. 2010. Statistical Methods Group of the Cochrane Collaboration.

COLE, J. W., et al. 2012. Rare variants in ischemic stroke: an exome pilot study. *PLoS One,* 7**,** e35591.

COLLABORATION, H. S. 2002. Hcys and risk of ischemic heart disease and stroke: a meta-analysis. *JAMA,* 288**,** 2015-22.

CONDE, L., et al. 2011. Novel associations for coronary artery disease derived from genome wide association studies are not associated with increased carotid intima-media thickness, suggesting they do not act via early atherosclerosis or vessel remodeling. *Atherosclerosis,* 219**,** 684-9.

CONGRAINS, A., et al. 2012. Genetic variants at the 9p21 locus contribute to atherosclerosis through modulation of ANRIL and CDKN2A/B. *Atherosclerosis,* 220**,** 449-55.

CONSORTIUM, I. H. 2005. A haplotype map of the human genome. *Nature,* 437**,** 1299-320.

CONSORTIUM, I. H. G. S. 2004. Finishing the euchromatic sequence of the human genome. *Nature,* 431**,** 931-45.

CONTI, M., et al. 2003. Cyclic AMP-specific PDE4 phosphodiesterases as critical components of cyclic AMP signaling. *J Biol Chem,* 278**,** 5493-6.

CORREIA, M., et al. 2004. Prospective community-based study of stroke in Northern Portugal: incidence and case fatality in rural and urban populations. *Stroke,* 35**,** 2048-53.

COTLARCIUC, I,. 2012. Bio-repository of DNA in stroke: a study protocol of three ancestral populations. *J R Soc Med Cardiovasc Dis,* 1.

CRUZ-FLORES, S., et al. 2011. Racial-ethnic disparities in stroke care: the American experience: a statement for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke,* 42**,** 2091-116.

DAHLOF, B., et al. 2005. Prevention of cardiovascular events with an antihypertensive regimen of amlodipine adding perindopril as required versus atenolol adding bendroflumethiazide as required, in the Anglo-Scandinavian Cardiac Outcomes Trial-Blood Pressure Lowering Arm (ASCOT-BPLA): a multicentre randomised controlled trial. *Lancet,* 366**,** 895-906.

DALLOL, A., et al. 2002. Tumour specific promoter region methylation of the human homologue of the Drosophila Roundabout gene DUTT1 (ROBO1) in human cancers. *Oncogene,* 21**,** 3020-8.

DANAEI, G., et al. 2009. The preventable causes of death in the United States: comparative risk assessment of dietary, lifestyle, and metabolic risk factors. *PloS Med,* 6**,** e1000058.

DANAEI, G., et al. 2006. Global and regional mortality from ischaemic heart disease and stroke attributable to higher-than-optimum blood glucose concentration: comparative risk assessment. *Lancet,* 368**,** 1651-9.

DAS, S. K. 2007. A Prospective Community-Based Study of Stroke in Kolkata, India. *Stroke,* 38**,** 906-910.

DAVIS, D. P., ROBERTSON, T. & IMBESI, S. G. 2006. Diffusion-weighted magnetic resonance imaging versus computed tomography in the diagnosis of acute ischemic stroke. *J Emerg Med,* 31**,** 269-77.

DE BAKKER, P. I., et al. 2008. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum Mol Genet,* 17**,** R122-8.

DEBETTE, S., et al. 2010. Genome-wide association studies of MRI-defined brain infarcts: meta-analysis from the CHARGE Consortium. *Stroke,* 41**,** 210-7.

DEEKS, J. J., ALTMAN, D. G. AND BRADBURN, M. J. 2008. Statistical Methods for Examining Heterogeneity and Combining Results from Several Studies in Meta-Analysis, in Systematic Reviews in Health Care: Meta-Analysis in Context. *BMJ Publishing Group, London, UK,* Second Edition (eds M. Egger, G. D. Smith and D. G. Altman).

DELLA-MORTE, D., et al. 2012. Genetics of ischemic stroke, stroke-related risk factors, stroke precursors and treatments. *Pharmacogenomics,* 13**,** 595-613.

DERSIMONIAN, R. & LAIRD, N. 1986. Meta-analysis in clinical trials. *Control Clin Trials,* 7**,** 177-88.

DESHMUKH, H. A., et al. 2012. Genome-wide association study of genetic determinants of LDL-c response to atorvastatin therapy: importance of Lp(a). *J Lipid Res,* 53**,** 1000-11.

DESMEULES, M., et al. 2005. Disparities in mortality patterns among Canadian immigrants and refugees, 1980-1998: results of a national cohort study. *J Immigr Health,* 7**,** 221-32.

DESTEFANO, A. L., et al. 2006. Genome-wide scan for white matter hyperintensity: the Framingham Heart Study. *Stroke,* 37**,** 77-81.

DEVLIN, B. & ROEDER, K. 1999. Genomic control for association studies. *Biometrics,* 55**,** 997-1004.

DHAMIJA, R.K., et al.2009. Hcys and Lipoprotein (a) correlation in Ischemic stroke patients. *Journal of the Neurological Sciences,* 281**,** 64–68.

DHILLON, V. S., SHAHID, M. & HUSAIN, S. A. 2007. Associations of MTHFR DNMT3b 4977 bp deletion in mtDNA and GSTM1 deletion, and aberrant CpG island hypermethylation of GSTM1 in non-obstructive infertility in Indian men. *Mol Hum Reprod,* 13**,** 213-22.

DINDAGUR, N., KRUTHIKA-VINOD, T. P. & CHRISTOPHER, R. 2006. Thrombophilic gene polymorphisms in puerperal cerebral veno-sinus thrombosis. *J Neurol Sci,* 249**,** 25-30.

DINDAGUR, N., KRUTHIKA-VINOD, T. P. & CHRISTOPHER, R. 2007. Factor V gene A4070G mutation and the risk of cerebral veno-sinus thrombosis occurring during puerperium. *Thromb Res,* 119**,** 497-500.

DING, H., et al. 2010. Confirmation of genomewide association signals in Chinese Han population reveals risk loci for ischemic stroke. *Stroke,* 41**,** 177-80.

DONNAN, G. A., FISHER, M., MACLEOD, M. & DAVIS, S. M. 2008. Stroke. *Lancet,* 371**,** 1612-23.

DRISCOLL, M. C., et al. 2003. Stroke risk in siblings with sickle cell anemia. *Blood,* 101**,** 2401-4.

DZIEDZIC, T., et al. 2002. Intracerebral hemorrhage triggers interleukin-6 and interleukin-10 release in blood. *Stroke,* 33**,** 2334-5.

EASTON, J. D., et al. 2009. Definition and evaluation of transient ischemic attack: a scientific statement for healthcare professionals from the American Heart Association/American Stroke Association Stroke Council; Council on Cardiovascular Surgery and Anaesthesia; Council on Cardiovascular Radiology and Intervention; Council on Cardiovascular Nursing; and the Interdisciplinary Council on Peripheral Vascular Disease. The American Academy of Neurology affirms the value of this statement as an educational tool for neurologists. *Stroke,* 40**,**2276-93.

EASTON, D. F. & EELES, R. A. 2008. Genome-wide association studies in cancer. *Hum Mol Genet,* 17**,** R109-15.

EBRAHIM, S., et al. 2010. The effect of rural-to-urban migration on obesity and diabetes in India: a cross-sectional study. *PloS Med,* 7**,** e1000268.

ECKSTEIN, H. H. 2012. Evidence-based management of carotid stenosis: recommendations from international guidelines. *J Cardiovasc Surg (Torino),* 53**,** 3-13.

EDWARDS, A. O., et al. 2005. Complement factor H polymorphism and age-related macular degeneration. *Science,* 308**,** 421-4.

EDWARDS, T. L., et al. 2010. Genome-wide association study confirms SNPs in SNCA and the MAPT region as common risk factors for Parkinson disease. *Ann Hum Genet,* 74**,** 97-109.

EGGER, M., DAVEY SMITH, G., SCHNEIDER, M. & MINDER, C. 1997. Bias in meta-analysis detected by a simple, graphical test. *BMJ,* 315**,** 629-34.

EHRET, G. B. 2010. Genome-wide association studies: contribution of genomics to understanding blood pressure and essential HTN. *Curr Hypertens Rep,* 12**,** 17-25.

EHRET, G. B., et al. 2011. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature,* 478**,** 103-9.

EKBOM, T., et al. 1992. The stroke preventive effect in elderly hypertensives cannot fully be explained by the reduction in office blood pressure--insights from the Swedish Trial in Old Patients with HTN (STOP-HTN). *Blood Press,* 1**,** 168-72.

ELLINOR, P. T., et al. 2012. Meta-analysis identifies six new susceptibility loci for AF. *Nat Genet,* 44**,** 670-5.

EUN-JU LEE, K. J. K., et al. 2011. Genome-wide scan of granular corneal dystrophy, type II: confirmation of chromosome 5q31 and identification of new co-segregated loci on chromosome 3q26.3. *Experimental and Molecular Medicine,* 43**,** 393-400.

EZZATI, M. & RIBOLI, E. 2012. Can noncommunicable diseases be prevented? Lessons from studies of populations and individuals. *Science,* 337**,** 1482-7.

FAIRHEAD, J. F., MEHTA, Z. & ROTHWELL, P. M. 2005. Population-based study of delays in carotid imaging and surgery and the risk of recurrent stroke. *Neurology,* 65**,** 371-5.

FALCONE, G. J., et al. 2012. Burden of risk alleles for HTN increases risk of intracerebral hemorrhage. *Stroke,* 43**,** 2877-83.

FATTORE, G., et al. 2012. The social and economic burden of stroke survivors in Italy: a prospective, incidence-based, multi-centre cost of illness study. *BMC Neurol,* 12**,** 137.

FEIGIN, V. L., et al. 2003. Stroke epidemiology: a review of population-based studies of incidence, prevalence, and case-fatality in the late 20th century. *Lancet Neurol,* 2**,** 43-53.

FEIGIN, V. L., et al. 2009. Worldwide stroke incidence and early case fatality reported in 56 population-based studies: a systematic review. *Lancet Neurol,* 8**,** 355-69.

FERRO, J. M., MASSARO, A. R. & MAS, J. L. 2010. Aetiological diagnosis of ischaemic stroke in young adults. *Lancet Neurol,* 9**,** 1085-96.

FLOSSMANN, E. & ROTHWELL, P. M. 2005. Family history of stroke in patients with transient ischemic attack in relation to HTN and other intermediate phenotypes. *Stroke,* 36**,** 830-5.

