UNIVERSIDADE DE LISBOA – FACULDADE DE CIÊNCIAS

DEPARTAMENTO DE BIOLOGIA VEGETAL

GENETIC FACTORS INVOLVED IN STROKE SUSCEPTIBILITY AND IN OUTCOME AT THREE MONTHS

Helena Isabel Gomes Pires Manso

DOUTORAMENTO EM BIOLOGIA

(Especialidade em Genética)

2011

UNIVERSIDADE DE LISBOA – FACULDADE DE CIÊNCIAS

DEPARTAMENTO DE BIOLOGIA VEGETAL

GENETIC FACTORS INVOLVED IN STROKE SUSCEPTIBILITY AND IN OUTCOME AT THREE MONTHS

Helena Isabel Gomes Pires Manso

Tese orientada pelo Prof. Doutor Pedro Silva (Faculdade de Ciências da Universidade de Lisboa) e pela Doutora Astrid Moura Vicente (Instituto Gulbenkian de Ciência e Instituto Nacional de Saúde Dr. Ricardo Jorge)

DOUTORAMENTO EM BIOLOGIA

(Especialidade em Genética)

2011

Abstract

Stroke is a significant cause of death and disability in developed countries. It is a multifactorial disease, resulting from the interplay between genes and well-known life-style/environmental risk factors. Numerous studies have attempted to identify the genetic risk factors predisposing to stroke, but few have investigated the genetic factors involved in stroke outcome. This work aimed at the identification of genes contributing to stroke and influencing patient's outcome after three months. Four inflammatory genes (*IL1B*, *IL6*, *MPO* and *TNF*) and two genes involved in the nitric oxide metabolism (*NOS1* and *NOS3*) were tested for association with stroke. The results suggest that the *IL6* and *MPO* genes influence stroke susceptibility through independent effects and non-additive interactions. Furthermore, they provided novel evidence for the involvement of the *NOS1* gene in stroke susceptibility. Several studies have shown the important impact of oxidative stress, inflammation, angiogenesis, neurogenesis, neurovascular damage and neurovascular remodeling for stroke-associated brain damage and/or stroke recovery. Association analyses were thus carried out to assess the role of candidate genes involved in inflammatory processes (*IL1B*, *IL6*, *MPO* and *TNF*) and oxidative stress (*NOS1* and *NOS3*), as well as matrix metalloproteinase genes (*MMP2* and *MMP9*) and growth factor genes (*BDNF*, *FGF2* and *VEGFA*) in patient's outcome at three months. *MMP2* genetic variants were found associated with patient's outcome, and the results also indicate that two epistatic interactions between the *BDNF* and *FGF2* genes and between the *FGF2* and *VEGFA* genes influence this phenotype. A genome-wide association study was performed in stroke outcome using DNA pooled samples to provide novel insights into the mechanisms involved in stroke recovery. The *BBS9* and *GLIS3* genes were found associated with patient's outcome at three months. Taken together, these results suggest that stroke susceptibility and outcome are modulated by a combination of main gene effects and gene-gene interactions, independently of stroke risk factors and/or severity parameters, highlighting the complexity of mechanisms predisposing to stroke and influencing recovery afterwards.

Keywords: stroke, association study, candidate genes, epistatic interactions, susceptibility, outcome.

Resumo

O Acidente Vascular Cerebral (AVC) é uma das principais causas de morte e incapacidade permanente nos países desenvolvidos, tendo importantes consequências económicas e sociais. O AVC é considerado uma doença complexa, que resulta de uma acção combinada entre genes e factores de risco ambientais ou de estilo de vida. Muitos estudos foram já levados a cabo com o intuito de identificar os factores genéticos de risco para o AVC, mas os resultados têm sido inconsistentes. A maioria desses estudos analisou o papel de genes candidatos; mais recentemente, foram também realizados estudos de associação ao nível de todo o genoma. Ao contrário da susceptibilidade, poucos estudos procuraram identificar quais os factores genéticos envolvidos na recuperação após um AVC.

Este trabalho de doutoramento pretendeu assim identificar genes que contribuam para a susceptibilidade ao AVC e que influenciem o estado de incapacidade funcional (*outcome*) do doente ao fim de três meses de recuperação. Para isso, foram utilizadas duas estratégias distintas. A primeira estratégia consistiu na análise do papel de genes candidatos nos dois fenótipos referidos anteriormente, genes esses que tinham sido escolhidos com base na sua função, nos resultados de estudos de associação anteriores e/ou nos resultados de estudos com modelos animais de AVC, factores de risco ou condições médicas associadas. Foi ainda realizado um estudo de associação ao nível de todo o genoma com o objectivo de identificar factores genéticos que influenciem o *outcome* do doente. Nesta segunda estratégia não existia uma hipótese *a priori* relativamente ao papel de um determinado gene no fenótipo, correspondendo por isso a uma análise não enviesada.

Vários estudos têm sugerido que a inflamação e o stress oxidativo desempenham um papel relevante na susceptibilidade ao AVC. Factores de risco já conhecidos, como aterosclerose, diabetes, obesidade e hipertensão, estão associados a um perfil inflamatório elevado. Além disso, é também conhecida a importância das enzimas sintase do óxido nítrico (NOS) para a aterosclerose e a regulação da pressão sanguínea. Atendendo a isso, foi testada a associação de quatro genes inflamatórios (*IL1B*, *IL6*, *MPO* e *TNF*) e de dois genes envolvidos no metabolismo do óxido nítrico (*NOS1* e *NOS3*) com o risco de AVC. Foram encontradas associações de variantes genéticas nos

genes inflamatórios *IL6* e *MPO* com a doença, assim como de uma interacção epistática entre eles contribuindo para o risco. Isto sugere que os dois genes influenciam a susceptibilidade ao AVC através de efeitos independentes e de efeitos de interacção não aditivos. Os resultados mostram ainda uma associação entre o AVC e variantes genéticas no gene *NOS1*, que codifica uma das isoformas de NOS. A análise de genes candidatos sugere assim que efeitos independentes dos genes inflamatórios ou de stress oxidativo *IL6*, *MPO* e *NOS1*, e efeitos não aditivos resultantes de interacções entre os genes *IL6* e *MPO* têm um impacto na susceptibilidade ao AVC. Estes resultados são compatíveis e reforçam as observações feitas em estudos *in vitro* e *in vivo* relativamente ao papel da inflamação e do stress oxidativo nesta doença.

É igualmente reconhecido o importante impacto de stress oxidativo, inflamação, angiogénese, neurogénese, dano e remodelação neurovasculares na lesão cerebral associada ao AVC e/ou na recuperação funcional dos doentes. Assim, foram também levados a cabo estudos de associação para avaliar o papel de genes candidatos envolvidos em processos inflamatórios (*IL1B*, *IL6*, *MPO* e *TNF*) e de stress oxidativo (*NOS1* e *NOS3*), assim como de genes das metaloproteinases da matriz (MMPs) (*MMP2* and *MMP9*) e de genes de factores de crescimento (*BDNF*, *FGF2* and *VEGFA*), no *outcome* do doente após três meses de recuperação. Relativamente ao gene *MMP2*, os resultados mostram a associação entre variantes genéticas deste gene e o *outcome* do doente após três meses de recuperação. Vários estudos anteriores demonstraram que a ruptura da barreira hemato-encefálica associada ao AVC está relacionada com a expressão e activação de MMPs, levando a hemorragia, edema e morte celular. No entanto, outros estudos sugerem que a actividade destas proteínas pode ser benéfica na angiogénese e na remodelação neurovascular em fases tardias de recuperação, o que poderá contribuir para a recuperação funcional do doente. Neste trabalho de doutoramento foram ainda identificadas duas interacções epistáticas entre os genes *BDNF* e *FGF2* e entre os genes *FGF2* e *VEGFA* em associação com o *outcome* do paciente. Estes três genes codificam factores de crescimento que são partilhados pelo sistema nervoso e vascular e que afectam a homeostasia e desenvolvimento dos dois sistemas. Os factores de crescimento têm importantes funções ao nível da angiogénese, neurogénese e protecção neuronal, influenciando o estado neurológico dos doentes e a recuperação dos mesmos após o AVC. Tendo em conta estes resultados, a análise de genes candidatos sugere que o gene *MMP2* e interacções epistáticas entre os genes *BDNF* e *FGF2*, e entre os genes *FGF2* e *VEGFA* têm um impacto no *outcome* do

doente ao fim de três meses de recuperação. Estes resultados são também compatíveis com estudos *in vitro* e *in vivo* que tinham demonstrado previamente a importância das MMPs e dos factores de crescimento na recuperação após um AVC.

Foi feito um rastreio genómico com o objectivo de identificar factores genéticos que influenciem o *outcome* do doente. Para isso foi testada a associação de mais de 250 mil polimorfismos, localizados ao longo de todo o genoma, com o *outcome* do doente. Uma das vantagens desta estratégia é a de possibilitar a descoberta de novos mecanismos envolvidos neste fenótipo. Para realizar este estudo de uma forma economicamente eficiente foram analisados conjuntos (*pools*) de amostras de DNA de doentes. Após a identificação dos marcadores mais importantes com base em quatro estratégias distintas, esses resultados foram validados por genotipagem individual. Este estudo permitiu a identificação de uma associação entre os genes *BBS9* e *GLIS3* e o *outcome* do doente. Sabe-se que pacientes com a síndrome *Bardet-Biedl* têm mutações no gene *BBS9*. A obesidade é uma das manifestações clínicas mais importantes desta síndrome e foi demonstrado que, após um AVC, os danos neurológicos e as lesões cerebrais são menores em ratinhos em regime de restrição calórica do que em ratinhos sem esta restrição, o que sugere que o excesso de calorias consumidas/obesidade poderá influenciar o *outcome* após AVC. Relativamente ao gene *GLIS3*, que codifica um factor de transcrição, tinham sido anteriormente reportadas associações deste gene com a diabetes. Esta doença é um factor de risco para o AVC e está também associada com maior risco de morte e incapacidade funcional de doentes quando o AVC ocorre. Apesar de não ser imediatamente perceptível qual o papel dos genes *BBS9* e *GLIS3* no *outcome* dos doentes, é possível que estes genes tenham uma influência indirecta nesse fenótipo através de um efeito na diabetes e obesidade.

Em conclusão, os resultados obtidos durante este trabalho de doutoramento sugerem que tanto a susceptibilidade ao AVC como o *outcome* do doente são modulados por uma combinação de efeitos de genes independentes e de interacções entre genes. Isto indica que os mecanismos envolvidos na predisposição a esta doença e na recuperação posterior dos doentes poderão ser bastante complexos. É de salientar que, com este trabalho de doutoramento, o número de genes candidatos analisados, até ao momento, na área da genética do *outcome* praticamente duplicou. Além disso, pela primeira vez, foi realizado um estudo de associação ao nível de todo o genoma com o *outcome* do doente. No futuro, será desejável aumentar a dimensão da amostra, em especial dos doentes com informação sobre recuperação. Deverão, ainda, ser realizados

estudos para identificar as variantes genéticas causais que estão na base das associações encontradas com a susceptibilidade ao AVC e com o *outcome* do doente.

Palavras-chave: Acidente Vascular Cerebral, estudo de associação, genes candidatos, interacções epistáticas, susceptibilidade, recuperação.

Acknowledgements

First, I would like to acknowledge my supervisor Dr. Astrid M. Vicente for giving me the opportunity to work with her and to undertake this project. Her support, encouragement, suggestions and advices over these years were of great value for my work. I would also like to thank her for giving me the opportunity to attend four international meetings in the field of human genetics, where I could present my work, establish new scientific collaborations, be aware of the most recent advances in this research area and learn many things that could be applied to my work.

To Professor Pedro Silva, my academic supervisor, for his helpful advices and careful revision of this thesis.

To Dr. Sofia A. Oliveira and Dr. Henrique Teotónio, my thesis committee, for their suggestions and advices.

To our collaborators, Dr. Sofia A. Oliveira and her group members. In particular, I would like to thank Tiago Krug for his suggestions and for his friendship.

To Professor Lisete Sousa for her advices on statistical analyses.

To all the Instituto Gulbenkian de Ciência (IGC) staff that help me in this work, especially Isabel Marques, João Costa and João Garcia for their availability, professionalism and kindness.

To all stroke patients and controls whose contribution made this work possible.

To Fundação para a Ciência e Tecnologia for funding.

To each one of my colleagues that during these PhD years worked with me, in particular Catarina Correia, Filipa Sequeira, Inês Sousa, João Sobral, Margarida Espada and Tiago Magalhães. Thanks for your friendship and companionship, for your support, advices,

and suggestions, and for making this group such a wonderful and pleasant place to work. A special thanks to João for his help in the lab work, and to Tiago for teaching me R.

