**Supplementary Online Material**

**Individual Study Description – Discovery Analysis**

# Australian Stroke Genetics Collaborative (ASGC)

ASGC stroke cases comprised European-ancestry stroke patients admitted to four clinical centres across Australia (The Neurosciences Department at Gosford Hospital, Gosford, New South Wales (NSW); the Neurology Department at John Hunter Hospital, Newcastle, NSW; The Queen Elizabeth Hospital, Adelaide; and the Royal Perth Hospital, Perth) between 2003 and 2008. Stroke was defined by WHO criteria as a sudden focal neurologic deficit of vascular origin, lasting more than 24 hours and confirmed by imaging such as computerised tomography (CT) and/or magnetic resonance imaging (MRI) brain scan. Other investigative tests such as electrocardiogram, carotid doppler and transoesophageal echocardiogram were conducted to define IS mechanism as clinically appropriate. Cases were excluded from participation if aged <18 years, diagnosed with haemorrhagic stroke or transient ischemic attack rather than IS, or were unable to undergo baseline brain imaging. Based on these criteria, a total of 1230 IS cases were included in the current study. IS subtypes were assigned using TOAST criteria, based on 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 Caucasian ancestry and residing in theHunter Region, NSW, Australia. Detailed recruitment methods for the HCS have been previously described. Briefly, participants were randomly selected from the NSW State electoral roll and contacted by mail between 2004 and 2007. Consenting participants completed five detailed self-report questionnaires and attended the HCS data collection centre, at which time a series of clinical measures were obtained. A total of 1280 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 relevant institutional ethics committees.

# Bio-Repository of DNA in Stroke (BRAINS)

The Bio-Repository of DNA in Stroke (BRAINS) is an international study recruiting highly phenotyped patients with stroke. For the purposes of the current work all patients were Caucasians.

Diagnosis of stroke was confirmed using positive imaging (MRI or CT) and ischemic stroke subtypes were assigned using TOAST criteria, based on clinical, imaging and risk factor data. Controls were European-Ancestry, stroke-free participants from the shared WTCCC controls, a prospectively collected cohort of individuals born in 1958 (1958 Birth Cohort). The cohort has been described in detail elsewhere.

# The Genetics of Early Onset Stroke (GEOS) Study, Baltimore, USA

GEOS is a population-based case-control study designed to identify genes associated with earlyonset stroke in patients with first-ever ischemic stroke aged 15-49 years from the greater Baltimore- Washington area between 1992 and 2008. Only patients of European descent are included in this meta-analysis.

Cases were identified through discharge surveillance from 59 participating hospitals and direct physician referral from a defined geographic region. Abstracted medical records were reviewed and adjudicated for ischemic stroke subtype by two neurologists, with discrepancies resolved by a third neurologist. Controls with no history of ischemic stroke were identified through random digit dialing and were frequency-matched to cases based on sex, age, geographic location and, during the later study periods, ethnicity.

# Heart Protection Study (HPS)

The Heart Protection Study (HPS) was a large randomized trial involving individuals at increased risk of vascular events. Between 1994-1997 20,536 men and women aged 40-80 years were recruited from 69 collaborating hospitals in the United Kingdom (with ethics committee approval). Participants were eligible for inclusion provided they had non-fasting blood total cholesterol concentrations of at least 135 mg/dL (3.5 mmol/L) and either a previous diagnosis of coronary disease, ischemic stroke, other occlusive disease of non-coronary arteries, diabetes mellitus, or (if were men 65 years or older) treated hypertension. None of them was on statin therapy. At the initial screening visit, all participants provided written consent and began a “runin” phase involving 4 weeks of placebo followed by 4 to 6 weeks of 40 mg simvastatin daily, after which compliant and eligible individuals were randomly allocated 40 mg simvastatin daily or matching placebo for approximately 5 years. Individuals entering HPS with a clinical diagnosis of ischemic stroke were used as cases in the METASTROKE study. Individuals entering HPS with pre-existing diabetes but no history of cerebrovascular disease, coronary heart disease or peripheral vascular disease were used as controls.

**The Ischemic Stroke Genetics Study (ISGS)/ Siblings With Ischemic Stroke Study (SWISS)** The Siblings with Ischemic Stroke Study (SWISS) is a multicenter affected sibling pair study enrolling probands with ischemic stroke at 66 US medical centers and 4 Canadian medical centers. All 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. Additionally all probands were required to have at least one living sibling with a history of stroke. Siblings were enrolled using proband-initiated contact or direct contact when permitted by Institutional Review Boards. Clinical exclusion criteria mirrored that in ISGS. Concordant siblings had their diagnosis of ischemic stroke confirmed by review of medical records by a central vascular neurology committee. Subtype diagnoses were assigned to the index strokes of probands and concordant siblings according to TOAST criteria. Readily available US controls were utilized, including stroke-free participants from the Baltimore Longitudinal Study of Aging and the National Institute of Neurological Diseases and Stroke neurologically normal control series taken from the Coriell Cell Repositories. All controls had been previously genotyped and described in detail elsewhere.