FLOSSMANN, E. & ROTHWELL, P. M. 2006. Family history of stroke does not predict risk of stroke after transient ischemic attack. *Stroke,* 37**,** 544-6.

FLOSSMANN, E., SCHULZ, U. G. & ROTHWELL, P. M. 2004. Systematic review of methods and results of studies of the genetic epidemiology of ischemic stroke. *Stroke,* 35**,** 212-27.

FLOSSMANN, E., SCHULZ, U. G. & ROTHWELL, P. M. 2005. Potential confounding by intermediate phenotypes in studies of the genetics of ischaemic stroke. *Cerebrovasc Dis,* 19**,** 1-10.

FOULKES, M. A., et al. 1988. The Stroke Data Bank: design, methods, and baseline characteristics. *Stroke,* 19**,** 547-54.

FRATTOLA, A., et al. 1993. Prognostic value of 24-hour blood pressure variability. *J Hypertens,* 11**,** 1133-7.

FRAZER, K. A., et al. 2007. A second generation human haplotype map of over 3.1 million SNPs. *Nature,* 449**,** 851-61.

FUKUTAKE, T. 2011. Cerebral autosomal recessive arteriopathy with subcortical infarcts and leukoencephalopathy (CARASIL): from discovery to gene identification. *J Stroke Cerebrovasc Dis,* 20**,** 85-93.

FUKUTAKE, T. & HIRAYAMA, K. 1995. Familial young-adult-onset arteriosclerotic leukoencephalopathy with alopecia and lumbago without arterial HTN. *Eur Neurol,* 35**,** 69-79.

GLESSNER, J. T., et al. 2009. Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature,* 459**,** 569-73.

GOHIL, R., PECK, G. & SHARMA, P. 2009. The genetics of venous thromboembolism. A meta-analysis involving approximately 120,000 cases and 180,000 controls. *Thromb Haemost,* 102**,** 360-70.

GOLDSTEIN, R.J., et al. 1976. Computer stroke registry and diagnosis program. *Neurology,* 25.

GOTO, Y., et al. 1992. Mitochondrial myopathy, encephalopathy, lactic acidosis, and stroke-like episodes (MELAS): a correlative study of the clinical features and mitochondrial DNA mutation. *Neurology,* 42**,** 545-50.

GOURIE-DEVI, M. 2004. Prevalence of neurological disorders in Bangalore, India: a community-based study with comparison between urban and rural areas. *Neuroepidemiology,* 23**,** 261–268.

GOVERMENT OF INDIA, C. B. O. H. I. 2008. National Health Profile 2008 *New Delhi: Directorate General of Health Sciences***,** 102.

GRETARSDOTTIR, S., et al. 2008. Risk variants for AF on chromosome 4q25 associate with ischemic stroke. *Ann Neurol,* 64**,** 402-9.

GRETARSDOTTIR, S., et al. 2003. The gene encoding phosphodiesterase 4D confers risk of ischemic stroke. *Nat Genet,* 35**,** 131-8.

GSCHWENDTNER, A., et al. 2009. Sequence variants on chromosome 9p21.3 confer risk for atherosclerotic stroke. *Ann Neurol,* 65**,** 531-9.

GUDBJARTSSON, D. F., et al. 2007. Variants conferring risk of AF on chromosome 4q25. *Nature,* 448**,** 353-7.

GUDBJARTSSON, D. F., et al. 2009. A sequence variant in ZFHX3 on 16q22 associates with AF and ischemic stroke. *Nat Genet,* 41**,** 876-8.

GUNARATHNE, A. 2009. Ischemic Stroke in South Asians A Review of the Epidemiology, Pathophysiology, and Ethnicity-Related Clinical Features. *Stroke,* 40**,** e415-e423.

GUPTA, A., et al. 2012. Cerebrovascular reserve and stroke risk in patients with carotid stenosis or occlusion: a systematic review &meta-analysis. *Stroke,* 43**,** 2884-91.

GURUMUKHANI, J. K., URSEKAR, M. & SINGHAL, B. S. 2004. Cerebral autosomal dominant arteriopathy with subcortical infarcts and leucoencephalopathy (CADASIL): a case report with review of literature. *Neurol India,* 52**,** 99-101.

HAAST, R. A., GUSTAFSON, D. R. & KILIAAN, A. J. 2012. Sex differences in stroke. *J Cereb Blood Flow Metab,* 32**,** 2100-7.

HALL, S. 2000. Frequency and allelic association of common variants in the lipoprotein lipase gene in different ethnic groups: The Wandsworth Heart and Stroke Study. *Genetic Epidemiology,* 18**,** 203–216.

HAMMAR, N., et al. 2002. Migration and mortality: a 20 year follow up of Finnish twin pairs with migrant co-twins in Sweden. *J Epidemiol Community Health,* 56**,** 362-6.

HANKEY, G. J., et al. 2013. Effect of B vitamins and lowering Hcys on cognitive impairment in patients with previous stroke or transient ischemic attack: a prespecified secondary analysis of a randomized, placebo-controlled trial and meta-analysis. *Stroke,* 44**,** 2232-9.

HANNULA-JOUPPI, K., et al. 2005. The axon guidance receptor gene ROBO1 is a candidate gene for developmental dyslexia. *PLoS Genet,* 1**,** e50.

HARA, K., et al. 2009. Association of HTRA1 mutations and familial ischemic cerebral small-vessel disease. *N Engl J Med,* 360**,** 1729-39.

HARVEY, A., et al. 2012. The future of technologies for personalised medicine. *N Biotechnol,* 29**,** 625-33.

HASAN, N., et al. 2012. Towards the identification of blood biomarkers for acute stroke in humans: a comprehensive systematic review. *Br J Clin Pharmacol,* 74**,** 230-40.

HASSAN, A., et al. 2002. Planning genetic studies in human stroke: sample size estimates based on family history data. *Neurology,* 58**,** 1483-8.

HATA, J., et al. 2007. Functional SNP in an Sp1-binding site of AGTRL1 gene is associated with susceptibility to brain infarction. *Hum Mol Genet,* 16**,** 630-9.

HELGADOTTIR, A., et al. 2012. Apolipoprotein(a) genetic sequence variants associated with systemic atherosclerosis and coronary atherosclerotic burden but not with venous thromboembolism. *J Am Coll Cardiol,* 60**,** 722-9.

HELGADOTTIR, A., et al. 2004. The gene encoding 5-lipoxygenase activating protein confers risk of MI and stroke. *Nat Genet,* 36**,** 233-9.

HEUSCHMANN, P. U., et al. 2009. Incidence of stroke in Europe at the beginning of the 21st century. *Stroke,* 40**,** 1557-63.

HIGGINS, J. P. & THOMPSON, S. G. 2002. Quantifying heterogeneity in a meta-analysis. *Stat Med,* 21**,** 1539-58.

HINGORANI, A. D. & CASAS, J. P. 2012. The interleukin-6 receptor as a target for prevention of coronary heart disease: a mendelian randomisation analysis. *Lancet,* 379**,** 1214-24.

HOLDT, L. M. & TEUPSER, D. 2012. Recent studies of the human chromosome 9p21 locus, which is associated with atherosclerosis in human populations. *Arterioscler Thromb Vasc Biol,* 32**,** 196-206.

HOLLANDER, M., et al. 2002. Carotid plaques increase the risk of stroke and subtypes of cerebral infarction in asymptomatic elderly: the Rotterdam study. *Circulation,* 105**,** 2872-7.

HOLLIDAY, E. G., et al. 2012. Common variants at 6p21.1 are associated with large artery atherosclerotic stroke. *Nat Genet,* 44**,** 1147-51.

HOLMES.,M.V., et al. 2014a. Mendelian randomization of blood lipids for coronary heart disease. Eur Heart J, [Epub ahead of print]

HOLMES, M.V., et al. 2014b. Causal effects of body mass index on cardiometabolic traits and events: a Mendelian randomization analysis. *Am J Hum Genet*. 6;94(2):198-208.

HOLMES, M. V., et al. 2011. Effect modification by population dietary folate on the association between MTHFR genotype, Hcys, and stroke risk: a meta-analysis of genetic studies and randomised trials. *Lancet,* 378**,** 584-94.

HONG, H., et al. 2012. Technical reproducibility of genotyping SNP arrays used in genome-wide association studies. *PLoS One,* 7**,** e44483.

HOZO, S. P., DJULBEGOVIC, B. & HOZO, I. 2005. Estimating the mean and variance from the median, range, and the size of a sample. *BMC Med Res Methodol,* 5**,** 13.

HSU, F. C., et al. 2011. Transcobalamin 2 variant associated with poststroke Hcys modifies recurrent stroke risk. *Neurology,* 77**,** 1543-50.

HUANG, T., et al. 2013.Association of homocysteine with type 2 diabetes: a meta-analysis implementing Mendelian randomization approach. *BMC Genomics*, 14, 867.

HUGHES, K. & ONG, C. N. 2000. Hcys, folate, vitamin B12 &cardiovascular risk in Indians, Malays &Chinese in Singapore. *J Epidemiol Community Health,* 54**,** 31-4.

HUMPHREYS, K., et al. 2011. The genetic structure of the Swedish population. *PLoS One,* 6**,** e22547.

HUNT, K. J., et al. 2002. Genetic basis of variation in carotid artery plaque in the San Antonio Family Heart Study. *Stroke,* 33**,** 2775-80.

ICHTCHENKO, K., et al. 1995. Neuroligin 1: a splice site-specific ligand for beta-neurexins. *Cell,* 81**,** 435-43.

IKRAM, M. A., et al. 2009. Genomewide association studies of stroke. *N Engl J Med,* 360**,** 1718-28.

INABA, Y., CHEN, J. A. & BERGMANN, S. R. 2012. Carotid plaque, compared with carotid intima-media thickness, more accurately predicts coronary artery disease events: a meta-analysis. *Atherosclerosis,* 220**,** 128-33.

INTERNATIONAL CONSORTIUM FOR BLOOD PRESSURE GENOME-WIDE ASSOCIATION, S., EHRET, et al. 2011. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature,* 478**,** 103-9.

INTERNATIONAL STROKE GENETICS, C. & WELLCOME TRUST CASE-CONTROL, C. 2010. Failure to validate association between 12p13 variants and ischemic stroke. *N Engl J Med,* 362**,** 1547-50.

INTERNATIONAL STROKE GENETICS, C., WELLCOME TRUST CASE CONTROL, C., BELLENGUEZ, C., et al. 2012. Genome-wide association study identifies a variant in HDAC9 associated with large vessel ischemic stroke. *Nat Genet,* 44**,** 328-33.