To my friends, in particular Joana Gonçalves, Ângela Gaspar, Carina Sousa, Helena Albuquerque, Daniela Santos, Miguel Jorge, Nuno Cláudio, Soraia Oliveira, Carlos Serra and Pedro Faria.

To Ricardo for his love, support and encouragement in the last months of my PhD work.

To my family. This thesis is dedicated to my parents and brother for their love, unconditional support and precious advices, and for always being there for me.

Table of contents

Abstract iii
Resumo v
Acknowledgements ix
Table of contents xi
Statement of work xiii
List of manuscripts xiv
Abbreviations xvi
Chapter 1. Introduction 1
1.1 Stroke 3
1.1.1 Definition 3
1.1.2 Incidence and prevalence 3
1.1.3 Clinical aspects, diagnosis and pathophysiology 4
1.2 Etiology 7
1.2.1 Stroke risk factors 7
1.2.2 Evidence for genetic liability and the multifactorial model for stroke 8
1.2.3 Endophenotypes 10
1.3 Strategies for genetic analyses in stroke 11
1.3.1 Linkage studies 11
1.3.1.1 Single gene disorders 12
1.3.1.2 Common stroke 13
1.3.2 Association studies 14
1.3.2.1 Candidate genes 16
1.3.2.2 Genome-wide association studies 19
1.3.3 Gene-gene interactions 22
1.4 Stroke outcome and functional recovery 24
1.5 Objectives and thesis outline 28
References 30
Chapter 2. Matrix metalloproteinase genes 43
Chapter 3. Inflammatory genes 65
Chapter 4. Growth factor genes 95

Chapter 5. Nitric oxide synthase genes 111
Chapter 6. Genome-wide association study in stroke outcome................................ 143
Chapter 7. General discussion and final considerations 175
7.1 General discussion 177
7.2 Concluding remarks and future perspectives 185
References.. 191
Appendix I. Other manuscript contributions 213

Statement of work

The individuals participating in this study were recruited at several hospitals in mainland Portugal and Madeira Islands. Clinical assessment and collection of blood samples from stroke patients and controls were performed at those hospitals. 556 patients and 446 controls were recruited at Instituto Nacional de Saúde Dr. Ricardo Jorge (INSA) in the context of earlier studies. The remaining individuals (116 patients and 84 controls) were recruited at Instituto Gulbenkian de Ciência (IGC). DNA extraction from blood samples was performed at INSA and IGC by specialized technicians.

The experiments were designed, performed and analyzed by the author with contributions/suggestions from collaborators. The author participated in the genotyping, statistical analyses and/or discussion of the results of the manuscripts where she contributed as a co-author (see Appendix I).

List of manuscripts

Published as first author:

- **Manso H**, Krug T, Sobral J, Albergaria I, Gaspar G, Ferro JM, Oliveira SA, Vicente AM (2010) Variants of the Matrix Metalloproteinase-2 but not the Matrix Metalloproteinase-9 genes significantly influence functional outcome after stroke. BMC Med Genet, 11:40.

- **Manso H**, Krug T, Sobral J, Albergaria I, Gaspar G, Ferro JM, Oliveira SA, Vicente AM (2011) Variants in the inflammatory *IL6* and *MPO* genes modulate stroke susceptibility through main effects and gene-gene interactions. J Cereb Blood Flow Metab [Epub ahead of print].

Submitted:

- **Manso H**, Krug T, Sobral J, Albergaria I, Gaspar G, Ferro JM, Oliveira SA, Vicente AM (2011) Evidence for genetic interactions among growth factor genes in stroke outcome (Submitted)

- **Manso H**, Krug T, Sobral J, Albergaria I, Gaspar G, Ferro JM, Oliveira SA, Vicente AM (2011) Variants within the nitric oxide synthase 1 gene are associated with stroke susceptibility (Submitted)

Unpublished results:

- **Manso H**, Krug T, Sobral J, Albergaria I, Gaspar G, Ferro JM, Oliveira SA, Vicente AM (2011) A genome-wide association study using DNA pooling identifies *BBS9* and *GLIS3* as novel *loci* influencing patient's outcome after stroke

Published as co-author:

- Rosa A, Fonseca BV, Krug T, **Manso H**, Gouveia L, Albergaria I, Gaspar G, Correia M, Baptista MV, Simões RM, Pinto AN, Taipa R, Ferreira C, Fontes JR, Silva MR, Gabriel JP, Matos I, Lopes G, Ferro JM, Vicente AM, Oliveira SA (2008). Mitochondrial haplogroup H1 is protective for ischemic stroke in Portuguese patients. BMC Medical Genetics, 9:57.

- Domingues-Montanari S, Fernández-Cadenas I, del Rio-Espinola A, Corbeto N, Krug T, **Manso H**, Gouveia L, Sobral J, Mendioroz M, Fernández-Morales J, Alvarez-Sabin J, Ribó M, Rubiera M, Obach V, Martí-Fàbregas J, Freijo M, Serena J, Ferro JM, Vicente AM, Oliveira SA, Montaner J (2010). Association of a genetic variant in the *ALOX5AP* gene with higher risk of ischemic stroke – a case-control, meta-analysis and functional study. Cerebrovasc Disease, 29:528-537.

- Krug T, **Manso H**, Gouveia L, Sobral J, Xavier JM, Albergaria I, Gaspar G, Correia M, Baptista MV, Simões RM, Pinto AN, Taipa R, Ferreira C, Fontes JR, Silva MR, Gabriel JP, Matos I, Lopes G, Ferro JM, Vicente AM, Oliveira SA (2010). Kalirin: a novel genetic risk factor for ischemic stroke. Hum Genet 127:513-523.

Submitted as co-author:

- Krug T, Gabriel JP, Taipa R, Gouveia L, Fonseca BV, **Manso H**, Albergaria I, Gaspar G, Ferro JM, Vicente AM, Silva MR, Matos I, Lopes G, Oliveira SA (2011) Tetratricopeptide repeat domain 7B emerges as a novel risk factor for ischemic stroke following a multifactorial approach. (Submitted)

Abbreviations

ACE – angiotensin-converting enzyme

AGT – angiotensinogen

ALOX5AP – arachidonate 5-lipoxygenase-activating protein

APJ – apelin receptor protein

APLNR – apelin receptor

APOE/Apoe – apolipoprotein E

AVC – acidente vascular cerebral

Aβ-PP – amyloid beta precursor protein

BBB – blood-brain barrier

BBS – Bardet-Biedl syndrome

BBS9 – Bardet-Biedl syndrome 9

BDNF – brain-derived neurotrophic factor

BI – Barthel Index

CADASIL – cerebral autosomal dominant arteriopathy with subcortical infarcts and leucoencephalopathy

CAV1/Cav1 – caveolin-1

CBS – cystathionine-β-synthase

CELSR1 – cadherin, EGF LAG seven-pass G-type receptor 1

CI – confidence interval

CNS – Central Nervous System

CNV – copy number variant

COL12A1 – collagen type XII alpha 1

CRP – C-reactive protein

CT – computed tomography

CVC – cross-validation consistency

ECM – extracellular matrix

eNOS – endothelial nitric oxide synthase

F12 – coagulation factor XII

F13A1 – coagulation factor XIII, A1 polypepetide

F2 – coagulation factor II

F5 – coagulation factor V

F7 – coagulation factor VII

FCT – Fundação para a Ciência e a Tecnologia

FDA – Food and Drug Administration

FDR – false discovery rate

FGA – fibrinogen alpha chain

FGB – fibrinogen beta chain

FGF2 – fibroblast growth factor 2

GLIS3 – GLIS family zinc finger 3

GOS – Glasgow Outcome Scale

GP1BA – glycoprotein Ib, alpha polypeptide

GSTO1 – glutathione S-transferase omega 1

GWAS – genome-wide association studies

HBGF-2 – heparin-binding growth factor 2

HbS – haemoglobin S

HCHWA-D – hereditary cerebral haemorrhage with amyloidosis - the Dutch type

HCHWA-I – hereditary cerebral haemorrhage with amyloidosis - the Icelandic type

HDL – high-density lipoprotein

HPGDS – hematopoietic prostaglandin D synthase

HWE – Hardy-Weinberg equilibrium

I/D – insertion/deletion

ICH – intracerebral hemorrhage

IG – information gain

IGC – Instituto Gulbenkian de Ciência

IL1B/IL1-beta – interleukin 1 beta

IL6/IL-6 – interleukin 6

IMT – intima-media thickness

iNOS – inducible nitric oxide synthase

INSA – Instituto Nacional de Saúde Dr. Ricardo Jorge

IPF1 – insulin promoter factor 1

ITGA2 – integrin alpha 2

ITGA2B – integrin alpha 2b

ITGB3 – integrin beta 3

KO – knockout

LD – linkage disequilibrium

LDL – low-density lipoprotein

LDL-R – low-density lipoprotein receptor

LPL – lipoprotein lipase

MAF – minor allele frequency

MCAO – middle cerebral artery occlusion

MDR – multifactor-dimensionality reduction

MELAS – syndrome of mitochondrial myopathy, encephalopathy, lactic acidosis and stroke-like episodes

MMP2/MMP-2 – matrix metalloproteinase 2

MMP9/MMP-9 – matrix metalloproteinase 9

MMP – matrix metalloproteinase

MPO – myeloperoxidase

MR – magnetic resonance

MRI – magnetic resonance imaging

mRS – modified Rankin Scale

MTHFR – methylenetetrahydrofolate reductase

NIHSS – National Institute of Health Stroke Scale

NINJ2 – ninjurin 2

nNOS – neuronal nitric oxide synthase

NO – nitric oxide

NOS – nitric oxide synthase

NOS1/Nos1 – nitric oxide synthase 1

NOS3/Nos3 – nitric oxide synthase 3

OR – Odds ratio

OR6S1 – olfactory receptor 6S1

OR-MDR – Odds ratio-based MDR

PDE4D – phosphodiesterase 4D

PDLIM5 – PDZ and LIM domain 5

PITX2 – paired-like homeodomain 2

PKCη – protein kinase C η

PLAT – plasminogen activator, tissue

PLZF – promyelocytic leukemia zinc finger

PNS – Peripheral Nervous System

PON1 – paraoxonase 1

POU5F1 – POU domain class 5 transcription factor 1

PTHB1 – parathyroid hormone-responsive B1

QC – quality control

RAS – relative allele signal

SD – standard deviation

SERPINE1 – serpin peptidase inhibitor, clade E, member 1

sGCalpha1 – alpha 1 subunit of soluble guanylate cyclase

SIS – Stroke Impact Scale

SNP – single nucleotide polymorphism

SNPSpD – single nucleotide polymorphism spectral decomposition

SpD – spectral decomposition

TBA – testing balanced accuracy

TIA – transient ischemic attack

TIMP – tissue inhibitor of metalloproteinases

TNF – tumor necrosis factor

TOAST – Trial of ORG 10172 in Acute Stroke Treatment

tPA – tissue plasminogen activator

UTR – untranslated region

VDR/RXR – vitamin D hormone receptor/retinoid X receptor

VEGFA/VEGF-A – vascular endothelial growth factor A

VWF – von Willebrand factor

WMH – white matter hyperintensities

yrs – years

ZFHX3 – zinc finger homeobox 3

Chapter 1. Introduction

1.1 Stroke

1.1.1 Definition

Stroke is a major health problem in developed countries, with a very high incidence and mortality. Furthermore, it is also a major cause of significant disability. With increasing life expectancy, this public health problem tends to worsen, with important economic and social consequences. Stroke is defined by the World Health Organization [1988] as “rapidly developing clinical signs of focal (or global) disturbance of cerebral function, with symptoms lasting 24h or longer, or leading to death, with no apparent cause other than of vascular origin”. Stroke represents a collection of different processes, with different clinical phenotypes, etiological mechanisms and risk factor profiles that ultimately result in cellular death in the brain. Different stroke subtypes have also different degrees of heritability and genetic risk factors profiles [Markus 2003].

1.1.2 Incidence and prevalence

Over the past four decades, age-adjusted stroke incidence rates have decreased in high income countries, but have increased in low to middle income countries, exceeding the rate observed in the more developed countries [Feigin et al. 2009]. The annual incidence of stroke is estimated to be about 1.1 million in the European Union, Switzerland, Iceland and Norway, and approximately 6 million people in these countries are stroke survivors [Truelsen et al. 2005]. In 2004, the cost of stroke for the European society was approximately 22 billion euros, but it might be larger since estimations were based on incidence data [Andlin-Sobocki et al. 2005]. In the same year, the total cost of stroke in Portugal was approximately 853 million euros, and this country had one of the highest stroke incidence and prevalence rates in Europe (people aged 25 years or older) [Andlin-Sobocki et al. 2005; Truelsen et al. 2005]. Annual incidence rates in rural and urban areas were 2.02 (95% confidence interval [CI]: 1.69-2.34) and 1.73 (95% CI: 1.53-1.92), respectively, per 1,000 inhabitants [Correia et al. 2004].