The Ischemic Stroke Genetic Study (ISGS) is a multicenter study where inpatient cases were recruited from five United States academic medical centers. Cases are adult men and women over the age of 18 years diagnosed with first-ever ischemic stroke confirmed by a study neurologist on the basis of history, physical examination and CT or MR imaging of the brain who were enrolled within 30 days of onset of stroke symptoms. Cases exclusion criteria include: a mechanical aortic or mitral valve at the time of the index ischemic stroke, central nervous system vasculitis, bacterial endocarditis, cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL), Fabry disease, homocystinuria, mitochondrial encephalopathy with lactic acidosis and stroke-like episodes (MELAS), or sickle cell anemia. Stroke severity at enrollment was assessed using the NIH Stroke Scale with the diagnostic evaluation including head CT (95%) or MRI (83%), electrocardiography (92%), cervical arterial imaging (86%), and echocardiography (74%). Medical records from all cases were centrally reviewed by a vascular neurology committee and assigned ischemic stroke subtype diagnoses according to criteria from the Trial of ORG10172 (TOAST), the Oxfordshire Community Stroke Project, and the BaltimoreWashington Young Stroke Study.

# The MGH Genes Affecting Stroke Risk and Outcome Study (MGH-GASROS)

Cases were all consecutive patients aged 18 years 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, as well as on the inpatient Medical and Vascular Surgical services from January 2003 to July 2008. Only patients of European ancestry (confirmed by principal component analysis using genome-wide SNP data) were included in the present analysis. Ischemic 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. Patients with specific vascular disorders (vasculitis, subacute bacterial endocarditis, fibromuscular dysplasia, vasospasm) were excluded from the study. All subjects were evaluated by a neurologist upon presentation and provided informed consent. Clinical and laboratory data were collected during the admission for qualifying ischemic stroke event. Diagnostic work-up included: head CT (100%), brain MRI (90%), cervical and intracranial vessel imaging using CTA or MRA (75%), carotid and/or transcranial ultrasound (24%), echocardiography (86%), and Holter monitoring (16%). Controls were recruited among the stroke-free adults presenting to the MGH outpatient clinics and matched with the stroke cases on the basis of age, sex and ancestry information obtained from principal component analysis of GWAS data.

# Milano

This study includes consecutive Italian patients referred to Besta Institute from 2000 to 2009 with stroke and included in the Besta Cerebrovascular Diseases Registry (CEDIR). Ischemic stroke cases, first ever or recurrent, confirmed on brain imaging, were selected for this study. All cases were of self-reported Caucasian ancestry and had clinically relevant diagnostic workup performed. All cases were phenotyped by an experienced stroke neurologist according to TOAST criteria, based on relevant clinical imaging and available information on cardiovascular risk factors. Controls are Italian individuals enrolled within the PROCARDIS Study, with no personal or sibling history of coronary heart disease before age 66 years.

# Wellcome Trust Case-Control Consortium 2 (WTCCC2)

The WTCCC2 samples were genotyped as part of the WTCCC 2 ischemic stroke study. Stroke cases included samples recruited by investigators at St. George's University London (SGUL) and University of Oxford in the UK and the Department of Neurology, Klinikum Großhadern, LudwigMaximilians- 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. The University of Edinburgh collection comprised 727 ischaemic stroke cass,

onsecutively collected as part of the Edinburgh Stroke Study. Cases were of self-reported Caucasian ancestry, with ischaemic stroke subtypes determined according to TOAST criteria based on relevant clinical and imaging data. 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, Klinikum Großhadern, 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/)](http://www.b58cgene.sgul.ac.uk/), and ascertained as part of the national child development study (http://www.cls.ioe.ac.uk/studies.asp?section=000100020003). 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/en/english.html). 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.

# VISP

The VISP trial (P.I. James Toole, MD, Wake Forest University School of Medicine (WFU); R01

NS34447) was a multi-center, double-blind, randomized, controlled clinical trial that enrolled patients aged 35 or older with Homocysteine 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 dose folic acid, vitamin B6 and vitamin B12 reduced recurrent cerebral infarction (1° endpoint), and nonfatal myocardial infarction (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. Enrollment 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.