IOANNIDIS, J. P., NTZANI, E. E. & TRIKALINOS, T. A. 2004. 'Racial' differences in genetic effects for complex diseases. *Nat Genet,* 36**,** 1312-8.

IOANNIDIS, J. P. & PANAGIOTOU, O. A. 2011. Comparison of effect sizes associated with biomarkers reported in highly cited individual articles and in subsequent meta-analyses. *JAMA,* 305**,** 2200-10.

JAIN, S., RAM, H., KUMARI, S. & KHULLAR, M. 2003. Plasma Hcys levels in Indian patients with essential HTN and their siblings. *Ren Fail,* 25**,** 195-201.

JANSSENS, A.C., et al. 2009. An empirical comparison of meta-analyses of published gene-disease associations versus consortium analyses. Genetics in Medicine, 11(3): 153-62.

JASHARI, F., et al. 2012. Coronary and carotid atherosclerosis: Similarities and differences. *Atherosclerosis,*227(2), 193-200.

JENSEN, M. B., et al. 2009. Potential biomarkers for the diagnosis of stroke. *Expert Rev Cardiovasc Ther,* 7**,** 389-93.

REFERENCES

JERRARD-DUNNE, P., et al. 2003. Evaluating the genetic component of ischemic stroke subtypes: a family history study. *Stroke,* 34**,** 1364-9.

JIANG, Y. H., BRESSLER, J. & BEAUDET, A. L. 2004. Epigenetics and human disease. *Annu Rev Genomics Hum Genet,* 5**,** 479-510.

JOHANSSON, A., ENROTH, S. & GYLLENSTEN, U. 2013. Continuous Aging of the Human DNA Methylome Throughout the Human Lifespan. *PLoS One,* 8**,** e67378.

JOHN C. et al. 2000. Methylenetetrahydrofolate Reductase 677 C3T Mutationand Coronary Heart Disease Risk in UK Indian Asians. *Arterioscler Thromb Vasc Biol.,* 20**,** 2448-2452.

JOHNSON, C. J., et al. 1995. Interrater reliability of an etiologic classification of ischemic stroke. *Stroke,* 26**,** 46-51.

JOHNSON, T., et al. 2011. Blood pressure loci identified with a gene-centric array. *Am J Hum Genet,* 89**,** 688-700.

JOHNSTON, S. C., et al. 2009. Global variation in stroke burden and mortality: estimates from monitoring, surveillance, and modelling. *Lancet Neurol,* 8**,** 345-54.

JONES, T. H., et al. 1981. Thresholds of focal cerebral ischemia in awake monkeys. *J Neurosurg,* 54**,** 773-82.

JOOD, K., et al. 2005. Fibrinolytic gene polymorphism and ischemic stroke. *Stroke,* 36**,** 2077-81.

JORDAN, J. D. & POWERS, W. J. 2012. Cerebral autoregulation and acute ischemic stroke. *Am J Hypertens,* 25**,** 946-50.

JOUSILAHTI, P., et al. 1997. Parental history of cardiovascular disease and risk of stroke. A prospective follow-up of 14371 middle-aged men and women in Finland. *Stroke,* 28**,** 1361-6.

JOUTEL, A., et al. 1996. Notch3 mutations in CADASIL, a hereditary adult-onset condition causing stroke and dementia. *Nature,* 383**,** 707-10.

KALEV-ZYLINSKA, M. L., et al. 2013. Stroke patients develop antibodies that react with components of N-methyl-D-aspartate receptor subunit 1 in proportion to lesion size. *Stroke,* 44**,** 2212-9.

KALITA, J. 2009. Relationship of Hcys with other risk factors and outcome of ischemic stroke. *Clinical Neurology and Neurosurgery,* 111**,** 364–367.

KALITA, J., et al. 2011. A study of ACE and ADD1 polymorphism in ischemic and hemorrhagic stroke. *Clin Chim Acta,* 412**,** 642-6.

KALITA, J., et al. 2006. Methylenetetrahydrofolate reductase gene polymorphism in Indian stroke patients. *Neurol India,* 54**,** 260-3.

KAMSTRUP, P. R., et al. 2012. Genetic evidence that lipoprotein(a) associates with atherosclerotic stenosis rather than venous thrombosis. *Arterioscler Thromb Vasc Biol,* 32**,** 1732-41.

KASIMAN, K., et al. 2012. Familial effects on ischemic stroke: the role of sibling kinship, sex, and age of onset. *Circ Cardiovasc Genet,* 5**,** 226-33.

KASNER, S. E. 2006. Clinical interpretation and use of stroke scales. *Lancet Neurol,* 5**,** 603-12.

KATO, N., et al J. 2011. Meta-analysis of genome-wide association studies identifies common variants associated with blood pressure variation in east Asians. *Nat Genet,* 43**,** 531-8.

KAUL, S. & MUNSHI, A. 2012. Genetics of ischemic stroke: Indian perspective. *Neurol India,* 60**,** 498-503.

KELEMEN, L. E. 2004. Associations of plasma Hcys and the methylenetetrahydrofolatereductase C677T polymorphism with carotid intima media thicknessamong South Asian, Chinese and European Canadians. *Atherosclerosis,* 176**,** 361–370.

KELLY, P. J., et al. 2003. Stroke in young patients with hyperHcysmia due to cystathionine beta-synthase deficiency. *Neurology,* 60**,** 275-9.

KHAJA, A. M. & GROTTA, J. C. 2007. Established treatments for acute ischaemic stroke. *Lancet,* 369**,** 319-30.

KHAN, N. A., et al. 2013. Risk factors, quality of care and prognosis in South Asian, East Asian and white patients with stroke. *BMC Neurol,* 13**,** 74.

KHAN, T. A., et al. 2013.Apolipoprotein E genotype, cardiovascular biomarkers and risk of stroke: systematic review and meta-analysis of14,015 stroke cases and pooled analysis of primary biomarker data from up to 60,883individuals. *Int J Epidemiol*, 42, 475-92.

KIKUYA, M., et al. 2000. Prognostic significance of blood pressure and heart rate variabilities: the Ohasama study. *HTN,* 36**,** 901-6.

KIM, D.S., et al. 2014. Rare coding variation in paraoxonase-1 is associated with ischemic stroke in the NHLBI Exome Sequencing Project. *J. Lipid Res*, 55:(6), 1173-1178.

REFERENCES

KAIN, K., et al.2005. Coagulationfactor XIII Bsubunitantigen, FXIIIVal34Leu genotype andischaemic stroke in South Asians. *Thrombosis and Haemostasis,* 93**,** 394-395.

KISSELA, B. M., et al. 2012. Age at stroke: temporal trends in stroke incidence in a large, biracial population. *Neurology,* 79**,** 1781-7.

KITTNER, S. J., et al. 1998. Cerebral infarction in young adults: the Baltimore-Washington Cooperative Young Stroke Study. *Neurology,* 50**,** 890-4.

KLEINJAN, D. A. & VAN HEYNINGEN, V. 2005. Long-range control of gene expression: emerging mechanisms and disruption in disease. *Am J Hum Genet,* 76**,** 8-32.

KONIG, I. R., et al. 2008. Predicting long-term outcome after acute ischemic stroke: a simple index works in patients from controlled clinical trials. *Stroke,* 39**,** 1821-6.

KOPYTA, I. A., et al. 2012. Polymorphisms of genes encoding coagulation factors II, V, VII, and XIII in relation to pediatric ischemic stroke: family-based and case-control study. *Neurologist,* 18**,** 282-6.

KOSTULAS, K., et al. 2007. PDE4D and ALOX5AP genetic variants and risk for Ischemic CVDin Sweden. *J Neurol Sci,* 263**,** 113-7.

KOSTULAS, K., et al. 1999. An angiotensin-converting enzyme gene polymorphism suggests a genetic distinction between ischaemic stroke and carotid stenosis. *Eur J Clin Invest,* 29**,** 478-83.

KOSTULAS, N., et al. 2009. Common CCR 5 polymorphism in stroke: the CCR 5 delta32 polymorphism differentiates cardioembolism from other aetiologies of ischaemic cerebrovascular diseases. *Scand J Immunol,* 70**,** 475-80.

KRUG, T., et al. 2012. TTC7B emerges as a novel risk factor for ischemic stroke through the convergence of several genome-wide approaches. *J Cereb Blood Flow Metab,* 32**,** 1061-72.

KUBO, M., et al. 2007. A nonsynonymous SNP in PRKCH (protein kinase C eta) increases the risk of cerebral infarction. *Nat Genet,* 39**,** 212-7.

KUMAR, J., et al. 2005. Hcys levels are associated with MTHFR A1298C polymorphism in Indian population. *J Hum Genet,* 50**,** 655-63.

KUNUTSOR, S.K., et al. 2014. Vitamin D and high blood pressure: causal association or epiphenomenon? *Eur J Epidemiol*, 29:1–14.

KUSUMA, Y., GUPTA, S. & PANDAV, C. 2009. Migration and HTN: a cross-sectional study among neo-migrants and settled-migrants in Delhi, India. *Asia Pac J Public Health,* 21**,** 497-507.

LAMBERT, C. G. & BLACK, L. J. 2012. Learning from our GWAS mistakes: from experimental design to scientific method. *Biostatistics,* 13**,** 195-203.

LANKTREE, M. B., DICHGANS, M. & HEGELE, R. A. 2010. Advances in genomic analysis of stroke: what have we learned and where are we headed? *Stroke,* 41**,** 825-32.

LANKTREE, M. B., HEGELE, R. A., YUSUF, S. & ANAND, S. S. 2009. Multi-ethnic genetic association study of carotid intima-media thickness using a targeted cardiovascular SNP microarray. *Stroke,* 40**,** 3173-9.

LANSKA, D. J. 1997. Geographic distribution of stroke mortality among immigrants to the United States. *Stroke,* 28**,** 53-7.

LAO, O., et al. 2008. Correlation between genetic and geographic structure in Europe. *Curr Biol,* 18**,** 1241-8.

LAW, M.R., WALD, N.J., RUDNICKA, A.R.2003. Quantifying effect of statins on low-density lipoprotein cholesterol, ischemic heart disease, and stroke: systematic review and meta-analysis. *BMJ*, 326, 1423.

LAW, M. R., et al. 2009. Use of blood pressure lowering drugs in the prevention of cardiovascular disease: meta-analysis of 147 randomised trials in the context of expectations from prospective epidemiological studies. *BMJ,* 338**,** b1665.

LAWLOR, D.A., et al. 2008. Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. *Statist. Med.,* 27:1133–1163.