After stroke, one third of surviving patients remain disabled and one in seven requires institutional care [Hankey et al. 2002]. Moreover, among 30-day survivors of first-ever stroke, only about 50% survive 5 years [Hankey et al. 2002]. The cumulative risk of a recurrent stroke is about 40% during the first 10 years after a first-ever event; and the case fatality of a recurrent stroke is almost twice that of a first-ever stroke [Hardie et al. 2004].

1.1.3 Clinical aspects, diagnosis and pathophysiology

Stroke can be broadly divided into ischemic and hemorrhagic, with the latter comprising intracerebral hemorrhage (ICH) and subarachnoid hemorrhage. It is believed that ischemic and hemorrhagic stroke result from both common and different determinants [Humphries and Morgan 2004]. Since ICH corresponds to approximately 70% of hemorrhagic strokes [Lloyd-Jones et al. 2009], *hemorrhagic stroke* will be used in this thesis as a synonym of *ICH* unless otherwise specified. According to Amarenco et al. [2009], stroke should also include cerebral venous thrombosis and spinal cord stroke. Stroke has several sudden symptoms including numbness or weakness of the face, arm or leg, especially in one side of the body; severe headache with no apparent cause; confusion, trouble speaking or understanding, and trouble seeing in one or both eyes.

The only reliable method to distinguish ischemic stroke from ICH is brain imaging, including computed tomography (CT) and magnetic resonance (MR). CT is widely available and can be performed quickly, which are two important advantages of this method [Saenger and Christenson 2010]. Although CT is often normal in hyperacute ischemic stroke, it is particularly valuable to exclude the existence of hemorrhages and tumors [Warlow et al. 2003]. Early CT can identify ICH, but cannot reliably distinguish between primary ICH and hemorrhagic transformation of an ischemic stroke [Davenport and Dennis 2000]. MR is more sensitive than CT and can help the diagnosis in patients who delay seeking treatment, although some infarctions are equally not visible [Davenport and Dennis 2000; Saenger and Christenson 2010; Warlow et al. 2003]. In contrast with CT, MR is not readily available in some countries and it can not be safely used in many acutely ill patients [Davenport and Dennis 2000; Warlow et al. 2003].

Many studies have investigated the potential use of several blood-based protein biomarkers for stroke risk prediction, diagnosis or prognosis [Baird 2006; Saenger and Christenson 2010]; some of these biomarkers can be specially useful for stroke patients that have normal or ambiguous results in brain imaging. Lipoprotein-associated phospholipase A2 was approved by the US Food and Drug Administration (FDA) for long-term prognostic risk for stroke and coronary heart disease. Other promising biomarkers include the asymmetric dimethylarginine, matrix metalloproteinase-9 and glial fibrillary acidic protein [reviewed in Saenger and Christenson 2010].

Ischemic stroke

Ischemic stroke corresponds to 80-90% of total stroke cases [Bhatnagar et al. 2010; Lloyd-Jones et al. 2009] and is caused by a sudden interruption of a cerebral artery or, less often, by a decrease in cerebral blood flow due to stenosis [Davenport and Dennis 2000; NINDS 2004]. Ischemia refers to the loss of oxygen and nutrients for brain cells due to inadequate blood flow. If neurological deficits last less than 24h, it is considered a transient ischemic attack (TIA). Although TIA is less dangerous than stroke, early administration of thrombolytic therapies is important to attenuate the short-term risk of ischemic stroke, cardiovascular events and death that is associated with TIAs [Saenger and Christenson 2010].

In white people, approximately 50% of ischemic strokes are caused by atherothromboembolism, 25% correspond to lacunar infarcts due to occlusion of small cerebral arteries, 20% are caused by emboli from the myocardium (resulting from concurrent myocardial infarction, atrial fibrillation, mitral stenosis, etc.) and the remainder are due to rarer causes [Davenport and Dennis 2000; Saenger and Christenson 2010; Warlow et al. 2003]. A wide variety of classification schemes have been used to classify subtypes of ischemic stroke, with different weaknesses and strengths [Amarenco et al. 2009; Kirshner 2009]. Despite differences regarding nomenclature, weight placed on stroke risk factors, clinical features, and brain imaging techniques, these classification systems differentiate between strokes of large vessel (atherothrombotic), cardioembolic, small vessel (lacunar), other determined causes and cryptogenic strokes (undetermined etiology) [Amarenco et al. 2009; Kirshner 2009]. The Trial of ORG 10172 in Acute Stroke Treatment (TOAST) [Adams et al. 1993] has been the most widely used classification system.

Ischemic stroke triggers a series of events called the ischemic cascade. Cellular bioenergetic failure, excitotoxicity, oxidative stress, stress signalling, inflammation, blood-brain barrier dysfunction, hemostatic activation, apoptosis, angiogenesis, survival and gene expression are activated by cerebral ischemia, resulting in a series of biochemical, hemodynamic and neurophysiological alterations [Deb et al. 2010; Mehta et al. 2007; Mitsios et al. 2006; Saenger and Christenson 2010]. The adequate delivery of nutrients and oxygen to the cells is affected during ischemic stroke. Neurons in the infarcted core die in a short time due to necrosis. In contrast, the majority of neurons remains viable in the penumbra region for a longer period and can be saved if reperfused in time. If the ischemic state persists for an extended period of time, secondary cell death develops gradually in the penumbra region mainly due to excitotoxicity and apoptosis. Duration, severity and location of the ischemic insult influence the extent of damage [Deb et al. 2010; Mitsios et al. 2006].

Thrombolysis with recombinant tissue plasminogen activator (tPA) is the only approved therapy for acute ischemic stroke. The serine protease tPA cleaves plasminogen into active plasmin, whose primary function in plasma is fibrin digestion [Adibhatla and Hatcher 2008]. After degradation of the cross-linked fibrin, the clot becomes soluble and can suffer further proteolysis by other enzymes, thus restoring blood flow [Adibhatla and Hatcher 2008]. Only 2-8% of all stroke patients are treated with tPA due to delay in seeking medical care and an extensive list of contraindications, including prior myocardial infarction, hypertension or evidence of ICH [Kleindorfer et al. 2004; Wu and Grotta 2010]. Nevertheless, recent studies suggest that the time window for thrombolysis can be safely extended without major risk of hemorrhage or death [reviewed in Wu and Grotta 2010]. Although there is no thrombolytic alternative to tPA, major advances have occurred in stroke prevention, especially in management of atrial fibrillation and carotid stenosis, two conditions that increase stroke risk [Wu and Grotta 2010]. Clinical trials have demonstrated the efficacy of warfarin and carotid endarterectomy (in both symptomatic and asymptomatic patients with high-grade stenosis) in the treatment of atrial fibrillation and carotid artery atherosclerosis, respectively, and secondary stroke prevention [Wu and Grotta 2010].

Hemorrhagic stroke

Hemorrhagic stroke corresponds to approximately 10-15% of stroke cases and occurs due to a rupture of a cerebral blood vessel. It is more frequently fatal than ischemic stroke [Davenport and Dennis 2000; Lloyd-Jones et al. 2009; Sahni and Weinberger 2007]. Primary ICH is mainly due to small vessel disease, which is often associated with hypertension and amyloid angiopathy [Davenport and Dennis 2000; Sahni and Weinberger 2007]. Vascular abnormalities (like aneurysms and arteriovenous malformations), hemorrhagic conversion of an ischemic stroke, tumor, etc are responsible for secondary ICH [Davenport and Dennis 2000].

The outcome after ICH is influenced by the location of the injury and by the hemorrhagic volume: early fatality occurs if the hemorrhagic volume exceeds a certain limit, but delayed fatality may also happen as a consequence of the secondary brain injury that results from the hematoma [Xi et al. 2006]. The harmful effects of hemorrhagic stroke are due to hypoxia, increased intracranial pressure resulting from hematoma enlargement, which may additionally restrict cerebral blood flow, and from the chemical toxicity in brain parenchyma and vasculature caused by the blood that is released from the vessel [Deb et al. 2010].

Nowadays, there is no specific therapy to improve outcome after ICH [Sahni and Weinberger 2007]. Early hematoma removal through craniotomy revealed no benefit and, although the activated recombinant Factor VII (an antifibrinolytic agent) limits hematoma expansion, it was unable to reduce disability or mortality after stroke [Broderick 2005; Sahni and Weinberger 2007].

1.2 Etiology

1.2.1 Stroke risk factors

Age, gender and race are non-modifiable risk factors that play a role in the development of stroke. Stroke incidence increases with age and is higher in males than females but only at younger ages [Lloyd-Jones et al. 2009; Truelsen et al. 2005].

Furthermore, comparing to white people, black people have approximately twice the risk of suffering a stroke [Lloyd-Jones et al. 2009]. There is also evidence that the prevalence of stroke subtypes varies according to the race and ethnicity of stroke patients [Kirshner 2009].

The estimated stroke risk also increases with the number of modifiable risk factors that are present in an individual [Lloyd-Jones et al. 2009]. These risk factors can be changed or controlled through pharmacological or surgical interventions and life-style adjustments, as primary or secondary stroke prevention measures. Modifiable stroke risk factors include hypertension, atrial fibrillation, carotid stenosis, increased blood cholesterol, cigarette smoking, alcohol consumption, diabetes, physical inactivity and obesity [Hankey 2006; Lloyd-Jones et al. 2009]. Randomized controlled trials have shown that treating hypertension, hypercholesterolemia, carotid stenosis and atrial fibrillation reduce the incidence of ischemic stroke [Hankey 2006].

1.2.2 Evidence for genetic liability and the multifactorial model for stroke

The fact that major clinical and life-style risk factors account for approximately 60% of the population-attributable risk for stroke [Whisnant 1997] suggests the existence of other risk factors that may also contribute to this disease. Studies in twins, families and animal models provided evidence that stroke has a genetic component [Dichgans 2007]. Stroke is very common in old people, which makes it difficult to collect enough twin pairs and increases the probability of twins dying from other unrelated disease [Flossmann et al. 2004; Lloyd-Jones et al. 2009]. Therefore, conducting twin studies in stroke is challenging, and these studies have been few and with small sample sizes [Flossmann et al. 2004]. A meta-analysis showed that monozygotic twins were only 1.6 times more likely to be concordant for stroke than dizygotic twins, suggesting that the genetic influence on stroke risk is small [Flossmann et al. 2004]. The heritability for stroke death and for stroke hospitalization or stroke death was estimated as 0.32 and 0.17, respectively [Bak et al. 2002]. These studies, however, did not analyze stroke subtypes nor assessed potential confounders. In addition, many studies have shown that a positive family history of stroke is a risk factor for this disease, even when accounting for other known risk factors [reviewed in Flossmann et al. 2004]. This effect is likely to be moderate (odds ratio=1.76 [95%

confidence interval=1.70-1.90], but there was significant heterogeneity between studies [Flossmann et al. 2004]. There is also evidence that genetic factors are more important in stroke occurring early in life, with the relative influence of genetics on stroke susceptibility decreasing with age [Jerrard-Dunne et al. 2003; Jood et al. 2005; Schulz et al. 2004]. The effect of genetic factors on stroke risk may also depend on stroke subtypes [Jerrard-Dunne et al. 2003; Jood et al. 2005; Polychronopoulos et al. 2002; Schulz et al. 2004].

There are some rare Mendelian forms of stroke arising from single-gene defects that are specially important in young patients with no known clinical/life-style risk factors [Dichgans 2007; Gulcher et al. 2005; Hassan and Markus 2000]. Stroke may be the prevailing manifestation or appear in conjunction with other phenotypes [Dichgans 2007]. These disorders are frequently associated with specific stroke subtypes and the mode of inheritance is variable [Dichgans 2007; Hassan and Markus 2000]. Cerebral Autosomal Dominant Arteriopathy with Subcortical Infarcts and Leucoencephalopathy (CADASIL), Fabry’s disease, Sickle-cell disease, Marfan syndrome and syndrome of mitochondrial myopathy, encephalopathy, lactic acidosis and stroke-like episodes (MELAS) are examples of single-gene disorders with ischemic stroke. Hereditary cerebral hemorrhage with amyloidosis (HCHWA) – the Icelandic type (HCHWA-I) and the Dutch type (HCHWA-D) – are rare forms of hemorrhagic stroke arising from single-gene defects.