Control data for comparison with VISP stroke cases were obtained through the database of genotypes and phenotypes (dbGAP) High Density SNP Association Analysis of Melanoma: Case-Control and Outcomes Investigation (phs000187.v1.p1; R01CA100264, 3P50CA093459, 5P50CA097007, 5R01ES011740, 5R01CA133996, HHSN268200782096C; PIs Christopher Amos, Qingyi Wei, Jeffrey E. Lee).

# WHI

The WHI Hormone Trial HT (WHI-HT) consisted of two separate clinical trials in postmenopausal women ages 50 to 79 years at baseline—a trial of combined estrogen and progestin (Estrogen plus Progestin or E+P) in women who had an intact uterus at baseline (n=16,608) and a trial of estrogen (Estrogen Alone or E-Alone) in women who had a prior hysterectomy at baseline (n=10,739). Postmenopausal women who gave written informed consent were enrolled in the WHI at 40 clinical centers in the United States. Exclusions for safety reasons included prior diagnosis of breast cancer or other cancers within the past 10 years (except nonmelanoma skin cancer). Women with systolic blood pressure (SBP) of 200 mm Hg or higher or diastolic blood pressure (DBP) of 105 mm Hg or higher were advised to see their physician within a specified period depending on blood pressure level and were temporarily excluded from the clinical trials until their blood pressure was determined to be under control. Stroke diagnosis requiring and/or occurring during hospitalization was based on rapid onset of a neurological deficit attributable to an obstruction or rupture of an arterial vessel system. Hospitalized incident stroke events were identified by semiannual questionnaires and adjudicated following medical record review, which occurred both locally and centrally. Ischemic strokes were further classified by the central neurologist adjudicators according to the Oxfordshire and Trial of Org 10172 Acute Stroke Trial

(TOAST)

http://jama.jamanetwork.com.offcampus.lib.washington.edu/article.aspx?articleid=196626 - REFJOC30333-25 criteria to examine stroke subtypes. The TOAST classification focuses on the presumed underlying stroke mechanism and requires detailed investigations (such as brain computed tomography, magnetic resonance imaging, angiography, carotid ultrasound, and echocardiography).

**Individual Study Description – Replication Analysis**

# CADISP

The CADISP study includes subjects with ischemic stroke without cervical artery dissection. They were recruited from the same centers as cervical artery dissection patients for a specificity analysis. These were cases with a diagnosis of ischemic stroke, in which dissection had been formally ruled out according to CADISP inclusion criteria. Non-dissection ischemic stroke cases were frequency-matched on age (by 5-year intervals) and sex with dissection patients.

A total of 658 non-dissection ischemic stroke cases were included. We excluded 19 cases due to unavailability of geographically matched healthy controls, or due to non-European origin; of the remaining 639 ischemic stroke cases, 613 individuals had good quality DNA available and were genotyped at the Centre National de Génotypage CNG. Of these, a total of 555 non-dissection ischemic stroke cases aged < 60 years, who were successfully genotyped and met genotyping quality control criteria, were available for the METASTROKE analysis.

# Leuven

Cerebral ischemia, defined as a clinical stroke with imaging confirmation or a TIA with a new ischemic lesion on diffusion weighted MRI, who were admitted to the Stroke Unit of the University Hospitals in Leuven. All patients underwent brain imaging (MRI in 91% of patients, CT in the remainder) and a standardized protocol including carotid ultrasound or CT angiography and cardiac examination (echocardiography and Holter monitoring) in all patients.

# Lund Stroke Register, Sweden

Lund Stroke Register (LSR) since 2001 continuously enrolls patients aged 18 and older with firstever stroke, living in the primary uptake area of Skåne University Hospital, Lund. The study is mainly hospital-based but has a good coverage of the whole geographical population. All included patients are examined with CT/MR or autopsy of the brain. When clinically indicated, the patients are examined with ultrasound imaging of carotid arteries, echocardiography, and angiography. For the replication part of this study, first-ever ischemic stroke patients from LSR between 2012 and 2013 were included as well as control subjects from the same geographical uptake area recruited between 2001 and 2002. All patients were assessed by a neurologically trained physician regarding stroke type. The control subjects were individuals without stroke, randomly selected from the official Swedish Population Register and matched for age and gender to the stroke patients. Presence of hypertension, diabetes mellitus, heart disease, smoking and hypercholesterolemia was registered for patients and control subjects. Informed consent was obtained from all individuals or when they were not able to respond from their next-of-kin. The study was approved by the Ethics Committee of Lund University.

# Münster (Westphalian Stroke Cases and Controls from the Dortmund Health Study, Germany)

In Westphalia patients were recruited through hospitals participating in the regional Westphalian Stroke Register, located in the west of the country. For this analysis ischemic stroke patients recruited during the period 2000-2005 were included. The register’s standardized patient documentation form included major stroke type and severity, comorbidities, diagnostic and therapeutic details of the treatment process. Ischemic stroke was further subtyped according to the TOAST classification by the documenting physician. Patients who had experienced a transient ischaemic attack or a haemorrhagic stroke were excluded from this analysis.