LEE, L. J., et al. 2000. Impact on stroke subtype diagnosis of early diffusion-weighted magnetic resonance imaging and magnetic resonance angiography. *Stroke,* 31**,** 1081-9.

LEE, S., SHAFE, A. C. & COWIE, M. R. 2011. UK stroke incidence, mortality and cardiovascular risk management 1999-2008: time-trend analysis from the General Practice Research Database. *BMJ Open,* 1**,** e000269.

LEEK, J. T., et al. 2010. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat Rev Genet,* 11**,** 733-9.

LEISTNER, S., et al. 2008. Increased prevalence of microangiopathic brain lesions among siblings of patients with lacunar stroke. A prospective multicenter study. *Eur Neurol,* 59**,** 143-7.

LETTRE, G. & RIOUX, J. D. 2008. Autoimmune diseases: insights from genome-wide association studies. *Hum Mol Genet,* 17**,** R116-21.

LEVY, D., et al. 2009. Genome-wide association study of blood pressure and HTN. *Nat Genet,* 41**,** 677-87.

LEWANDOWSKI, C. & BARSAN, W. 2001. Treatment of acute ischemic stroke. *Ann Emerg Med,* 37**,** 202-16.

LI, M., BOEHNKE, M. & ABECASIS, G. R. 2006. Efficient study designs for test of genetic association using sibship data and unrelated cases and controls. *Am J Hum Genet,* 78**,** 778-92.

LIAO, D., et al. 1997. Familial history of stroke and stroke risk. The Family Heart Study. *Stroke,* 28**,** 1908-12.

LIEW, S. C., et al. 2012. Association of Methylentetraydrofolate Reductase (MTHFR) 677 C > T gene polymorphism and Hcys levels in psoriasis vulgaris patients from Malaysia: a case-control study. *Nutr J,* 11**,** 1.

LIM, S. S., et al. 2012. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet,* 380**,** 2224-60.

LIN, B. K., et al. 2006. Tracking the epidemiology of human genes in the literature: the HuGE Published Literature database. *Am J Epidemiol,* 164**,** 1-4.

LINDGREN, A., et al. 2005. Prevalence of stroke and vascular risk factors among first-degree relatives of stroke patients and control subjects. A prospective consecutive study. *Cerebrovasc Dis,* 20**,** 381-7.

LISABETH, L. D., et al. 2008. Stroke among siblings in a biethnic community. *Neuroepidemiology,* 31**,** 33-8.

LIU, M. E., et al. 2013. A functional polymorphism of PON1 interferes with microRNA binding to increase the risk of ischemic stroke and carotid atherosclerosis. *Atherosclerosis,* 228**,** 161-7.

LO, S. S., et al. 1991. Studies of diabetic twins. *Diabetes Metab Rev,* 7**,** 223-38.

LUTHRA, K., et al. 2002. Apolipoprotein E gene polymorphism in cerebrovascular disease: a case-control study. *Clin Genet,* 62**,** 39-44.

IQBAL, M.P, et al. 2005. Lack of association of methylenetetrahydrofolate reductase 677C>T mutationwith coronary artery disease in a Pakistani population. *Journal of Molecular and Genetic Medicine,* 1**,** 26-32

MACCLELLAN, L. R., et al. 2006. Familial aggregation of ischemic stroke in young women: the stroke prevention in young women study. *Genet Epidemiol,* 30**,** 602-8.

MACINTYRE, G., BAILEY, J., HAVIV, I. & KOWALCZYK, A. 2010. is-rSNP: a novel technique for in silico regulatory SNP detection. *Bioinformatics,* 26**,** i524-30.

MAHONEY, F. I. & BARTHEL, D. W. 1965. Functional Evaluation: The Barthel Index. *Md State Med J,* 14**,** 61-5.

MAJUMDAR, V., NAGARAJA, D., KARTHIK, N., CHRISTOPHER, R. 2009. Association of ENOS gene polymorphisms with early-onset stroke in South Indians. *J. of Atherosclerosis and thrombosis,* 17(1), 45-53.

MAJUMDAR, V., NAGARAJA, D. & CHRISTOPHER, R. 2010. Association of the functional KL-VS variant of Klotho gene with early-onset ischemic stroke. *Biochem Biophys Res Commun,* 403**,** 412-6.

MAMATHA, S. N., NAGARAJA, D., PHILIP, M. & CHRISTOPHER, R. 2011. Asymmetric dimethylarginine as a risk marker for early-onset ischemic stroke in Indian population. *Clin Chim Acta,* 412**,** 139-42.

MANCIA, G., et al. 2007. Long-term prognostic value of blood pressure variability in the general population: results of the Pressioni Arteriose Monitorate e Loro Associazioni Study. *HTN,* 49**,** 1265-70.

MANOLIO, T. A., et al. 2004. Genetics of ultrasonographic carotid atherosclerosis. *Arterioscler Thromb Vasc Biol,* 24**,** 1567-77.

MANOLIO, T. A., et al. 2009. Finding the missing heritability of complex diseases. *Nature,* 461**,** 747-53.

MANWARING, N., et al. 2007. Population prevalence of the MELAS A3243G mutation. *Mitochondrion,* 7**,** 230-3.

MARCHINI, J., et al. 2004. The effects of human population structure on large genetic association studies. *Nat Genet,* 36**,** 512-7.

MARKUS, H., 2003. Stroke Genetics. *Oxford University Press,*Chapter 1.

MARJOT, T., YADAV, S., HASAN, N., BENTLEY, P. & SHARMA, P. 2011. Genes associated with adult cerebral venous thrombosis. *Stroke,* 42**,** 913-8.

MATARIN, M., et al. 2008. Association of integrin alpha2 gene variants with ischemic stroke. *J Cereb Blood Flow Metab,* 28**,** 81-9.

MATARIN, M., et al. 2008. Whole genome analyses suggest ischemic stroke and heart disease share an association with polymorphisms on chromosome 9p21. *Stroke,* 39**,** 1586-9.

MATARIN, M., et al. 2008. Structural genomic variation in ischemic stroke. *Neurogenetics,* 9**,** 101-8.

MATSUSHITA, T., et al. 2009. Lack of association between variations of PDE4D and ischemic stroke in the Japanese population. *Stroke,* 40**,** 1245-51.

MAYOSI, B. M., et al. 2005. Genotype at the -174G/C polymorphism of the interleukin-6 gene is associated with common carotid artery intimal-medial thickness: family study and meta-analysis. *Stroke,* 36**,** 2215-9.

MCALLISTER, S., et al. 2013. Do different types of financial support after illness or injury affect socio-economic outcomes? A natural experiment in New Zealand. *Soc Sci Med,* 85**,** 93-102.

MCCARROLL, S. A. & ALTSHULER, D. M. 2007. Copy-number variation and association studies of human disease. *Nat Genet,* 39**,** S37-42.

MCCOLGAN, P. & SHARMA, P. 2008. The genetics of carotid dissection: meta-analysis of a MTHFR/C677T common molecular variant. *Cerebrovasc Dis,* 25**,** 561-5.

MCCOLGAN, P., THANT, K. Z. & SHARMA, P. 2010. The genetics of sporadic ruptured and unruptured intracranial aneurysms: a genetic meta-analysis of 8 genes and 13 polymorphisms in approximately 20,000 individuals. *J Neurosurg,* 112**,** 714-21.

MCEVOY, M., et al. 2010. Cohort profile: The Hunter Community Study. *Int J Epidemiol,* 39**,** 1452-63.

MEDICI, F., et al. 1999. Concordance rate for type II T2Din monozygotic twins: actuarial analysis. *Diabetologia,* 42**,** 146-50.

MEHNDIRATTA, M. M., et al. 2013. Meeting the challenges of stroke in India. *Neurology*, 80, 2246-7.

MEJIA MOHAMED, E. H., TAN, K. S., ALI, J. M. & MOHAMED, Z. 2011. TT genotype of the methylenetetrahydrofolate reductase C677T polymorphism is an important determinant for Hcys levels in multi-ethnic Malaysian ischaemic stroke patients. *Ann Acad Med Singapore,* 40**,** 186-6.

MESCHIA, J. F. 2011. New information on the genetics of stroke. *Curr Neurol Neurosci Rep,* 11**,** 35-41.

MESCHIA, J. F., et al. 2003. The Ischemic Stroke Genetics Study (ISGS) Protocol. *BMC Neurol,* 3**,** 4.

MESCHIA, J. F., et al. 2000. Verifying the stroke-free phenotype by structured telephone interview. *Stroke,* 31**,** 1076-80.

MESCHIA, J. F., et al. 2006. The Siblings With Ischemic Stroke Study (SWISS): a progress report. *Clin Med Res,* 4**,** 12-21.

MESCHIA, J. F., et al. 2011. Siblings with ischemic stroke study: results of a genome-wide scan for stroke loci. *Stroke,* 42**,** 2726-32.

MESCHIA, J. F. & TOURNIER-LASSERVE, E. 2013. Advances in stroke: genetics 2012. *Stroke,* 44**,** 309-10.

MESCHIA, J. F.,et al. 2013. Stroke Genetics Network (SiGN) Study: Design and Rationale for a Genome-Wide Association Study of Ischemic Stroke Subtypes. *Stroke.* 44, 2694-2702.

MICHEAL, S., et al.2009. MTHFR gene C677T and A1298C polymorphisms andHcys levels in primary open angle and primary closedangle glaucoma. *Molecular Vision,* 15**,** 2268-2278.

MISHRA, N. K., et al. 2010. Thrombolysis in very elderly people: controlled comparison of SITS International Stroke Thrombolysis Registry and Virtual International Stroke Trials Archive. *BMJ,* 341**,** c6046.

MISRA, U. K., et al. 2010. MTHFR gene polymorphism and its relationship with plasma Hcys and folate in a North Indian population. *Biochem Genet,* 48**,** 229-35.

MODI, M.,et al. 2005. HyperHcysmia as a risk factor for ischemic stroke: An Indian scenario. *Neurology India,* 53**,** 297-301.

MOHR, J. P., et al. 1995. Magnetic resonance versus computed tomographic imaging in acute stroke. *Stroke,* 26**,** 807-12.

MOHR, J. P., et al. 1978. The Harvard Cooperative Stroke Registry: a prospective registry. *Neurology,* 28**,** 754-62.

MONET-LEPRETRE, M., et al. 2009. Distinct phenotypic and functional features of CADASIL mutations in the Notch3 ligand binding domain. *Brain,* 132**,** 1601-12.