Although Mendelian forms of stroke are important, they only account for a small percentage of cases and classical patterns of inheritance cannot be demonstrated for most stroke patients. In contrast with monogenic forms, most stroke cases represent a complex trait that is likely to be polygenic in etiology, with many *loci* modulating different pathophysiological mechanisms [Dichgans 2007; Hassan and Markus 2000]. Genetic factors may contribute to conventional risk factors, like hypertension, to intermediate phenotypes, such as atherosclerosis, or act directly on stroke risk [Dichgans 2007; Hassan and Markus 2000]. Environmental factors, gene-gene interactions and gene-environment interactions may also act at these different levels [Dichgans 2007]. Furthermore, it is likely that different genetic and environmental factors predispose different individuals to specific subtypes. These subtypes have different degrees of heritability and risk factors profiles, and the majority of recurrent strokes have the same mechanism as the incident stroke [Jerrard-Dunne et al. 2003; Jood et al. 2005; Kirshner 2009; Polychronopoulos et al. 2002; Schulz et al. 2004]. Due

to this complexity, genetic studies on common stroke have been less successful than those on monogenic forms of stroke [Dichgans 2007]. Nevertheless, and because of wider distribution of complex stroke, the impact of these genetic findings on a population may be larger and contribute to new preventative measures and/or therapeutic treatments [Dichgans 2007].

1.2.3 Endophenotypes

The pathophysiology and clinical presentation of stroke is diverse. Since this disease is the end result of different processes and each of these processes may be under the influence of different genetic factors, this complexity can make difficult to detect associations between stroke and specific genes. Therefore, one of the strategies to reduce the complexity of the phenotype is to study intermediate phenotypes or endophenotypes. Intermediate phenotypes are associated with stroke and represent particular components of the disease process, but are not compulsory stages towards stroke. Affected individuals may display those traits before the stroke event and the number of individuals with intermediate phenotypes may exceed the number of available stroke patients. In comparison to stroke, fewer genes are likely to be involved, which may increase the chance of their identification. Genes influencing any of these endophenotypes should be separately assessed for their effect on stroke [Dichgans 2007; Markus 2004].

Carotid intima-media thickness (IMT), for example, is an indicative of subclinical atherosclerosis and is a strong predictor of large artery stroke [Humphries and Morgan 2004]. It is associated with a specific process (atherosclerosis), which increases homogeneity and power. Heritability of carotid IMT is estimated in 30-40% [Humphries and Morgan 2004]. White matter hyperintensities (WMH) on MR brain scans, also known as leukoaraiosis, are associated with small artery ischemic stroke. Leukoaraiosis is also highly heritable (55-80%) [Paternoster et al. 2009]. Other intermediate phenotypes associated with stroke include presence of microbleeds or multiple lacunae on brain imaging.

1.3 Strategies for genetic analyses in stroke

1.3.1 Linkage studies

Usually, the first strategy to dissect the genetics of a trait is linkage analysis, which allows the identification of broad genomic regions containing a disease gene in the absence of a previous biological hypothesis [Dawn Teare and Barrett 2005]. This strategy has been applied to both rare Mendelian forms of stroke and common stroke. Two *loci* are linked if, under independent inheritance, they are transmitted together from parent to offspring more often than would be expected [Dawn Teare and Barrett 2005; Mayeux 2005]. For a disease following a known Mendelian pattern of inheritance, one or few large family pedigrees with several affected individuals can be collected to investigate the genetic basis of a disease [Cui et al. 2010]. Parametric linkage methods are then used to try to identify the set of marker alleles that are segregating with the disease in families, which are assumed to be located near the disease gene [Mayeux 2005]. The linkage strategy is the most powerful to identify rare, high-risk alleles in Mendelian diseases [Mayeux 2005]. For multifactorial diseases, many genes and environmental factors, as well as complex interactions between them, can contribute to risk and the disease mode of inheritance is usually unknown. In this case, non-parametric or model-free methods are used, which rely on the principle that affected relatives in a pedigree would share more often haplotypes that are identical by descent in the region of a disease-causing gene than would be expected by chance [Cui et al. 2010; Dawn Teare and Barrett 2005]. Two commonly applied designs are the affected sib-pair design and the affected relative design [Cui et al. 2010]. In contrast with Mendelian diseases, linkage analysis of complex diseases can only identify large genomic regions, which often contain hundreds of genes, including many biological plausible candidates [Dawn Teare and Barrett 2005]. Thus, usually nonparametric methods have lower resolution to fine map a susceptibility *locus* [Belmont and Leal 2005]. One way to narrow the intervals of interest is to study the loss of heterozigosity in that region or to further fine map regions with strong linkage signals using

approaches that are based on linkage disequilibrium (LD) [Cui et al. 2010; Dawn Teare and Barrett 2005].

1.3.1.1 Single gene disorders

As already mentioned, single-gene defects are a rare but important cause of stroke, especially in younger patients. In some Mendelian conditions, stroke is the prevailing manifestation, while in others it is one of the associated clinical features. CADASIL, Fabry's disease, sickle-cell disease and MELAS are disorders with autosomal dominant, X-linked, autosomal recessive and maternal modes of inheritance, respectively [reviewed in Dichgans 2007; Guo et al. 2010; Hassan and Markus 2000]. The HCHWA-I and HCHWA-D are both autosomal dominant disorders.

CADASIL usually occurs in middle-aged individuals, which do not have vascular risk factors, and its clinical phenotype include recurrent strokes and TIAs and progressive dementia. Brain imaging shows similarities with sporadic small-vessel disease. This autosomal dominant disorder is caused by mutations in the *NOTCH3* gene, although the latter could not be detected for some persons with the typical phenotype [Joutel et al. 1996]. This gene encodes a cell-surface receptor that is expressed on vascular smooth-muscle cells and has a role in arterial development. The majority of patients have *NOTCH3* missense mutations, which cluster in exons 3-6 and affect the number of cysteine residues of the protein [Peters et al. 2005].

Fabry's disease is an X-linked recessive disorder. Deficiency of the lysosomal enzyme α-galactosidase leads to progressive accumulation of glycosphingolipids in the myocardium, vasculature, etc… resulting in complications such as stroke and myocardial infarction. Cerebrovascular symptoms can occur due to large-vessel or small vessel disease, or due to embolism from associated cardiac disease. Fabry's disease appears in childhood or adolescence and complications are observed in mid-adulthood. Most patients have missense or nonsense mutations in the α-galactosidase gene [Dichgans 2007].

Stroke is an important complication of sickle-cell disease, which is the most common cause of stroke in children [Switzer et al. 2006]. This disease can be caused by homozygous mutations in the β chain of the haemoglobin, which result in haemoglobin S (HbS), or by compound heterozygous states of HbS with other haemoglobinopathies

(e.g. α-thalassaemia) [Dichgans 2007]. Stroke is attributed to large-vessel or small-vessel disease and stroke recurrence is common. An abnormal interaction between sickled red blood cells and the vascular endothelium favours thrombus formation and vascular occlusion. The risk of stroke can be affected by modifier genes [Steinberg and Adewoye 2006].

MELAS is characterized by seizures, stroke-like episodes, lactic acidosis, amongst others, and is caused by mutations in mitochondrial DNA [Martínez-Fernández et al. 2001]. The phenotypic heterogeneity observed for mitochondrial disorders may be due to variable expression of mutated DNA in different tissues.

HCHWA-I and HCHWA-D are both autosomal dominant forms of amyloidosis that are characterized by recurrent strokes. Nearly 1/5 of strokes in Icelandic patients aged <35 years are due to HCHWA-I, which is caused by a non-synonym mutation in the gene encoding for cystatin C [Revesz et al. 2002]. In contrast with the Icelandic type, the Dutch type is characterized by dementia and is caused by a non-synonymous mutation in the amyloid β precursor protein (*Aβ-PP*) gene [Levy et al. 1990].

1.3.1.2 Common stroke

Gretardstottir and colleagues [2002] performed the first genome-wide linkage scan in common stroke. 476 Icelandic patients with TIA, ischemic or hemorrhagic stroke, and 438 of their relatives (179 families in total) were analyzed using a framework set of 1,000 microsatellite markers, and a major *locus*, mapped on chromosome 5 (5q12), was identified. After the linkage scan, a case-control association analysis of 864 patients with TIA, ischemic or hemorrhagic stroke, and 908 controls was performed to fine map this *locus*, with the strongest association being found between the phosphodiesterase 4D gene (*PDE4D*) and ischemic stroke [Gretarsdottir et al. 2003]. In addition, they observed that the *PDE4D* mRNA was significantly lower in affected individuals than in controls. The *PDE4D* gene encodes different isoforms of the enzyme PDE4D, which regulates the levels of the second messenger cAMP (cyclic adenosine monophosphate) [Houslay and Adams 2003]. Evidence suggests that cAMP plays a role in the proliferation of smooth muscle cells and macrophages, and possible in atherosclerosis and plaque stability [Matarin et al. 2010]. Therefore, the authors proposed that the *PDE4D* gene might be involved in the pathogenesis of stroke through

atherosclerosis. However, the association of this gene with ischemic stroke was only confirmed in some replication studies [Bersano et al. 2008] and a meta-analysis showed no association after exclusion of the original study, which suggests that the association observed in Iceland may be restricted to specific populations [Bevan et al. 2008].

Two genome-wide linkage scans searching for susceptibility genes for myocardial infarction and stroke were performed by the same group [Helgadottir et al. 2004]. 296 Icelandic families, including 713 individuals with myocardial infarction, and 164 Icelandic families that had been collected for a previous study [Gretarsdottir et al. 2002], including 342 patients with ischemic stroke or TIA, were independently analyzed. The authors observed linkage to the same *locus* on chromosome 13 (13q12-13). Subsequent analyses using 779 individuals with myocardial infarction, 702 stroke patients and 624 unrelated controls lead to the identification of a susceptibility gene for both diseases: arachidonate 5-lipoxygenase-activating protein (*ALOX5AP*). ALOX5AP converts unesterified arachidonic acid to leukotrienes [Dixon et al. 1990] and leukotrienes have been implicated in critical stages of atherosclerosis [Spanbroek et al. 2003]. Like *PDE4D*, replication studies have shown conflicting results [Bersano et al. 2008]. A meta-analysis failed to confirm the association of *ALOX5AP* with ischemic stroke [Zintzaras et al. 2009], but a different meta-analysis indicated that a SNP in this gene is associated with ischemic stroke [Domingues-Montanari et al. 2010]. These contradictory findings may be a result of different inclusion criteria used in each study. The former study analyzed more individuals, but also identified significant heterogeneity between studies, which was not observed by Domingues-Montanari and co-workers [2010].

A whole-genome linkage scan of common stroke was also performed in 56 families from northern Sweden, which had been selected from a relatively genetically homogenous region, but the study did not identify any major *locus* for ischemic stroke [Nilsson-Ardnor et al. 2007].

1.3.2 Association studies

Linkage analysis has successfully identified the genetic causes of many single-gene disorders. In polygenic stroke, however, the situation is more difficult due to several reasons: stroke is a late-onset disease, which reduces the possibility of making

genetic comparisons between living relatives, it has phenotypic and genetic heterogeneity, variable penetrance and it coexists with confounding risk factors [Hassan and Markus 2000]. The majority of genetic studies in stroke have, thus, employed association analysis.

Association analysis is considered one of the best methods to identify genetic factors contributing to complex traits [Mayeux 2005; Palmer and Cardon 2005]. Genetic association studies investigate if a specific marker allele is associated with a disease in a population. In a case-control study, for instance, this is performed by examining if a marker allele is more often present in affected individuals than in healthy controls [Cordell and Clayton 2005; Healy 2006]. In contrast to linkage, where different alleles may be associated with the trait in different families, association assumes that the same allele (or alleles) is associated with the trait in an identical manner across the whole population [Cordell and Clayton 2005].

Association studies can target genetic markers that are assumed to be the disease-causing variants and test these markers for association with a disease (direct association mapping), but the probability of selecting the true causative variant is low [Cordell and Clayton 2005; Orr and Chanock 2008; Palmer and Cardon 2005]. Usually, association studies analyze subsets of single nucleotide polymorphisms (SNPs) in a specific region or throughout the genome. In this case, an association of a marker allele with a disease is likely to be due to LD between that allele and a nearby causal variant (indirect association mapping) [Cordell and Clayton 2005; Orr and Chanock 2008; Palmer and Cardon 2005]. LD is the "non-random association of alleles at different *loci*", which is weakened every time recombination occurs between these *loci* in the population [Dawn Teare and Barrett 2005]. Thus, although association analysis has more power to detect low effect sizes than linkage analysis, it requires that large numbers of genetic markers be genotyped to cover a genomic region, since LD is usually observed for short chromosomal segments [Cordell and Clayton 2005; Mayeux 2005]. Due to the fact that association operates over short distances, this strategy has been used to fine map regions initially detected by linkage analysis [Cordell and Clayton 2005].