Controls were drawn from the population based, prospective Dortmund Health Study (DHS), conducted in the same region. Aim of the study was the assessment of the prevalence and incidence of different headache types as well as other chronic health conditions and to analyze their consequences on daily activities of those affected. Participants were randomly drawn from the registration office in the city of Dortmund, within the age range 25 to 74 years and stratified by gender. They participated in a face to face health interview and several physical examinations in the baseline assessment in 2003/4. Cases with a history of stroke (n=28) were excluded from this analysis.

Both studies were approved by the ethics committee of the University of Muenster. All participants gave their informed consent.

# Poland:Krakow

Patients were recruited in the stroke unit of the Jagiellonian University in Krakow, Poland (a single-center study). All stroke patients and controls were >18 years of age and were white. All patients had clinically relevant diagnostic workup performed, including brain imaging with computed tomography (CT) (100%) and/or magnetic resonance imaging (MRI) (8%) as well as ancillary diagnostic investigations including duplex ultrasonography of the carotid and vertebral arteries (85.2%), echocardiography (54.8%). MR-angiography, CT-angiography Holter monitoring, transesophageal echocardiography and blood tests for hypercoagulability were performed were indicated. Patients were classified into etiologic subtypes according to the Trial of Org 10172 in Acute Stroke Treatment (TOAST).

The control group included unrelated subjects taken from the population of southern Poland. Control subjects had no apparent neurological disease based on the findings in a structured questionnaire and a neurological examination.

The study was approved by local research ethics committees and informed consent was obtained from all participants.

# Portugal

The Portuguese stroke cases and controls used in this study were ascertained and collected as described previously. Briefly, unrelated patients with a clinical diagnosis of IS, who were under the age of 65 at stroke onset, were recruited through Neurology and Internal Medicine Departments throughout Portugal. All patients were seen, and all neuroradiology tests were reviewed by study neurologists. Trauma, tumors, infection, and other causes of neurological deficit were excluded. Since stroke is a late-onset disease, the control group was selected from a group of healthy volunteers with a higher mean age than the case group, thus minimizing the chances for mis-classification as “stroke-free”. Control individuals were verified to be free of stroke by direct interview before recruitment, but no brain imaging studies were performed. Part of the control samples were requested from Biobanco-IMM, Lisbon Academic Medical Center, Lisbon, Portugal. All participants were adults of Portuguese Caucasian origin. The research protocol was approved by the Ethics Committees of participating institutions, and all participants provided informed consent.

# Risk Assessment of Cerebrovascular Events (RACE) Study, Pakistan

RACE is a retrospective case-control study designed to identify and evaluate genetic, lifestyle and biomarker determinants of stroke and its subtype in Pakistan. Samples were recruited from six hospital centres in Pakistan. Cases were eligible for inclusion in the study if they: (1) are aged at least 18 years; (2) presented with a sudden onset of neurological deficit affecting a vascular territory with sustained deficit at 24 hours verified by medical attention within 72 hours after onset (onset is defined by when the patient was last seen normal and not when found with deficit);

(3) the diagnosis was supported by CT/MRI; and (4) presented with a Modified Rankin Score of < 2 prior to the stroke. TOAST and Oxfordshire classification systems were used to sub-phenotype all stroke cases. Control participants were individuals enrolled in the Pakistan Risk of Myocardial Infarction Study (PROMIS), a case/control study of acute MI based in Pakistan. Controls in PROMIS were recruited following procedures and inclusion criteria as adopted for RACE cases. In order to minimize any potential selection biases, PROMIS controls selected for this stroke study were frequency matched to RACE cases based on age and gender and were recruited in the following order of priority: (1) non-blood related or blood related visitors of patients of the outpatient department; (2) non-blood related visitors of stroke patients; (3) patients of the out-patient department presenting with minor complaints.

# USA

Ischemic case samples were collected from the Massachusetts General Hospital - Genes Affecting Stroke Risk and Outcomes Study (MGH-GASROS, Funded by NIH-NINDS P50NS051343), and the Greater Cincinnati/Northern Kentucky Stroke Study (NIH R01 NS30678). Control samples were collected from the Greater Cincinnati Foundation Grant (Cincinnati Control Cohort) and Genetic and Environmental Risk Factors for Hemorrhagic Stroke (GERFHS, funded by NIH R01 NS036695) study. For all samples, consent was obtained from a study subject and blood was drawn. DNA was extracted from the blood sample and quantified to calculate the amount of DNA available for genotyping. DNA was stored at -80 degrees until genotyped.