MONTANER, J. 2009. Blood biomarkers to guide stroke thrombolysis. *Front Biosci (Elite Ed),* 1**,** 200-8.

MORRISON, A. C., FORNAGE, M., LIAO, D. & BOERWINKLE, E. 2000. Parental history of stroke predicts subclinical but not clinical stroke: the Atherosclerosis Risk in Communities Study. *Stroke,* 31**,** 2098-102.

MORTON, N. E. & COLLINS, A. 1998. Tests and estimates of allelic association in complex inheritance. *Proc Natl Acad Sci U S A,* 95**,** 11389-93.

MUNSHI, A., et al. 2008. Angiotensin-converting enzyme Insertion/Deletion polymorphism and the risk of ischemis stroke in a south indian population. *J of Neurological Sciences,* 272**,** 1320-135.

MUNSHI, A., et al. 2009. Phosphodiesterase 4D (PDE4D) gene variants &the risk of ischemic stroke in South Indian population. *J of Neurological Sciences,* 285**,** 142-145.

MUNSHI, A., et al.2010. VNTR polymorphism in intron 4 of the eNOS gene and the risk of ischemic stroke in a South Asian population. *Brain Res Bull*, 82(5-6), 247-50.

MUNSHI, A., et al. 2009. Prothombin gene G20210A mutation is not a risk factor for ischemic stroke in a South Indian Hyderabadi Population. *Thromb Res,* 124**,** 245-7.

MUNSHI, A., et al. 2010. Interleukin-10-1082 promoter polymorphism and ischemic stroke risk in a South Indian population. *Cytokine,* 52**,** 221-4.

MUNSHI, A., et al. 2011. Association of tumor necrosis factor-alpha and matrix metalloproteinase-3 gene variants with stroke. *Eur J Neurol,* 18**,** 1053-9.

MUNSHI, A., et al. 2010. Estrogen receptor alpha genetic variants and the risk of stroke in a South Indian population from Andhra Pradesh. *Clin Chim Acta,* 411**,** 1817-21.

MUNSHI, A., et al. 2012. Association of 1347 G/A cytochrome P450 4F2 (CYP4F2) gene variant with HTN and stroke. *Mol Biol Rep,* 39**,** 1677-82.

MUNSHI, A., et al. 2010. Association of the -344C/T aldosterone synthase (CYP11B2) gene variant with HTN and stroke. *J Neurol Sci,* 296**,** 34-8.

MUNSHI, A. 2012.Genetic variation in MDR1, LPL and eNOS genes and the response to atorvastatin treatment in ischemic stroke. *Hum Genet*, 131, 1775-81.

MUQTADAR, H. & TESTAI, F. D. 2012. Single gene disorders associated with stroke: a review and update on treatment options. *Curr Treat Options Cardiovasc Med,* 14**,** 288-97.

MURRAY, C. J. & LOPEZ, A. D. 1997. Global mortality, disability, and the contribution of risk factors: Global Burden of Disease Study. *Lancet,* 349**,** 1436-42.

MURRAY, C. J., et al. 2012. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet,* 380**,** 2197-223.

MUSIALEK, P., et al. 2013. Multimarker approach in discriminating patients with symptomatic and asymptomatic atherosclerotic carotid artery stenosis. *J Clin Neurol,* 9**,** 165-75.

MYKKANEN, K., et al. 2009. Different clinical phenotypes in monozygotic CADASIL twins with a novel NOTCH3 mutation. *Stroke,* 40**,** 2215-8.

NARANG, A. P., et al. 2009. Hcys--risk factor for ischemic stroke? *Indian J Physiol Pharmacol,* 53**,** 34-8.

NARAYAN, G., et al. 2006. Promoter hypermethylation-mediated inactivation of multiple Slit-Robo pathway genes in cervical cancer progression. *Mol Cancer,* 5**,** 16.

NEWCOMBE, P. J., et al. 2009. Multilocus Bayesian meta-analysis of gene-disease associations. *Am J Hum Genet,* 84**,** 567-80.

NEWTON-CHEH, C., et al. 2009. Genome-wide association study identifies eight loci associated with blood pressure. *Nat Genet,* 41**,** 666-76.

NICOLAOU, M., et al. 2000. Genetic predisposition to stroke in relatives of hypertensives. *Stroke,* 31**,** 487-92.

NORSKOV, M. S., et al. 2011. Copy number variation in glutathione S-transferases M1 and T1 and ischemic vascular disease: four studies and meta-analyses. *Circ Cardiovasc Genet,* 4**,** 418-28.

O'DONNELL, M. J., et al. 2010. Risk factors for ischaemic and intracerebral haemorrhagic stroke in 22 countries (the INTERSTROKE study): a case-control study. *Lancet,* 376**,** 112-23.

O'LEARY, D. H., et al. 1999. Carotid-artery intima and media thickness as a risk factor for MI and stroke in older adults. Cardiovascular Health Study Collaborative Research Group. *N Engl J Med,* 340**,** 14-22.

OCSP. 1983. Incidence of stroke in Oxfordshire: first year's experience of a community stroke register. *Br Med J (Clin Res Ed),* 287**,** 713-7.

OKADA, Y., et al. 2012. Meta-analysis identifies multiple loci associated with kidney function-related traits in east Asian populations. *Nat Genet,* 44**,** 904-9.

OLSSON, S., et al. 2012. Genetic variation within the interleukin-1 gene cluster and ischemic stroke. *Stroke,* 43**,** 2278-82.

OPHERK, C., et al. 2004. Long-term prognosis and causes of death in CADASIL: a retrospective study in 411 patients. *Brain,* 127**,** 2533-9.

ORGANISATION, W. H. 2012. World Health Statistics.

OSLER, M., HOLST, C., PRESCOTT, E. & SORENSEN, T. I. 2001. Influence of genes and family environment on adult smoking behavior assessed in an adoption study. *Genet Epidemiol,* 21**,** 193-200.

PACI, M., NANNETTI, L., D'IPPOLITO, P. & LOMBARDI, B. 2011. Outcomes from ischemic stroke subtypes classified by the Oxfordshire Community Stroke Project: a systematic review. *Eur J Phys Rehabil Med,* 47**,** 19-23.

PANAGARIYA, A., SHARMA, B. & SHUBHAKARAN 2004. CADASIL in a family from north-west India. *J Assoc Physicians India,* 52**,** 580-1.

PANIGRAHI, I. 2006. Role of MTHFR C677T polymorphism in ischemic stroke. *Neurology India,* 54(1), 48-50.

PANTONI, L., et al. 2010. Comparison of clinical, familial, and MRI features of CADASIL and NOTCH3-negative patients. *Neurology,* 74**,** 57-63.

PATERNOSTER, L., CHEN, W. & SUDLOW, C. L. 2009. Genetic determinants of white matter hyperintensities on brain scans: a systematic assessment of 19 candidate gene polymorphisms in 46 studies in 19,000 subjects. *Stroke,* 40**,** 2020-6.

PAUL, N., et al. 2013. Depression Among Stroke Survivors: A Community-based, Prospective Study from Kolkata, India. *Am J Geriatr Psychiatry,* 21**,** 821-31.

PELTONEN, L., PALOTIE, A. & LANGE, K. 2000. Use of population isolates for mapping complex traits. *Nat Rev Genet,* 1**,** 182-90.

PERKEL, J. 2008. SNP genotyping: six technologies that keyed a revolution. *Nature Methods,* 5.

PETERSEN, L., NIELSEN, G. G., ANDERSEN, P. K. & SORENSEN, T. I. 2002. Case-control study of genetic and environmental influences on premature death of adult adoptees. *Genet Epidemiol,* 23**,** 123-32.

PETERSEN, L., SORENSEN, T. I., MORTENSEN, E. L. & ANDERSEN, P. K. 2010. Excess mortality rate during adulthood among Danish adoptees. *PLoS One,* 5**,** e14365.

PHADKE, S. R., MANDAL, K. & GIRISHA, K. M. 2009. Fabry disease: a treatable lysosomal storage disorder. *Natl Med J India,* 22**,** 20-2.

POTKIN, S. G., et al. 2009. A genome-wide association study of schizophrenia using brain activation as a quantitative phenotype. *Schizophr Bull,* 35**,** 96-108.

POTLURI, R., NATALWALA, A. 2009. Increasing prevalence of haemorrhagic stroke among South Asian patients in the United Kingdom from 1997 to 2005. *Journal of Clinical Neuroscience,* 16**,** 605–606.

POULTER, N. R., et al. 1990. The Kenyan Luo migration study: observations on the initiation of a rise in blood pressure. *BMJ,* 300**,** 967-72.

PRASAD, K., VIBHA, D. & MEENAKSHI. 2012. CVDin South Asia – Part I: A burning problem. *JRSM Cardiovascular Disease,* 1, 1-7.

PRICE, A. L., et al. 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet,* 38**,** 904-9.

PRITCHARD, J. K. & COX, N. J. 2002. The allelic architecture of human disease genes: common disease-common variant...or not? *Hum Mol Genet,* 11**,** 2417-23.

PURCELL, S., et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet,* 81**,** 559-75.

RADHA, A., et al. 2004. Prevalence of methylene tetrahydrofolate reductase polymorphism in South Indian population. *Current Science,* 86,3.

RADHIANA, H., et al. 2013. Non-contrast Computed Tomography in Acute Ischaemic Stroke: A Pictorial Review. *Med J Malaysia,* 68**,** 93-100.

RAO, R., et al. 2009. Ischaemic stroke subtypes and their genetic basis: a comprehensive meta-analysis of small and large vessel stroke. *Eur Neurol,* 61**,** 76-86.

RAO, S. 1971. Aspects of epidemiology of stroke in south India. *Proceedings of the First All India Conference on Stroke***,** 25–32.

RAWLINS, M. D. 2004. Cutting the cost of drug development? *Nat Rev Drug Discov,* 3**,** 360-4.

RAZVI, S. S., et al. 2005. The prevalence of cerebral autosomal dominant arteriopathy with subcortical infarcts and leucoencephalopathy (CADASIL) in the west of Scotland. *J Neurol Neurosurg Psychiatry,* 76**,** 739-41.

REICH, D., et al. 2009. Reconstructing Indian population history. *Nature,* 461**,** 489-94.

RHA, J. H., et al. 2013. Two-year direct medical costs of stroke in Korea: a multi-centre incidence-based study from hospital perspectives. *Int J Stroke,* 8**,** 186-92.

RIGAT, C.H.B., et al. 1990. An insertion/deletion polymorphism in the angiotensin I-converting enzyme gene accounting for half the variance of serum enzyme levels. *J Clin Invest.,* 86**,** 1343–1346.