Evidence suggests that most of the human genome consists of regions with little evidence of historical recombination (with high levels of LD between markers), where only a few haplotypes are observed, intercalated by hotspots of recombination (with low levels of LD) [Gabriel et al. 2002; Healy 2006]. Therefore, it is possible to genotype

only a subset of SNPs in a region, called haplotype tagging SNPs, to cover the majority of common genetic variation in a region and to indirectly track disease-causing variants [Consortium 2003; Healy 2006]. The International HapMap project (http://www.hapmap.org) [Consortium 2003] and, more recently, the 1000 Genomes Project (http://www.1000genomes.org) [Durbin et al. 2010], have tried to characterize the LD patterns in multiple populations, facilitating the discovery of variants that affect common diseases. Data from these projects can be used in indirect association testing of genetic markers located in candidate genes, in genomic regions identified in family-based linkage analysis or throughout the genome [Consortium 2003].

There are several study designs for association studies [reviewed in Cordell and Clayton 2005], but the population-based case-control design has become the approach of choice in association studies of many complex diseases [Palmer and Cardon 2005] and it was also used in this thesis. The problem associated with this study design is, however, that spurious associations may arise due to stratification and admixture within the population. This commonly occurs when the sample under study contains subgroups of different racial ethnicities and with different allele frequencies [Healy 2006]. Testing SNPs for Hardy-Weinberg equilibrium (HWE) is recommended, as departure from HWE may indicate population stratification or genotyping errors, for example [Fardo et al. 2009]. Careful selection of controls, genotyping of random genetic markers to assess population structure and correct for it, and genomic control may reduce confounding due to population stratification [Cordell and Clayton 2005; Palmer and Cardon 2005]. Additionally, family-based designs, in which parents or other unaffected family members are used as controls, is robust against population substructure and can be used to solve this problem [Cui et al. 2010]. Yet, family-based designs require that more individuals be genotyped and, particularly in late-onset diseases, parents and unaffected family members may be unavailable [Belmont and Leal 2005].

1.3.2.1 Candidate genes

The majority of genetic studies on human stroke have assessed the role of specific candidate genes in stroke susceptibility. The candidate-gene approach requires an *a priori* identification of genes that may be related to the pathogenesis of stroke, the development of intermediate phenotypes (e.g. intima-media thickening) or stroke risk

factors (e.g. hypertension and hyperlipidemia). Usually, these candidate genes are chosen based on their function, but may also be selected based on their genomic location [Domingues-Montanari et al. 2010]. There is an extensive list of tested candidate genes, which are involved in lipid metabolism, inflammation, homocystein metabolism, renin-angiotensin-aldosterone system, coagulation system, fibrinolitic system, platelet receptors, etc [reviewed in Bersano et al. 2008; Dichgans 2007; Domingues-Montanari et al. 2008; Gulcher et al. 2005; Guo et al. 2010]. This list includes:

- lipid metabolism: apolipoprotein E (*APOE*), paraoxonase 1 (*PON1*), lipoprotein lipase (*LPL*) and low-density lipoprotein receptor (*LDL-R*) genes,
- inflammation: C-reactive protein (*CRP*), tumour necrosis factor (*TNF*), interleukin 1 beta (*IL1B*) and interleukin 6 (*IL6*) genes,
- homocystein metabolism: methylenetetrahydrofolate reductase (*MTHFR*) and cystathionine-β-synthase (*CBS*) genes,
- renin-angiotensin-aldosterone system: angiotensin-converting enzyme (*ACE*) and angiotensinogen (*AGT*) genes,
- coagulation system: genes encoding the factor V Leiden (*F5*), prothrombin (*F2*), fibrinogen (*FGA*/*FGB*), factor VII (*F7*), factor XIII (*F13A1*), factor XII (*F12*), Von Willebrand factor (*VWF*),
- fibrinolitic system: genes encoding the plasminogen activator inhibitor 1 (*SERPINE1*),
- nitric oxide metabolism: nitric oxide synthase 3 (*NOS3*) gene,
- platelet receptors: integrin beta 3 (*ITGB3*), integrin alpha 2b (*ITGA2B*), integrin alpha 2 (*ITGA2*) genes.

Despite this extensive list and a large number of studies, the identification of variants for stroke susceptibility has been difficult, and the clinical utility and validity of many candidate genes have yet to be clarified. Most of the significant associations have small effect sizes (relative risk <1.5) and few associations have been consistently replicated across studies [Bersano et al. 2008; Domingues-Montanari et al. 2008]. Potential reasons for this lack of reproducibility include false-positive associations in the original study that were correctly non-replicated in subsequent studies, limited sample size of replication studies, methodological differences between studies (e.g. in study design or phenotype definition), and/or differences in genetic or environmental

background [Colhoun et al. 2003; Domingues-Montanari et al. 2008; Guo et al. 2010; Palmer and Cardon 2005]. The hypothesis that susceptibility *loci* for common stroke have small effects implies that thousands of individuals have to be tested to increase the chance of detecting an effect [Munafò and Flint 2004]. This number becomes very high when stroke subtypes are analyzed separately. Meta-analysis, a statistical tool that is used for combining results from different studies, increases the ability to detect small effects, determining whether a real effect is present or not [Anderson et al. 2010; Casas et al. 2004; Domingues-Montanari et al. 2010; Munafò and Flint 2004; Pereira et al. 2007; Sudlow et al. 2006; Wang et al. 2009]. This has the potential to solve some discrepancies in genetic association studies, and it can also detect between-study heterogeneity and publication bias [Munafò and Flint 2004]. Nevertheless, and although meta-analysis can confirm the involvement of a certain genetic variant when heterogeneity is taken into account, it does not substitute adequately powered studies [Munafò and Flint 2004]. In stroke, results from these analyses need to be interpreted with caution, because stroke risk factors are sometimes not included as covariates and sample sizes remain small when differences between studies (e.g. in ethnicity) are correctly taken into account [Matarin et al. 2010].

Some of the most studied candidate genes for stroke encode the factor V Leiden, prothrombin, MTHFR and ACE. Their possible relation with stroke is briefly described below.

The factor V Leiden, encoded by the *F5* gene (mapped on chromosomal region 1q23), is a glycoprotein that is involved in the coagulation process and is regulated by activated protein C, a protein which limits clot formation [Kalafatis et al. 1994]. The c.1691G>A variant leads to an amino-acid exchange (Arg506Gln), determining resistance to activated protein C [Bertina et al. 1994]. The consequence of this is increased thrombin generation and a hypercoagulable state due to less efficient degradation of mutated factor V by activated protein C when compared with normal factor V [Dahlback 1995], which may explain why carriers of this mutation possibly have increased risk of stroke [Bersano et al. 2008; Casas et al. 2004]. The prothrombin, encoded by the *F2* gene (mapped on chromosomal region 11p11-q12), is a glycoprotein that converts fibrinogen into fibrin. A variant in the 3'-untranslated region of the *F2* gene (c.20210G>A) is associated with increased prothrombin levels and thrombin formation, which may similarly lead to a procoagulant state [Franco et al. 1999]. This is

a plausible reason for the associations with stroke risk [Bersano et al. 2008; Casas et al. 2004].

The MTHFR enzyme catalyzes a reaction that is required for conversion of homocysteine in methionine. The c.677C>T in the *MTHFR* gene (mapped on chromosomal region 1p36.3) leads to an amino-acid substitution (Ala222Val) and is the most common variant associated with moderate hyperhomocysteinemia. This mutation makes the enzyme thermolabile, reducing the metabolism of homocysteine [Frosst et al. 1995]. Homocysteine serum concentration is associated with the risk of stroke in a dose-dependent manner [Wald et al. 2002] and association studies suggest that this *MTHFR* variant increases the risk of stroke [Bersano et al. 2008; Casas et al. 2004].

The ACE enzyme has an important role in blood pressure regulation and electrolyte balance. This enzyme converts angiotensin I to angiotensin II, which is involved in atherosclerotic processes and vasoconstriction, and is responsible for degradation of bradykinin, which in turn may stimulate vasodilator nitric oxide production [Kim and Iwao 2000]. An insertion/deletion (I/D) polymorphism in intron 16 of the *ACE* gene (mapped on chromosomal region 17q23) partly determines plasma and intracellular levels of ACE in healthy individuals and in patients with stroke [Sharma et al. 1994; Tiret et al. 1992]. Individuals with the DD genotype have an increase in ACE activity compared with homozygous for the I allele [Agerholm-Larsen et al. 2000] and may also have an increased risk of stroke [Bersano et al. 2008; Casas et al. 2004].

1.3.2.2 Genome-wide association studies

Often, there is incomplete knowledge of the biological pathways underlying complex traits, which limits the selection of candidate genes for association studies. Researchers may also want to test genomic regions for association with a given trait with no *a priori* reason. Data from the HapMap project [Consortium 2003] and the 1000 Genomes Project [Durbin et al. 2010] have enabled the development of array-based platforms that can be used in high-throughput genotyping of hundreds of thousands of common SNPs across the entire genome. It is thus possible to perform an association analysis in an unbiased way, which increases the chance of finding novel insights into disease pathophysiology. These genome-wide association studies (GWAS) have

successfully identified genetic variants involved in some complex human traits [Hindorff et al. 2009], but can also be problematic. The number of statistical tests performed in these studies is very large, which increases the potential for false positive results, and very stringent significance levels and replication of findings are required [Pearson and Manolio 2008]. GWAS have analyzed hundreds to thousands of people, but the detection of small overall effects, in many cases, requires larger sample sizes [Manolio et al. 2009]. One problem of increasing sample sizes by using a broad definition of the phenotype or by collecting samples from different ethnic populations is that the study sample may become less homogenous, which can have a negative impact on power. Several GWAS have been carried out in stroke [Lanktree et al. 2010].

The first GWAS in ischemic stroke was published in 2007 [Matarín et al. 2007]. Over 400,000 SNPs were assessed in 249 patients and 268 neurological controls, but none of them was significant after adjusting for multiple testing. This study was underpowered to detect *loci* with a moderate effect on stroke risk. The data obtained in this GWAS was also used to assess the role of copy number variants (CNVs) in ischemic stroke risk [Matarin et al. 2008]. No common genomic structural variation was unequivocally associated with ischemic stroke. Only one of the 45 CNVs that had not previously been reported in healthy individuals or that did not overlap with previously identified CNVs was present in more than one patient. However, this CNV was also detected in 5 of an additional 460 controls, which suggests that it is not a risk factor for ischemic stroke. The authors argued that other structural variants (smaller CNVs or CNVs located in genomic regions poorly covered by the genotyping arrays) may confer risk for stroke.

Kubo and colleagues [2007] identified a non-synonymous SNP in the protein kinase C eta (*PRKCH*) gene in association with lacunar infarction in two independent Japanese samples. First, 52,608 SNPs were genotyped in 188 patients with cerebral infarction and 188 controls. In a second phase, the 1,098 SNPs with the smallest *P*-values were genotyped in the remaining sample (924 individuals with cerebral infarction and the same number of controls). A non-synonymous SNP in *PRKCH*, possibly affecting protein kinase C η (PKCη) activity, was found significantly associated with lacunar infarction and this association was subsequently replicated in an independent sample of 1,137 cases with lacunar infarction and 1,875 controls. The associated SNP is likely to be specific to Asian populations as its minor allele frequency is significantly reduced in other populations [Kubo et al. 2007]. PKCη is expressed in

cells located in human atherosclerotic lesions and its expression increases with lesion progression [Kubo et al. 2007]. These results support a role for *PRKCH* in ischemic stroke (lacunar) pathogenesis, possibly through atherosclerosis. This group also identified an association with brain infarction for a SNP located in the 5'-flanking region of the apelin receptor gene (*APLNR*), which was likely to regulate the expression of this gene [Hata et al. 2007]. Apelin is the endogenous ligand of the apelin receptor protein (APJ) and has some functions in the control of blood pressure [Kagiyama et al. 2005; Seyedabadi et al. 2002]. A different GWAS was also carried out in Japanese individuals (131 ischemic stroke patients and 135 controls) by an independent group. Approximately 520,000 SNPs were assayed, followed by genotyping of 100 SNPs in 705 ischemic stroke patients and 3426 controls [Yamada et al. 2009]. Two non-synonymous SNPs in the cadherin, EGF LAG seven-pass G-type receptor 1 (*CELSR1*) gene were associated with ischemic stroke. A SNP near *CELSR2* (a *CELSR1*'s homologue) had previously been associated with the serum concentration of low density lipoprotein cholesterol and the prevalence of myocardial infarction [Kathiresan et al. 2008; Kathiresan et al. 2009], but the *CELSR1* SNPs were not related to lipid profiles [Yamada et al. 2009].