RISCH, N. 1990. Linkage strategies for genetically complex traits. III. The effect of marker polymorphism on analysis of affected relative pairs. *Am J Hum Genet,* 46**,** 242-53.

ROGER, V. L., et al. 2012. Executive summary: heart disease and stroke statistics--2012 update: a report from the American Heart Association. *Circulation,* 125**,** 188-97.

ROLFS, A., et al. 2005. Prevalence of Fabry disease in patients with cryptogenic stroke: a prospective study. *Lancet,* 366**,** 1794-6.

ROMERO, J. R. & WOLF, P. A. 2013. Epidemiology of Stroke: Legacy of the Framingham Heart Study. *Glob Heart,* 8**,** 67-75.

ROSAND, J., et al. 2006. Many hypotheses but no replication for the association between PDE4D and stroke. *Nat Genet,* 38**,** 1091-2; author reply 1092-3.

ROSE, F. C. 1985. Problems in the management of stroke. *Ann Acad Med Singapore,* 14**,** 12-5.

ROSS, O.A., et al. 2013. NOTCH3 Variants and Risk of Ischemic Stroke. *PLoSONE,*8(9): e75035.

ROST, N. S., et al. 2010. Determinants of white matter hyperintensity volume in patients with acute ischemic stroke. *J Stroke Cerebrovasc Dis,* 19**,** 230-5.

ROSTILA, M., SAARELA, J. & KAWACHI, I. 2013. Fatal stroke after the death of a sibling: a nationwide follow-up study from Sweden. *PLoS One,* 8**,** e56994.

ROTHWELL, P. M. 2010. Limitations of the usual blood-pressure hypothesis and importance of variability, instability, and episodic HTN. *Lancet,* 375**,** 938-48.

ROTHWELL, P. M., et al. 2010. Effects of beta blockers and calcium-channel blockers on within-individual variability in blood pressure and risk of stroke. *Lancet Neurol,* 9**,** 469-80.

ROTHWELL, P. M., et al. 2010. Prognostic significance of visit-to-visit variability, maximum systolic blood pressure, and episodic HTN. *Lancet,* 375**,** 895-905.

SACCO, R. L. 2001. Race-Ethnic Disparities in the Impact of Stroke Risk Factors, The Northern Manhattan Stroke Study. *Stroke***,** 32:1725.

SACCO, R. L., et al. 2013. An updated definition of stroke for the 21st century: a statement for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke,* 44**,**2064-89.

SCHMIDT, H., et al. 2011. Genetic variants of the NOTCH3 gene in the elderly and magnetic resonance imaging correlates of age-related cerebral small vessel disease. *Brain*, 134, 3384-97.

SAHA, S.P., et al. 2003. Epidemiological study of neurological disorders in a rural population of Eastern India. *J Indian Med Assoc,* 101**,** 299 –304.

SALEHEEN, D., et al. 2005. Association of phosphodiesterase 4D gene with ischemic stroke in a Pakistani population. *Stroke,* 36**,** 2275-7.

SAPOSNIK, G., et al. 2010. Risk of premature stroke in recent immigrants (PRESARIO): population-based matched cohort study. *Neurology,* 74**,** 451-7.

SCHULZ, U. G., FLOSSMANN, E. & ROTHWELL, P. M. 2004. Heritability of ischemic stroke in relation to age, vascular risk factors, and subtypes of incident stroke in population-based studies. *Stroke,* 35**,** 819-24.

SEN, S., et al. 2013. Racial-ethnic differences in stroke risk factors and subtypes: results of a prospective hospital-based registry. *Int J Neurosci*, 123(8), 568-74.

SESHADRI, S., et al. 2006. The lifetime risk of stroke: estimates from the Framingham Study. *Stroke,* 37**,** 345-50.

SESHADRI, S., et al. 2010. Parental occurrence of stroke and risk of stroke in their children: the Framingham study. *Circulation,* 121**,** 1304-12.

SEVER, P. S., et al. 2001. Rationale, design, methods and baseline demography of participants of the Anglo-Scandinavian Cardiac Outcomes Trial. ASCOT investigators. *J Hypertens,* 19**,** 1139-47.

SHARMA, P., YADAV, S. & MESCHIA, J. F. 2013. Genetics of ischaemic stroke. *J Neurol Neurosurg Psychiatry*. (Epub ahead of publication)

SHARMA, V., et al. 2011. Association of C3435T multi drug resistance gene-1 polymorphism with aspirin resistance in ischemic stroke and its subtypes. *J Neurol Sci*, 315(1-2), 72-6.

SHAW, J. E., SICREE, R. A. & ZIMMET, P. Z. 2010. Global estimates of the prevalence of diabetes for 2010 and 2030. *Diabetes Res Clin Pract,* 87**,** 4-14.

SHETTY, P. 2012. Public health: India's diabetes time bomb. *Nature,* 485, S14-S16.

SIMS, K., et al. 2009. Stroke in Fabry disease frequently occurs before diagnosis and in the absence of other clinical events: natural history data from the Fabry Registry. *Stroke,* 40**,** 788-94.

SKOL, A. D., et al. 2006. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet,* 38**,** 209-13.

SOMARAJAN, B. I., et al. 2010. A study of alpha1 antichymotrypsin gene polymorphism in Indian stroke patients. *J Neurol Sci,* 290**,** 57-9.

SOMARAJAN B.I., et al. 2011. Evaluation of MTHFR C677T polymorphism in ischemic and hemorrhagic stroke patients. A case-control study in a Northern Indian population. *J Neurol Sci.,* 15**,** 67-70.

SORENSEN, T. I., et al. 1988. Genetic and environmental influences on premature death in adult adoptees. *N Engl J Med,* 318**,** 727-32.

SOTOODEHNIA, N., et al. 2010. Common variants in 22 loci are associated with QRS duration and cardiac ventricular conduction. *Nat Genet,* 42**,** 1068-76.

SPENCE, J. D. & HEGELE, R. A. 2004. Noninvasive phenotypes of atherosclerosis: similar windows but different views. *Stroke,* 35**,** 649-53.

SPENCE, J. D., et al. 2001. Vitamin Intervention for Stroke Prevention (VISP) trial: rationale and design. *Neuroepidemiology,* 20**,** 16-25.

SPIELMAN, R. S., MCGINNIS, R. E. & EWENS, W. J. 1993. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent T2D(IDDM). *Am J Hum Genet,* 52**,** 506-16.

SREEDHARAN, S. E., et al. 2013. Employment status, social function decline and caregiver burden among stroke survivors. A South Indian study. *J Neurol Sci,* 332**,** 97-101.

STANKOVIC, S., ASANIN, M. & MAJKIC-SINGH, N. 2012. *Genetics of Ischemic Stroke: Emphasis on Candidate-Gene Association Studies.*

STERN, J.M. & SIMES, R.J. 1997. Publication bias: evidence of delayed publication in a cohort study of clinical research projects. BMJ, Sep 13; 315(7109): 640-5.

STRONG, K., MATHERS, C. & BONITA, R. 2007. Preventing stroke: saving lives around the world. *Lancet Neurol,* 6**,** 182-7.

STUNKARD, A. J., et al. 1986. An adoption study of human obesity. *N Engl J Med,* 314**,** 193-8.

SUCKOW, A. T., et al. 2008. Expression of neurexin, neuroligin, and their cytoplasmic binding partners in the pancreatic beta-cells and the involvement of neuroligin in insulin secretion. *Endocrinology,* 149**,** 6006-17.

SUKLA, K. K. & RAMAN, R. 2012. Association of MTHFR and RFC1 gene polymorphism with hyperHcysmia and its modulation by vitamin B12 and folic acid in an Indian population. *Eur J Clin Nutr,* 66**,** 111-8.

SULTANA S., et al. 2011. Tumour necrosis factor alpha and interleukin 10 gene polymorphisms and the risk of ischemic stroke in south Indian population. *J. Genet.,* 90**,** 361–364.

SYLAJA, P. N., et al. 2006. Thrombolysis in patients older than 80 years with acute ischaemic stroke: Canadian Alteplase for Stroke Effectiveness Study. *J Neurol Neurosurg Psychiatry,* 77**,** 826-9.

TABANGIN, M. E., WOO, J. G. & MARTIN, L. J. 2009. The effect of minor allele frequency on the likelihood of obtaining false positives. *BMC Proc,* 3 Suppl 7**,** S41.

TAO HM, S. B., CHEN GZ. 2009. Meta-analysis of the ACE gene polymorphism in cerebral infarction. *Can J Neurol Sci.,* 36**,** 20-5.

TESLOVICH, T. M., et al. 2010. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature,* 466**,** 707-13.

THAPAR, A., et al. 2013. Diagnosis and Management of Carotid Atherosclerosis. *BMJ*.346, f1485.

THE INTERLEUKIN-6 RECEPTOR MENDELIAN RANDOMISATION ANALYSIS (IL6R MR) CONSORTIUM. 2012. The interleukin-6 receptor as a target for prevention of coronary heart disease: a Mendelian randomization analysis. *Lancet,* 379, 1214-24.

THOMPSON, J. F., et al. 2005. An association study of 43 SNPs in 16 candidate genes with atorvastatin response. *Pharmacogenomics J*, 5, 352-8.

THOM, T., et al. 2006. Heart disease and stroke statistics--2006 update: a report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. *Circulation,* 113**,** e85-151.

TISZLAVICZ, Z., et al. 2012. Genetic polymorphisms of human beta-defensins in patients with ischemic stroke. *Acta Neurol Scand,* 126**,** 109-15.

TOUZE, E. & ROTHWELL, P. M. 2007. Heritability of ischaemic stroke in women compared with men: a genetic epidemiological study. *Lancet Neurol,* 6**,** 125-33.

TOUZE, E. & ROTHWELL, P. M. 2008. Sex differences in heritability of ischemic stroke: a systematic review and meta-analysis. *Stroke,* 39**,** 16-23.

TOYOOKA, K. 2011. Fabry disease. *Curr Opin Neurol,* 24**,** 463-8.

TRAYLOR, M., et al. 2013. Using phenotypic heterogeneity to increase the power of genome-wide association studies: application to age at onset of ischaemic stroke subphenotypes. *Genet Epidemiol,* 37**,** 495-503.

TRAYLOR, M., et al. 2012. Genetic risk factors for ischaemic stroke and its subtypes (the METASTROKE Collaboration): a meta-analysis of genome-wide association studies. *Lancet Neurol,* 11**,** 951-962.

TRIPATHI, R., et al. 2010. Association of Hcys and methylene tetrahydrofolate reductase(MTHFR C677T) gene polymorphism with coronary artery disease (CAD)in the population of North India. *Genetics and Molecular Biology,* 33**,** 224-228.

TROEDSON, C., et al. 2013. Systemic lupus erythematosus due to C1q deficiency with progressive encephalopathy, intracranial calcification and acquired moyamoya cerebral vasculopathy. *Lupus,* 22**,** 639-43.

TSAI, C. F., THOMAS, B. & SUDLOW, C. L. 2013. Epidemiology of stroke and its subtypes in Chinese vs white populations: A systematic review. *Neurology,* 81**,** 264-72.

TU, J. V. 2010. Reducing the global burden of stroke: INTERSTROKE. *Lancet,* 376**,** 74-5.

TURNER, S. T., et al. 2005. Genomic susceptibility loci for brain atrophy in hypertensive sibships from the GENOA study. *HTN,* 45**,** 793-8.

VANDERWEELE, T.J., et al. 2014. Methodological Challenges in Mendelian Randomization. *Epidemiology*, 25: 427–435.

VAN EEDEN, M., et al. 2012. The economic impact of stroke in The Netherlands: the euro-Restore4Stroke study. *BMC Public Health,* 12**,** 122.

VAN SWIETEN, J. C.,. 1988. Interobserver agreement for the assessment of handicap in stroke patients. *Stroke,* 19**,** 604-7.

VOIGHT, B. F., et al. 2010. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat Genet,* 42**,** 579-89.

WAIN, L. V., et al. 2011. Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. *Nat Genet,* 43**,** 1005-11.

WALD, D. S., LAW, M. & MORRIS, J. 2003. Serum Hcys and the severity of coronary artery disease. *Thromb Res,* 111**,** 55-7.

WANG, X., et al. 2009. A meta-analysis of candidate gene polymorphisms and ischemic stroke in 6 study populations: association of lymphotoxin-alpha in nonhypertensive patients. *Stroke,* 40**,** 683-95.

WANG, Y., et al. 2013. LRIG1 is a triple threat: ERBB negative regulator, intestinal stem cell marker and tumour suppressor. *Br J Cancer,* 108**,** 1765-70.

WARD, L. D. & KELLIS, M. 2012. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res,* 40**,** D930-4.

WARDLAW, J. M., WARLOW, C. P. & COUNSELL, C. 1997. Systematic review of evidence on thrombolytic therapy for acute ischaemic stroke. *Lancet,* 350**,** 607-14.

WEAVER, I. C., et al. 2004. Epigenetic programming by maternal behavior. *Nat Neurosci,* 7**,** 847-54.

WEBB, A. J., FISCHER, U. & ROTHWELL, P. M. 2011. Effects of beta-blocker selectivity on blood pressure variability and stroke: a systematic review. *Neurology,* 77**,** 731-7.

WELIN, L., et al. 1987. Analysis of risk factors for stroke in a cohort of men born in 1913. *N Engl J Med,* 317**,** 521-6.

WHISNANT, J. P. 1997. Modeling of risk factors for ischemic stroke. The Willis Lecture. *Stroke,* 28**,** 1840-4.

WHITE, H., et al. 2005. Ischemic stroke subtype incidence among whites, blacks, and Hispanics: the Northern Manhattan Study. *Circulation,* 111**,** 1327-31.

WHITELEY, W., CHONG, W. L., SENGUPTA, A. & SANDERCOCK, P. 2009. Blood markers for the prognosis of ischemic stroke: a systematic review. *Stroke,* 40**,** e380-9.

WILCOX, W. R., et al. 2008. Females with Fabry disease frequently have major organ involvement: lessons from the Fabry Registry. *Mol Genet Metab,* 93**,** 112-28.

WILLER, C. J., LI, Y. & ABECASIS, G. R. 2010. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics,* 26**,** 2190-1.

WILLIAMS, F. M., et al. 2013. Ischemic stroke is associated with the ABO locus: the EuroCLOT study. *Ann Neurol,* 73**,** 16-31.

WILLIAMS, G. R. 2001. Incidence and characteristics of total stroke in the United States. *BMC Neurol,* 1**,** 2.

WIVIOTT, S. D., ANTMAN, E. M. & BRAUNWALD, E. 2010. Prasugrel. *Circulation,* 122**,** 394-403.

WON, S. J., KIM, D. Y. & GWAG, B. J. 2002. Cellular and molecular pathways of ischemic neuronal death. *J Biochem Mol Biol,* 35**,** 67-86.

WORLD HEALTH ORGANIZATION. 2002. The Atlas of Heart Disease and Stroke. www.who.int/cardiovascular diseases/resources/atlas/en/. URL accessed on 28.09.2013.

WORLD HEALTH ORGANIZATION. 2010. Global status report on noncommunicable diseases. www.whqlibdoc.who.int/publications/2011/9789240686458_eng.pdf. URL accessed on 28.09.2013.

WORLD HEALTH ORGANIZATION. 2013.A global brief on Hypertension. http://apps.who.int/iris/bitstream/10665/79059/1/WHO_DCO_WHD_2013.2_eng.pdf.URL accessed on 28.09.2013.

WORRALL, B. B., CHEN, D. T. & MESCHIA, J. F. 2001. Ethical and methodological issues in pedigree stroke research. *Stroke,* 32**,** 1242-9.

WORRALL, B. B. & MYCHALECKYJ, J. C. 2006. PDE4D and stroke: a real advance or a case of the Emperor's new clothes? *Stroke,* 37**,** 1955-7.

WOZNIAK, M. A., et al. 2010. Frequency of unrecognized Fabry disease among young European-American and African-American men with first ischemic stroke. *Stroke,* 41**,** 78-81.

XU, X., LI, X., LI, J., OU, R. & SHENG, W. 2010. Meta-analysis of association between variation in the PDE4D gene and ischemic cerebral infarction risk in Asian populations. *Neurogenetics,* 11**,** 327-33.

YADAV, S., et al. 2013. The first Indian-origin family with genetically proven cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL). *J Stroke Cerebrovasc Dis,* 22**,** 28-31.

YADAV, S., et al. 2013. Genome-Wide Analysis of Blood Pressure Variability and Ischemic Stroke. *Stroke.* (Epub ahead of publication)

YADAV, S., et al. 2013. Detailed analysis of gene polymorphisms associated with ischemic stroke in South Asians. *PLoS One,* 8**,** e57305.

YADAV, S., et al. 2011. Bio-Repository of DNA in stroke (BRAINS): a study protocol. *BMC Med Genet,* 12**,** 34.

YAMADA, Y., et al. 2009. Identification of CELSR1 as a susceptibility gene for ischemic stroke in Japanese individuals by a genome-wide association study. *Atherosclerosis,* 207**,** 144-9.

YANG, J., et al. 2010. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet,* 42**,** 565-9.

YATSU, F. M. & SHALTONI, H. M. 2004. Implications of silent strokes. *Curr Atheroscler Rep,* 6**,** 307-13.

YOON, D., et al. 2011. Meta-Analysis of Homogeneous Subgroups Reveals Association between PDE4D Gene Variants &Ischemic Stroke. *Neuroepidemiology,* 36**,** 213-222.

ZAKOPOULOS, N. A., et al. 2005. Time rate of blood pressure variation is associated with increased common carotid artery intima-media thickness. *HTN,* 45**,** 505-12.

ZEGGINI, E., et al. 2007. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science,* 316**,** 1336-41.

ZHANG, C., et al. 2013. Effect of B-vitamin Supplementation on Stroke: a Meta-Analysis of Randomized Controlled Trials. *PLoS ONE,* 8(11): e81577.

ZHANG, F., et al. 2010. Mechanisms for nonrecurrent genomic rearrangements associated with CMT1A or HNPP: rare CNVs as a cause for missing heritability. *Am J Hum Genet,* 86**,** 892-903.

ZIVIN, J. A. & CHOI, D. W. 1991. Stroke therapy. *Sci Am,* 265**,** 56-63.

ZUO, L., et al. 2012. Genome-wide association study of alcohol dependence implicates KIAA0040 on chromosome 1q. *Neuropsychopharmacology,* 37**,** 557-66.

ZUO, L., et al. 2012. Genome-wide search for replicable risk gene regions in alcohol and nicotine co-dependence. *Am J Med Genet B Neuropsychiatr Genet,* 159B**,** 437-44.

ZUREIK, M., et al. 1999. Differential association of common carotid intima-media thickness and carotid atherosclerotic plaques with parental history of premature death from coronary heart disease : the EVA study. *Arterioscler Thromb Vasc Biol,* 19**,** 366-71.

APPENDICES

Appendix 1: Laboratory Techniques

A1.1 DNA Isolation

Blood samples were collected by the in-house phlebotomist in EDTA coated tubes and stored at -20 °C. Frozen blood samples were thawed in a 37 °C water bath with mild agitation and stored on ice till the start of the procedure. DNA isolation was performed using the Qiagen Gentra Puregene Blood kit (Catalogue number: 158445). 3X RBC lysis solution was added to the blood sample and inverted 10 times. Sample was incubated for 5 min at room temperature (15-25°C) and inverted at least once during incubation. White blood cells were isolated by centrifugation (2.000xg, 5 min). The supernatant was carefully discarded by pouring, leaving only about 200 µl of the residual liquid and the WBC pellet. The pellet was completely dispersed by vigorous vortexing and equivalent starting sample volume of the cell lysis solution was added. Sample was vigorously vortexed for 10s and incubated at 37 °C for 5 min. 1 µl of RNase A solution/ 200 µl of starting sample volume was added and incubated for 15 min at 37 °C followed by 3 min on ice. 333 µl of protein precipitation solution/ ml of starting sample volume was added and vigorously vortexed for 20s. Sample was incubated on ice for 5 min and centrifuged (2.000xg, 5 min). Supernatant was carefully decanted into isopropanol and mixed gently by inverting 50 times till the DNA threads were visible. The sample was centrifuged (2.000xg, 3 min), supernatant discarded and tube drained by inverting on a clean piece of absorbant paper. Pellet was washed by rinsing with 10 ml 70% ethanol, followed by centrifugation (2.000xg, 3 min). DNA pellets were air-dried for 15 min before re-suspension in DNA hydration solution. Tubes were incubated at 65°C for 1 hour to dissolve the DNA and incubated overnight at room temperature with gentle shaking. Samples were centrifuged and stored at -20°C.