The Icelandic group that had previously performed the first whole-genome linkage scan also carried out a GWAS in ischemic stroke [Gretarsdottir et al. 2008]. Over 300,000 SNPs were genotyped in 1,661 patients and 10,815 controls from Iceland, followed by replication of the most significant associations in two European samples (2,224 cases and 2,583 controls). Two SNPs on chromosomal region 4q25, which had previously been reported to associate with atrial fibrillation [Gudbjartsson et al. 2007], were further tested in additional European samples (2,327 patients and 16,760 controls). One SNP was significantly associated with ischemic stroke and the evidence for association increased when the analysis was restricted to the cardioembolic stroke subtype, which is a major complication of atrial fibrillation. The other SNP was also associated with cardioembolic stroke. These two genetic variants are close to the paired-like homeodomain 2 gene (*PITX2*), which encodes a transcriptional activator that is important for development of the sinoatrial node in mice, the natural pacemaker of the heart [Faucourt et al. 2001; Mommersteeg et al. 2007]. In a different study from the same group [Gudbjartsson et al. 2007], a variant on chromosomal region 16q22 was associated with both atrial fibrillation and ischemic stroke. This SNP is located in the zinc finger homeobox 3 (*ZFHX3*) gene, which encodes a transcription factor, and this

gene has been associated with neuronal and skeletal muscle differentiation [Berry et al. 2001].

Ikram and colleagues [2009] performed a GWAS in a cohort of 19,602 white participants. Over an average follow-up of eleven years, 1,544 persons had a stroke (1,164 ischemic strokes in total). Two intergenic SNPs in significant LD with each other and located in close proximity to the ninjurin 2 gene (*NINJ2*) were significantly associated with stroke and the evidence for association increased when the analysis was restricted to ischemic patients. One SNP was subsequently associated with stroke in a cohort of 2,430 black participants (comprising 215 stroke patients) and in a case-control sample of 4,265 white people (652 stroke patients). The second SNP was only associated in the sample of white participants possibly because of low LD between the two SNPs in the second sample. *NINJ2* encodes the ninjurin2 protein, an adhesion molecule that is expressed in glia [Araki and Milbrandt 2000]. It was proposed that the level of expression of this protein affects how the brain tolerates ischemic insults [Ikram et al. 2009].

It is possible to draw several conclusions from the results obtained in GWAS of stroke. First, the results have been inconsistent, as the same *locus* was not identified in two independent studies at a genome-wide significance level (approximately $P<10^{-7}$), but there is also significant heterogeneity in study designs and in ethnic backgrounds of individuals. Second, the effect sizes identified are modest. Third, no association has been found between stroke and any *locus* with previous evidence from candidate gene studies.

1.3.3 Gene-gene interactions

Complex diseases likely result from the effect of several genes, and possibly on interactions with other genes [Ritchie 2011]. The contribution of one *locus* to a phenotype may therefore depend on the genotype at a second *locus*, or in the extreme, the genetic background of an individual [Moore and Williams 2005; Tyler et al. 2009]. Identifying these non-additive or epistatic interactions in genetic studies is important, as any positive finding could shed light on the biological and biochemical pathways contributing to disease [Cordell 2009].

Several methods have been used to detect epistasis [reviewed in Cordell 2009; Ritchie 2011]. One possibility is to test for interactions that occur between two or more known or hypothetical genetic risk factors (e.g. to replicate a previous finding) using regression models. It is more common, however, to search for *loci* that may interact in genotype data from many genetic variants, obtained in candidate gene or GWAS. In this situation, there is probably no previous evidence linking these genetic variants with a specific disease, so one may wish to perform an exhaustive search of 2-*loci* or higher-order interactions, which has the important disadvantage of increasing the potential for false positives. Furthermore, in a genome scan, while searching for 2-*loci* interactions is computationally intensive but feasible, searching for 3-*loci* to higher-level interactions is impractical [Cordell 2009; Marchini et al. 2005]. To overcome this problem, two-stage procedures have been proposed, in which a subset of *loci* are selected in a first stage based on single-*locus* significance thresholds, and an exhaustive search of two-*locus* interactions is performed in the filtered dataset. Alternatively, genetic markers can be selected in a first stage based on experimental knowledge of biological networks or protein-protein interactions [Emily et al. 2009; Ritchie 2011]. Data-mining methods are another alternative to investigate the existence of gene-gene interactions. The multifactor dimensionality reduction (MDR) method is one of the most commonly used data-mining approaches [Ritchie et al. 2001]. It is a model free and nonparametric method, and can detect genetic interactions in the absence of independent main effects [Ritchie et al. 2003; Ritchie et al. 2001]. This method reduces the dimensionality of the data from N dimensions to one dimension by pooling multilocus genotypes into high and low risk groups, and subsequently tests the new, one dimensional variable for its ability to classify and predict disease status. If 10 cross-validation intervals are chosen, the MDR divides the data into 10 parts, develops the interaction model in 9 of these parts, and assesses the model in the remaining 1/10 of the data. The multiple testing issue is addressed by combining this cross-validation strategy and permutation testing [Ritchie et al. 2001]. An important disadvantage of this method is that, for higher-order interactions, it is best suited for the analysis of relatively small number of *loci* [Cordell 2009].

Susceptibility to ischemic stroke may be modulated by non-linear gene-gene interactions [Liu et al. 2009; Shen et al. 2007]. Liu and colleagues (2009), in particular, investigated the existence of gene-gene interactions between five candidate genes and stroke and found that individuals with a combination of polymorphisms in three of these

genes had an increased risk of thrombotic stroke. After gene-gene interactions have been detected, it is desirable to elucidate how these interactions at the DNA level can influence phenotypes in an individual through biochemical processes that are dependent on biomolecular interactions [Cordell 2009; Moore 2003; Moore and Williams 2005]. It is challenging to make inferences about biological processes based on statistical models of interactions, especially in humans, which cannot be used in experimental studies [Moore and Williams 2005]. Nevertheless, interpretation of results can be greatly enhanced by testing for interactions between genes encoding proteins in the same pathway or with similar functions [Ritchie 2011], as was performed in this thesis. Although this will bias the analysis and possibly miss novel interactions between SNPs, it will also increase the likelihood that any identified interaction will be biologically plausible and interpretable [Ritchie 2011].

Epistasis is a plausible explanation for the lack of success of many association studies, which have mainly used a single-*locus* analysis strategy [Cordell 2009]. If a genetic variant is individually tested for association with a phenotype but functions through complex mechanisms involving other genes, its effects may be missed [Cordell 2009]. The existence of these interactions may also explain why positive results from linkage and association studies of complex diseases are frequently not replicated in subsequent studies or show inconsistencies [Greene et al. 2009; Moore 2003]. For example, Greene and co-workers [2009] showed that, under an epistatic model, in which SNP_A and SNP_B are interacting *loci*, the power to replicate a main effect at SNP_A can be greatly affected by a small change in minor allele frequency at SNP_B. Furthermore, an initially found protective allele at *locus* A may be replicated as a risk allele due to differences in allele frequencies at *locus* B [Greene et al. 2009].

1.4 Stroke outcome and functional recovery

As already referred, a significant proportion of stroke survivors are left disabled. These patients have an increased probability of having a second stroke episode and/or of dying in the following years [Hankey et al. 2002; Hardie et al. 2004]. Patients with large-vessel ischemic stroke lose 120 million neurons per hour, which may explain the motor weakness and sensory disturbances that most individuals exhibit [Lakhan et al. 2009]. Comparing to the normal rate of neuronal loss during aging, this means that the

brain ages 3.6 years per hour in the absence of any treatment [Lakhan et al. 2009]. Despite this loss, a slow but consistent recovery occurs in the brain afterwards due to several physiological processes [Hurtado et al. 2006]. Edema resolution and/or reperfusion of the ischemic penumbra are plausible explanations for the spontaneous recovery that is observed in the first days after ischemic stroke [Hurtado et al. 2006]. Formation of new synapses from the surviving neurons and redundancy of brain circuits, allowing functional compensation of damaged areas, are possible mechanisms of brain plasticity, which also influence patient's recovery [Hurtado et al. 2006; Lakhan et al. 2009]. Moreover, formation of new neurons (neurogenesis) and blood vessels (angiogenesis) contribute to ameliorate neurological deficits and is correlated with patient's survival time, respectively [Jin et al. 2006; Krupinski et al. 1993; Nakatomi et al. 2002; Slevin et al. 2006]. Although the recovery mechanisms after hemorrhagic stroke have been less studied, it is already known that neurogenesis is induced after this type of stroke [Shendure and Ji 2008].

Table 1. Modified Rankin Scale (mRS)

mRS	Symptoms
0	No symptoms
1	No significant disability, despite symptoms (able to perform all usual activities and duties)
2	Slight disability (unable to perform all previous activities but able to look after own affairs without assistance)
3	Moderate disability (require some help, but able to walk without assistance)
4	Moderately severe disability (unable to walk without assistance and unable to attend to own bodily needs without assistance)
5	Severe disability (bedridden, incontinent and require constant nursing care and attention)
6	Dead

Adapted from [Kasner 2006]

The modified Rankin Scale (mRS) is one of the most widely used scales to assess stroke outcome [van Swieten et al. 1988]. It measures the global disability of a patient and has been employed in clinical practice to evaluate the effect of stroke on patient's activities and life-style [Kasner 2006]. mRS has been used as a primary end point in randomized clinical trials and can also be used to guide rehabilitation procedures [Kasner 2006]. This scale has seven different grades, with mRS=0 indicating no symptoms and mRS=6 indicating death (Table 1). In statistical analyses, the mRS outcomes are often dichotomized as either good or poor, but the cut-off for good/poor outcome is subject of controversy [Weisscher et al. 2008]. Other scales that have been employed to assess stroke outcome include the Barthel Index (BI), the Glasgow Outcome Scale (GOS) and the Stroke Impact Scale (SIS). The National Institutes of Health stroke scale (NIHSS) has been used as an initial assessment tool of stroke severity, and is predictive of long-term outcome [Kasner 2006]. This scale assesses level of consciousness, language (aphasia), coordination (ataxia), hemi-inattention (neglect), among other features. It is correlated with infarct volume, although other factors, including age and stroke location, contribute to this correlation [Kasner 2006].

Several factors are known to affect stroke outcome. Age, gender, diabetes, blood pressure, initial stroke severity, previous stroke, stroke type, volume of the infarction or hemorrhage, among others, have been identified as significant predictors of functional outcome and/or patient's mortality [Di Carlo et al. 2003; Megherbi et al. 2003; Tilling et al. 2001; Weimar et al. 2006; Weimar et al. 2002; Wong et al. 2005]. Moreover, it was shown that a family history of stroke is associated with stroke outcome, but not with stroke severity or mortality at 90 days, indicating that genetic factors may also influence stroke outcome [Jood et al. 2005; Lisabeth et al. 2005]. Studies with animal models, which recreate human ischemic stroke and ICH, also support this hypothesis [Atochin et al. 2010; Chang et al. 2011; Grossetete and Rosenberg 2008; Hyakkoku et al. 2010; Jeffs et al. 1997; Leker et al. 2007; Schäbitz et al. 2007; Tsuji et al. 2005]. Ischemic stroke can be induced in animals through occlusion of the middle cerebral artery (MCAO) using a clip or by injecting small blood clots [Small and Buchan 2000]. ICH, on the other hand, can be recreated through infusion of bacterial collagenase, which disrupts the basal lamina of blood vessels causing blood to flood the surrounding tissue [MacLellan et al. 2010]. Jeffs and colleagues [1997] performed a genome scan in rats to identify the genetic factors responsible for large infarct volumes after MCAO. They identified a quantitative trait *locus* that accounted for approximately 70% of total

variance in infarct volume [Jeffs et al. 1997]. Other studies have used KO mice to assess the influence of specific genes in stroke outcome. For instance, Caveolin-1 (CAV1) may have a deleterious role after ICH, as *Cav-1* KO mice had smaller injury volumes and milder neurological deficits than wild-type mice after the insult [Chang et al. 2011]. In a different study, mice deficient in the alpha 1 subunit of soluble guanylate cyclase (sGCalpha1(-/-)) had larger infarct volumes and worse neurological deficits after ischemic stroke than wild-type mice, suggesting that sGCalpha1 is beneficial after stroke [Atochin et al. 2010].