A1.2 Genotyping and Quality Control

Processing was carried out in accordance with the Infinium HD Super Assay (Rev.C, June 2010) protocol (Illumina Inc, San Diego, USA). Briefly, in a deep well plate 200ng of high quality genomic DNA was whole genome amplified overnight (37°C, 20-24 hours), then fragmented (37°C for 1 hour and 15mins in hybridization oven), precipitated and resuspended in hybridization buffer. Samples were denatured then taken from the plate and loaded onto the chips using a liquid handling robot (Freedom Evo, Tecan Ltd, Switzerland). Hybridization took place overnight at 48°C (16-24 hours). The process of single base extension (SBE) and staining was carried out using the liquid handling robot at 44°C and 32°C respectively. The probes on the chip were extended by a single hapten-labelled dideoxynucleotide (ddNTP) base complementary to the hybridized DNA. ddATP and ddTTP bases were labeled with DNP (2,4-Dinitrophenol), ddCTP and ddGTP were labeled with Biotin. The DNA sample was then stripped off the chip using formamide. The staining procedure involves signal amplification by multi-layer immune histochemical staining. Streptavidin detected the haptens simultaneously and an anti-DNP primary antibody conjugated to green and red fluorophores respectively (STM reagent, Illumina). They were then counterstained with biotinylated anti-steptavidin antibody and a DNP-labeled secondary antibody to the anti-DNP antibody (ATM reagent, Illumina) to amplify the fluorescent signals. The last layer of stain was the STM, containing fluorophores to allow signal detection. Finally the stained chips were coated in XC4 reagent, a glue-like substance to protect the dyes, and scanned using the iScan scanner with autoloader (IlluminaInc, San Diego, USA).

The robot was used for hybridization and staining to ensure that samples are not mixed up during transfer from plate to chip and to ensure that the pipetting and incubation for the staining procedure is accurate and consistent across the batch. An initial quality check was carried out on the genomic DNA before processing to ensure appropriate concentration and there was no loss of integrity.

Intensity data were quality controlled using the genotyping module in the Illumina Genome studio software. Samples were assessed for their call rate, which was set at ≥ 95% individually, and average across the batch. Control probes on the chip were checked to identify any possible processing errors: Every array contains both sample

dependent and sample independent control probes. Sample independent probes assess the quality of the processing; sample dependent probes also assess the quality of the DNA. The B-allele frequency plots (BAF) were also checked to identify contaminated samples. This plot would show more than three modes if the sample had been contaminated at source. A noisy BAF plot may also suggest degradation of the DNA sample.

Appendix 2: Publications, Presentations and Grants

A2.1 Publications

*1. **Sunaina Yadav**, Ioana Cotlarciuc, Patricia B. Munroe, Muhammad S Khan, Michael A Nalls, Steve Bevan, Yu-Ching Cheng, Wei-Min Chen, Rainer Malik, Nina S McCarthy, Elizabeth G Holliday, Douglas Speed, Nazeeha Hasan, Mateusz Pucek, Paul E. Rinne, Peter Sever, Alice Stanton, Denis Shields, Jane M Maguire, Mark McEvoy, Rodney J Scott, Luigi Ferrucci, Mary J Macleod, John Attia, Hugh S Markus, Michele M Sale, Bradford B Worrall, Braxton D Mitchell, Martin Dichgans, Cathy Sudlow, James F Meschia, Peter M Rothwell, Mark Caulfield, Pankaj Sharma, & International Stroke Genetics Consortium. Genome wide analysis of blood pressure variability and ischemic stroke. *Stroke*. 2013 Aug 8. (Epub ahead of print). PMID: 23929743

*2. Pankaj Sharma, **Sunaina Yadav**, James F. Meschia. Genetics of ischemic stroke. *J Neurol Neurosurg Psychiatry*. 2013 Apr 25. (Epub ahead of print). PMID: 23620417

*3. **Yadav S**, Hasan N, Marjot T, Khan MS, Prasad K, Bentley P, Sharma P. Detailed analysis of gene polymorphisms associated with ischemic stroke in South Asians. *PLoS One*. 2013; 8(3): e57305. PMID: 23505425

4. Traylor M, Farrall M, Holliday EG, Sudlow C, Hopewell JC, Cheng YC, Fornage M, Ikram MA, Malik R, Bevan S, Thorsteinsdottir U, Nalls MA, Longstreth W, Wiggins KL, **Yadav S**, Parati EA, Destefano AL, Worrall BB, Kittner SJ, Khan MS, Reiner AP, Helgadottir A, Achterberg S, Fernandez-Cadenas I, Abboud S, Schmidt R, Walters M, Chen WM, Ringelstein EB, O'Donnell M, Ho WK, Pera J, Lemmens R, Norrving B, Higgins P, Benn M, Sale M, Kuhlenbäumer G, Doney AS, Vicente AM, Delavaran H, Algra A, Davies G, Oliveira SA, Palmer CN, Deary I, Schmidt H, Pandolfo M, Montaner J, Carty C, de Bakker PI, Kostulas K, Ferro JM, van Zuydam NR, Valdimarsson E, Nordestgaard BG, Lindgren A, Thijs V, Slowik A, Saleheen D, Paré G, Berger K, Thorleifsson G; The Australian Stroke Genetics Collaborative, Wellcome Trust Case Control Consortium 2 (WTCCC2), Hofman A, Mosley TH, Mitchell BD, Furie K, Clarke R, Levi C, Seshadri S, Gschwendtner A, Boncoraglio GB, Sharma P, Bis JC, Gretarsdottir S, Psaty BM, Rothwell PM, Rosand J, Meschia JF, Stefansson K, Dichgans M, Markus HS; on behalf of the International Stroke Genetics Consortium. Genetic risk factors for ischaemic stroke and its subtypes (the

METASTROKE Collaboration): a meta-analysis of genome-wide association studies. *Lancet Neurol.* 2012 Nov; 11(11): 951-962. PMID: 23041239

5. Sian-Tsung Tan, Abtehale Al-Hussaini, **Sunaina Yadav**, Joban Sehmi, Mika Ala-Korpela, Paul Elliott, John C Chambers, Jaspal S Kooner. Does ApoB/A1 Ratio Explain the Increased Risk of Cardiovascular Disease in Indian Asians Compared to Europeans?*Circulation.* 2012; 125; AP191

6. Cheng YC, Anderson CD, Bione S, Keene K, Maguire JM, Nalls M, Rasheed A, Zeginigg M, Attia J, Baker R, Barlera S, Biffi A, Bookman E, Brott TG, Brown RD Jr, Chen F, Chen WM, Ciusani E, Cole JW, Cortellini L, Danesh J, Doheny K, Ferrucci L, GraziaFranzosi M, Frossard P, Furie KL, Golledge J, Hankey GJ, Hernandez D, Holliday EG, Hsu FC, Jannes J, Kamal A, Khan MS, Kittner SJ, Koblar SA, Lewis M, Lincz L, Lisa A, Matarin M, Moscato P, Mychaleckyj JC, Parati EA, Parolo S, Pugh E, Rost NS, Schallert M, Schmidt H, Scott RJ, Sturm JW, **Yadav S**, Zaidi M, Boncoraglio GB, Levi CR, Meschia JF, Rosand J, Sale M, Saleheen D, Schmidt R, Sharma P, Worrall B, Mitchell BD; GARNET Collaborative Research Group; GENEVA Consortium; on behalf of the International Stroke Genetics Consortium. Are MI-Associated Single-Nucleotide Polymorphisms Associated With Ischemic Stroke? *Stroke.* 2012 Apr; 43(4): 980-986. PMID: 22363065

7. **Yadav S**, Bentley P, Srivastava P, Prasad K, Sharma P.The First Indian-Origin Family with Genetically Proven Cerebral Autosomal Dominant Arteriopathy with Subcortical Infarcts and Leukoencephalopathy (CADASIL).*J Stroke Cerebrovasc Dis.* 2011 Jul 5. PMID: 21737310

8. Marjot T, **Yadav S**, Hasan N, Bentley P, Sharma P. Genes associated with adult cerebral venous thrombosis. *Stroke.* 2011 Apr; 42(4):913-8. Epub 2011 Feb 24.PMID: 21350198

9. **Yadav S**, Schanz R, Maheshwari A, Khan MS, Slark J, de Silva R, Bentley P, Froguel P, Kooner J, Shrivastav P, Prasad K, Sharma P. Bio-Repository of DNA in stroke (BRAINS): a study protocol. *BMC Med Genet.*2011 Mar 2; 12:34.PMID: 21366918

A2.2 Oral Presentations

*1. **Sunaina Yadav**, Ioana Cotlarcuic, Pankaj Sharma. Genome Wide Association Studies for Blood Pressure Variability and Ischemic Stroke. London Cardiovascular Society Young Investigators Award 2013 (Top 3 finalist). *Royal Medical Society*, February 2013

A2.3 Poster Presentations

1. Marjot T, **Yadav S**, Sharma P. Genes associated with adult cerebral venous thrombosis. *UK Stroke Forum 2010*

*2. **Sunaina Yadav**, Ioana Cotlarcuic, Pankaj Sharma. Genome Wide Association Studies for Blood Pressure Variability and Ischemic Stroke. *European Stroke Conference 2013*

A2.4 Grants

*1. United Kingdom-India Education and Research Initiative (UKIERI) and British Council PhD scholarship. January 2010-January 2013

*2. Welcome Trust Advanced Courses Bursary award. Welcome Trust Genome campus (Cambridge), August 2010

* Publications, Presentations and Grants relevant to this thesis

Appendix 3: Copyright Permissions

A3.1 Summary of permission for third party copyright works

Page Number	Type of work: text, figure, map, etc.	Source work	Copyright holder & year	Permission to re-use
Introduction Pages 31-81	Text	J Neurol Neurosurg Psychiatry (2013), PMID: 23620417	© Sharma et all 2013 published by BMJ journals	✓ (Appendix 3.2)
Chapter 2 Pages 83-115	Text, images and table	PLoS One (2013), PMID: 23505425	©Yadav et al 2013 published by PLoS One	✓ (Appendix 3.3)
Chapter 3 Pages 116-147	Text, images and table	Stroke (2013), PMID: 23929743	© Stroke 2013	✓ (Appendix 3.4)

A3.2 Journal of Neurology, Neurosurgery & Psychiatry copyright license

http://group.bmj.com/products/journals/instructions-for-authors/wholly_owned_licence.pdf

URL accessed on 16th September 2013

A3.3 PLoS One copyright license

http://www.plosone.org/static/editorial#copyright

URL accessed on 16th September 2013

A3.4 Stroke copyright license

http://www.ahajournals.org/site/rights/

URL accessed on 23rd September 2013