Despite evidence of a role of genetics in stroke outcome and recovery, the investigation of the genetic factors involved in stroke outcome in humans is still in its infancy, when compared with the large number of association studies assessing the role of candidate genes in stroke susceptibility. The *APOE* gene is the most tested candidate gene for stroke outcome [Meschia 2004]. As observed in genetic association studies of stroke susceptibility, some studies have found an association between *APOE* and stroke outcome, while others failed to confirm this association [McCarron et al. 2000; McCarron et al. 1998; Sarzynska-Dlugosz et al. 2007; Treger et al. 2003]. A meta-analysis found no association between presence of ε4+ genotypes and poor outcome after ischemic stroke or ICH, but there was significant heterogeneity between studies regarding the clinical scales used to assess patient's outcome and the time of assessment [Martínez-González and Sudlow 2006]. *APOE* is located on chromosome 19 and has three common alleles – ε2, ε3 and ε4, which encode the three major isoforms of the protein ApoE. This protein is the primary mediator of cholesterol and lipid transport in the brain and is associated with serum lipoprotein level. Furthermore, it has an important role in neurological diseases, such as Alzheimer's disease. Other candidate genes tested for a role in stroke outcome include those encoding for proteins involved in platelet adhesion and aggregation (*PLAT* [plasminogen activator, tissue] and *SERPINE1* genes), in metabolic pathways regulating oxidative stress in the brain (*MTHFR* and *GSTO1* [glutathione S-transferase omega 1] genes), in inflammatory pathways (*PTGS2* gene), and proteins of the thrombolytic system (*GP1BA* [glycoprotein Ib, alpha polypeptide] and *ITGB3* genes) [Maguire et al. 2010; Peddareddygari et al. 2009]. Two *PTGS2* SNPs and one *ITGB3* SNP were nominally associated with stroke outcome [Maguire et al. 2010]. Until now, no GWAS has been carried out to investigate the role of genetic factors on stroke outcome.

1.5 Objectives and thesis outline

The identification and characterization of the genetic factors underlying susceptibility to stroke and influencing patient's outcome afterwards should be a priority matter due to the high incidence, mortality and disability associated with stroke. The main objective of the present thesis was therefore to identify genetic factors involved in stroke susceptibility and in patient's outcome at three months. For this purpose, the following strategies were employed:

- Candidate genes were selected and tested for a role in stroke susceptibility using a population-based case-control design and taking advantage of a Portuguese biobank of 672 stroke patients and 530 healthy individuals.
- Searching for genes that contribute to stroke outcome constitutes one main originality of this thesis. Two approaches were followed to attain this goal: candidate genes and GWAS. A subset of 546 stroke patients was analyzed. Extensive clinical data during hospitalization and information on functional outcome at three months were available for each patient. Candidate genes were selected and tested for association with patient's outcome at three months. A pilot GWAS of approximately 250,000 SNPs was also carried out using a DNA pooling strategy. This made possible to perform an association analysis in an unbiased manner, with no prior selection of candidate genes, which increased the chance of finding novel insights into the mechanisms of stroke recovery, and in a cost-effective way.

The thesis outline is as follows:

- In chapter 2 of this work, genetic variants in two matrix metalloproteinase genes (*MMP2* and *MMP9*) were tested for association with patient's outcome at three months.
- Four inflammatory genes (*IL1B*, *IL6*, *MPO* and *TNF*) were tested for a role in stroke susceptibility and outcome and these analyses are described in chapter 3.
- In chapter 4, the association with stroke outcome was analyzed for genetic variants within three growth factor genes (*BDNF*, *FGF2* and *VEGFA*).
- Variants of two genes involved in the nitric oxide metabolism/oxidative stress (*NOS1* and *NOS3*) were tested for association with stroke susceptibility and outcome in chapter 5.

- In chapter 6, we sought to identify novel genes contributing to patient's outcome at three months, by performing a pilot GWAS in DNA pooled samples. A total of 262,264 SNPs, located throughout the genome, were assessed, followed by individual genotyping to validate results from the pooling stage.
- Chapter 7 comprises a general discussion of the results presented in chapters 2-6, as well as concluding remarks and future perspectives.
- The abstracts of published manuscripts for which I contributed as co-author are presented in Appendix 1.

References

Adams HP, Jr., Bendixen BH, Kappelle LJ, Biller J, Love BB, Gordon DL, Marsh EE, 3rd. (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

Adibhatla RM, Hatcher JF. (2008) Tissue plasminogen activator (tPA) and matrix metalloproteinases in the pathogenesis of stroke: therapeutic strategies. CNS Neurol Disord Drug Targets 7:243-253

Agerholm-Larsen B, Nordestgaard BG, Tybjaerg-Hansen A. (2000) ACE gene polymorphism in cardiovascular disease: meta-analyses of small and large studies in whites. Arterioscler Thromb Vasc Biol 20:484-492

Amarenco P, Bogousslavsky J, Caplan LR, Donnan GA, Hennerici MG. (2009) Classification of stroke subtypes. Cerebrovasc Dis 27:493-501

Anderson CD, Biffi A, Rost NS, Cortellini L, Furie KL, Rosand J. (2010) Chromosome 9p21 in ischemic stroke: population structure and meta-analysis. Stroke 41:1123-1131

Andlin-Sobocki P, Jönsson B, Wittchen HU, Olesen J. (2005) Cost of disorders of the brain in Europe. Eur J Neurol 12 Suppl 1:1-27

Araki T, Milbrandt J. (2000) Ninjurin2, a novel homophilic adhesion molecule, is expressed in mature sensory and enteric neurons and promotes neurite outgrowth. J Neurosci 20:187-195

Baird AE. (2006) Blood biologic markers of stroke: improved management, reduced cost? Curr Atheroscler Rep 8:267-275

Bak S, Gaist D, Sindrup SH, Skytthe A, Christensen K. (2002) Genetic liability in stroke: a long-term follow-up study of Danish twins. Stroke 33:769-774

Belmont JW, Leal SM. (2005) Complex phenotypes and complex genetics: an introduction to genetic studies of complex traits. Curr Atheroscler Rep 7:180-187

Berry FB, Miura Y, Mihara K, Kaspar P, Sakata N, Hashimoto-Tamaoki T, Tamaoki T. (2001) Positive and negative regulation of myogenic differentiation of C2C12 cells by isoforms of the multiple homeodomain zinc finger transcription factor ATBF1. J Biol Chem 276:25057-25065

Bersano A, Ballabio E, Bresolin N, Candelise L. (2008) Genetic polymorphisms for the study of multifactorial stroke. Hum Mutat 29:776-795

Bertina RM, Koeleman BP, Koster T, Rosendaal FR, Dirven RJ, de Ronde H, van der Velden PA, Reitsma PH. (1994) Mutation in blood coagulation factor V associated with resistance to activated protein C. Nature 369:64-67

Bevan S, Dichgans M, Gschwendtner A, KuhlenbÃ¤umer G, Ringelstein EB, Markus HS. (2008) Variation in the PDE4D gene and ischemic stroke risk: a systematic review and meta-analysis on 5200 cases and 6600 controls. Stroke 39:1966-1971

Bhatnagar P, Scarborough P, Smeeton NC, Allender S. (2010) The incidence of all stroke and stroke subtype in the United Kingdom, 1985 to 2008: a systematic review. BMC Public Health 10:539

Broderick JP. (2005) The STICH trial: what does it tell us and where do we go from here? Stroke 36:1619-1620

Casas JP, Hingorani AD, Bautista LE, Sharma P. (2004) Meta-analysis of genetic studies in ischemic stroke: thirty-two genes involving approximately 18,000 cases and 58,000 controls. Arch Neurol 61:1652-1661

Chang CF, Chen SF, Lee TS, Lee HF, Shyue SK. (2011) Caveolin-1 deletion reduces early brain injury after experimental intracerebral hemorrhage. Am J Pathol 178:1749-1761

Colhoun HM, McKeigue PM, Davey Smith G. (2003) Problems of reporting genetic associations with complex outcomes. Lancet 361:865-872

Consortium IH. (2003) The International HapMap Project. Nature 426:789-796

Cordell HJ. (2009) Detecting gene-gene interactions that underlie human diseases. Nat Rev Genet 10:392-404

Cordell HJ, Clayton DG. (2005) Genetic association studies. Lancet 366:1121-1131

Correia M, Silva MR, Matos I, Magalhães R, Lopes JC, Ferro JM, Silva MC. (2004) Prospective community-based study of stroke in Northern Portugal: incidence and case fatality in rural and urban populations. Stroke 35:2048-2053

Cui Y, Li G, Li S, Wu R. (2010) Designs for linkage analysis and association studies of complex diseases. Methods Mol Biol 620:219-242

Dahlback B. (1995) New molecular insights into the genetics of thrombophilia. Resistance to activated protein C caused by Arg506 to Gln mutation in factor V as a pathogenic risk factor for venous thrombosis. Thromb Haemost 74:139-148

Davenport R, Dennis M. (2000) Neurological emergencies: acute stroke. J Neurol Neurosurg Psychiatry 68:277-288

Dawn Teare M, Barrett JH. (2005) Genetic linkage studies. Lancet 366:1036-1044

Deb P, Sharma S, Hassan KM. (2010) Pathophysiologic mechanisms of acute ischemic stroke: An overview with emphasis on therapeutic significance beyond thrombolysis. Pathophysiology 17:197-218

Di Carlo A, Lamassa M, Baldereschi M, Pracucci G, Basile AM, Wolfe CD, Giroud M, Rudd A, Ghetti A, Inzitari D. (2003) Sex differences in the clinical presentation, resource use, and 3-month outcome of acute stroke in Europe: data from a multicenter multinational hospital-based registry. Stroke 34:1114-1119

Dichgans M. (2007) Genetics of ischaemic stroke. Lancet Neurol 6:149-161

Dixon RA, Diehl RE, Opas E, Rands E, Vickers PJ, Evans JF, Gillard JW, Miller DK. (1990) Requirement of a 5-lipoxygenase-activating protein for leukotriene synthesis. Nature 343:282-284

Domingues-Montanari S, Fernández-Cadenas I, del Rio-Espinola A, Corbeto N, Krug T, Manso H, Gouveia L, Sobral J, Mendioroz M, Fernández-Morales J, Alvarez-Sabin J, Ribó M, Rubiera M, Obach V, Martí-Fàbregas J, Freijo M, Serena J, Ferro JM, Vicente AM, Oliveira SA, Montaner J. (2010) Association of a genetic variant in the ALOX5AP with higher risk of ischemic stroke: a case-control, meta-analysis and functional study. Cerebrovasc Dis 29:528-537

Domingues-Montanari S, Mendioroz M, Rio-Espinola Ad, Fernández-Cadenas I, Montaner J. (2008) Genetics of stroke: a review of recent advances. Expert Rev. Mol. Diagn. 8:495-513

Durbin RM, Abecasis GR, Altshuler DL, Auton A, Brooks LD, Gibbs RA, Hurles ME, McVean GA, Consortium GP. (2010) A map of human genome variation from population-scale sequencing. Nature 467:1061-1073

Emily M, Mailund T, Hein J, Schauser L, Schierup MH. (2009) Using biological networks to search for interacting loci in genome-wide association studies. Eur J Hum Genet 17:1231-1240

Fardo DW, Becker KD, Bertram L, Tanzi RE, Lange C. (2009) Recovering unused information in genome-wide association studies: the benefit of analyzing SNPs out of Hardy-Weinberg equilibrium. Eur J Hum Genet 17:1676-1682

Faucourt M, Houliston E, Besnardeau L, Kimelman D, Lepage T. (2001) The pitx2 homeobox protein is required early for endoderm formation and nodal signaling. Dev Biol 229:287-306

Feigin VL, Lawes CM, Bennett DA, Barker-Collo SL, Parag V. (2009) Worldwide stroke incidence and early case fatality reported in 56 population-based studies: a systematic review. Lancet Neurol 8:355-369

Flossmann E, Schulz UG, Rothwell PM. (2004) Systematic review of methods and results of studies of the genetic epidemiology of ischemic stroke. Stroke 35:212-227

Franco RF, Trip MD, ten Cate H, van den Ende A, Prins MH, Kastelein JJ, Reitsma PH. (1999) The 20210 G-->A mutation in the 3'-untranslated region of the prothrombin gene and the risk for arterial thrombotic disease. Br J Haematol 104:50-54

Frosst P, Blom HJ, Milos R, Goyette P, Sheppard CA, Matthews RG, Boers GJ, den Heijer M, Kluijtmans LA, van den Heuvel LP. (1995) A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase. Nat Genet 10:111-113

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D. (2002) The structure of haplotype blocks in the human genome. Science 296:2225-2229

Greene CS, Penrod NM, Williams SM, Moore JH. (2009) Failure to replicate a genetic association may provide important clues about genetic architecture. PLoS One 4:e5639

Gretarsdottir S, Sveinbjornsdottir S, Jonsson HH, Jakobsson F, Einarsdottir E, Agnarsson U, Shkolny D, Einarsson G, Gudjonsdottir HM, Valdimarsson EM, Einarsson OB, Thorgeirsson G, Hadzic R, Jonsdottir S, Reynisdottir ST, Bjarnadottir SM, Gudmundsdottir T, Gudlaugsdottir GJ, Gill R, Lindpaintner K, Sainz J, Hannesson HH, Sigurdsson GT, Frigge ML, Kong A, Gudnason V, Stefansson K, Gulcher JR. (2002) Localization of a susceptibility gene for common forms of stroke to 5q12. Am J Hum Genet 70:593-603

Gretarsdottir S, Thorleifsson G, Manolescu A, Styrkarsdottir U, Helgadottir A, Gschwendtner A, Kostulas K, Kuhlenbäumer G, Bevan S, Jonsdottir T, Bjarnason H, Saemundsdottir J, Palsson S, Arnar DO, Holm H, Thorgeirsson G, Valdimarsson EM, Sveinbjörnsdottir S, Gieger C, Berger K, Wichmann HE, Hillert J, Markus H, Gulcher JR, Ringelstein EB, Kong A, Dichgans M, Gudbjartsson DF, Thorsteinsdottir U, Stefansson K. (2008) Risk variants for atrial fibrillation on chromosome 4q25 associate with ischemic stroke. Ann Neurol 64:402-409

Gretarsdottir S, Thorleifsson G, Reynisdottir ST, Manolescu A, Jonsdottir S, Jonsdottir T, Gudmundsdottir T, Bjarnadottir SM, Einarsson OB, Gudjonsdottir HM, Hawkins M, Gudmundsson G, Gudmundsdottir H, Andrason H, Gudmundsdottir AS, Sigurdardottir M, Chou TT, Nahmias J, Goss S, Sveinbjornsdottir S, Valdimarsson EM, Jakobsson F, Agnarsson U, Gudnason V, Thorgeirsson G, Fingerle J, Gurney M, Gudbjartsson D, Frigge ML, Kong A, Stefansson K, Gulcher JR. (2003) The gene encoding phosphodiesterase 4D confers risk of ischemic stroke. Nat Genet 35:131-138

Grossetete M, Rosenberg GA. (2008) Matrix metalloproteinase inhibition facilitates cell death in intracerebral hemorrhage in mouse. J Cereb Blood Flow Metab 28:752-763

Gudbjartsson DF, Arnar DO, Helgadottir A, Gretarsdottir S, Holm H, Sigurdsson A, Jonasdottir A, Baker A, Thorleifsson G, Kristjansson K, Palsson A, Blondal T, Sulem P, Backman VM, Hardarson GA, Palsdottir E, Helgason A, Sigurjonsdottir R, Sverrisson JT, Kostulas K, Ng MC, Baum L, So WY, Wong KS, Chan JC, Furie KL, Greenberg SM, Sale M, Kelly P, MacRae CA, Smith EE, Rosand J, Hillert J, Ma RC, Ellinor PT, Thorgeirsson G, Gulcher JR, Kong A, Thorsteinsdottir U, Stefansson K. (2007) Variants conferring risk of atrial fibrillation on chromosome 4q25. Nature 448:353-357

Gulcher JR, Gretarsdottir S, Helgadottir A, Stefansson K. (2005) Genes contributing to risk for common forms of stroke. Trends Mol Med 11:217-224

Guo JM, Liu AJ, Su DF. (2010) Genetics of stroke. Acta Pharmacol Sin 31:1055-1064

Hankey GJ. (2006) Potential new risk factors for ischemic stroke: what is their potential? Stroke 37:2181-2188

Hankey GJ, Jamrozik K, Broadhurst RJ, Forbes S, Anderson CS. (2002) Long-term disability after first-ever stroke and related prognostic factors in the Perth Community Stroke Study, 1989-1990. Stroke 33:1034-1040

Hardie K, Hankey GJ, Jamrozik K, Broadhurst RJ, Anderson C. (2004) Ten-year risk of first recurrent stroke and disability after first-ever stroke in the Perth Community Stroke Study. Stroke 35:731-735

Hassan A, Markus HS. (2000) Genetics and ischaemic stroke. Brain 123 (Pt 9):1784-1812

Hata J, Matsuda K, Ninomiya T, Yonemoto K, Matsushita T, Ohnishi Y, Saito S, Kitazono T, Ibayashi S, Iida M, Kiyohara Y, Nakamura Y, Kubo M. (2007) Functional SNP in an Sp1-binding site of AGTRL1 gene is associated with susceptibility to brain infarction. Hum Mol Genet 16:630-639

Healy DG. (2006) Case-control studies in the genomic era: a clinician's guide. Lancet Neurol 5:701-707

Helgadottir A, Manolescu A, Thorleifsson G, Gretarsdottir S, Jonsdottir H, Thorsteinsdottir U, Samani NJ, Gudmundsson G, Grant SF, Thorgeirsson G, Sveinbjornsdottir S, Valdimarsson EM, Matthiasson SE, Johannsson H, Gudmundsdottir O, Gurney ME, Sainz J, Thorhallsdottir M, Andresdottir M, Frigge ML, Topol EJ, Kong A, Gudnason V, Hakonarson H, Gulcher JR, Stefansson K. (2004) The gene encoding 5-lipoxygenase activating protein confers risk of myocardial infarction and stroke. Nat Genet 36:233-239

Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci U S A 106:9362-9367

Houslay MD, Adams DR. (2003) PDE4 cAMP phosphodiesterases: modular enzymes that orchestrate signalling cross-talk, desensitization and compartmentalization. Biochem J 370:1-18

Humphries SE, Morgan L. (2004) Genetic risk factors for stroke and carotid atherosclerosis: insights into pathophysiology from candidate gene approaches. Lancet Neurol 3:227-235

Hurtado O, Pradillo JM, Alonso-Escolano D, Lorenzo P, Sobrino T, Castillo J, Lizasoain I, Moro MA. (2006) Neurorepair versus neuroprotection in stroke. Cerebrovasc Dis 21 Suppl 2:54-63

Hyakkoku K, Hamanaka J, Tsuruma K, Shimazawa M, Tanaka H, Uematsu S, Akira S, Inagaki N, Nagai H, Hara H. (2010) Toll-like receptor 4 (TLR4), but not TLR3 or TLR9, knock-out mice have neuroprotective effects against focal cerebral ischemia. Neuroscience 171:258-267

Ikram MA, Seshadri S, Bis JC, Fornage M, DeStefano AL, Aulchenko YS, Debette S, Lumley T, Folsom AR, van den Herik EG, Bos MJ, Beiser A, Cushman M, Launer LJ, Shahar E, Struchalin M, Du Y, Glazer NL, Rosamond WD, Rivadeneira F, Kelly-Hayes M, Lopez OL, Coresh J, Hofman A, DeCarli C, Heckbert SR, Koudstaal PJ, Yang Q, Smith NL, Kase CS, Rice K, Haritunians T, Roks G, de Kort PL, Taylor KD, de Lau LM, Oostra BA, Uitterlinden AG, Rotter JI, Boerwinkle E, Psaty BM, Mosley TH, van Duijn CM, Breteler MM, Longstreth WT, Wolf PA. (2009) Genomewide association studies of stroke. N Engl J Med 360:1718-1728

Jeffs B, Clark JS, Anderson NH, Gratton J, Brosnan MJ, Gauguier D, Reid JL, Macrae IM, Dominiczak AF. (1997) Sensitivity to cerebral ischaemic insult in a rat model of stroke is determined by a single genetic locus. Nat Genet 16:364-367

Jerrard-Dunne P, Cloud G, Hassan A, Markus HS. (2003) Evaluating the genetic component of ischemic stroke subtypes: a family history study. Stroke 34:1364-1369

Jin K, Wang X, Xie L, Mao XO, Zhu W, Wang Y, Shen J, Mao Y, Banwait S, Greenberg DA. (2006) Evidence for stroke-induced neurogenesis in the human brain. Proc Natl Acad Sci U S A 103:13198-13202

Jood K, Ladenvall C, Rosengren A, Blomstrand C, Jern C. (2005) Family History in Ischemic Stroke Before 70 Years of Age. The Sahlgrenska Academy Study on Ischemic Stroke. Stroke 36:1383-1387

Joutel A, Corpechot C, Ducros A, Vahedi K, Chabriat H, Mouton P, Alamowitch S, Domenga V, Cécillion M, Marechal E, Maciazek J, Vayssiere C, Cruaud C, Cabanis EA, Ruchoux MM, Weissenbach J, Bach JF, Bousser MG, Tournier-Lasserve E. (1996) Notch3 mutations in CADASIL, a hereditary adult-onset condition causing stroke and dementia. Nature 383:707-710

Kagiyama S, Fukuhara M, Matsumura K, Lin Y, Fujii K, Iida M. (2005) Central and peripheral cardiovascular actions of apelin in conscious rats. Regul Pept 125:55-59

Kalafatis M, Rand MD, Mann KG. (1994) The mechanism of inactivation of human factor V and human factor Va by activated protein C. J Biol Chem 269:31869-31880

Kasner SE. (2006) Clinical interpretation and use of stroke scales. Lancet Neurol 5:603-612

Kathiresan S, Melander O, Guiducci C, Surti A, Burtt NP, Rieder MJ, Cooper GM, Roos C, Voight BF, Havulinna AS, Wahlstrand B, Hedner T, Corella D, Tai ES, Ordovas JM, Berglund G, Vartiainen E, Jousilahti P, Hedblad B, Taskinen MR, Newton-Cheh C, Salomaa V, Peltonen L, Groop L, Altshuler DM, Orho-Melander M. (2008) Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. Nat Genet 40:189-197

Kathiresan S, Voight BF, Purcell S, Musunuru K, Ardissino D, Mannucci PM, Anand S, Engert JC, Samani NJ, Schunkert H, Erdmann J, Reilly MP, Rader DJ, Morgan T, Spertus JA, Stoll M, Girelli D, McKeown PP, Patterson CC, Siscovick DS, O'Donnell CJ, Elosua R, Peltonen L, Salomaa V, Schwartz SM, Melander O, Altshuler D, Merlini PA, Berzuini C, Bernardinelli L, Peyvandi F, Tubaro M, Celli P, Ferrario M, Fetiveau R, Marziliano N, Casari G, Galli M, Ribichini F, Rossi M, Bernardi F, Zonzin P, Piazza A, Yee J, Friedlander Y, Marrugat J, Lucas G, Subirana I, Sala J, Ramos R, Meigs JB, Williams G, Nathan DM, MacRae CA, Havulinna AS, Berglund G, Hirschhorn JN, Asselta R, Duga S, Spreafico M, Daly MJ, Nemesh J, Korn JM, McCarroll SA, Surti A, Guiducci C, Gianniny L, Mirel D, Parkin M, Burtt N, Gabriel SB, Thompson JR, Braund PS, Wright BJ, Balmforth AJ, Ball SG, Hall AS, Linsel-Nitschke P, Lieb W, Ziegler A, König I, Hengstenberg C, Fischer M, Stark K, Grosshennig A, Preuss M, Wichmann HE, Schreiber S, Ouwehand W, Deloukas P, Scholz M, Cambien F, Li M, Chen Z, Wilensky R, Matthai W, Qasim A, Hakonarson HH, Devaney J, Burnett MS, et al. (2009) Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. Nat Genet 41:334-341

Kim S, Iwao H. (2000) Molecular and cellular mechanisms of angiotensin II-mediated cardiovascular and renal diseases. Pharmacol Rev 52:11-34

Kirshner HS. (2009) Differentiating ischemic stroke subtypes: risk factors and secondary prevention. J Neurol Sci 279:1-8

Kleindorfer D, Kissela B, Schneider A, Woo D, Khoury J, Miller R, Alwell K, Gebel J, Szaflarski J, Pancioli A, Jauch E, Moomaw C, Shukla R, Broderick JP, Institute N. (2004) Eligibility for recombinant tissue plasminogen activator in acute ischemic stroke: a population-based study. Stroke 35:e27-29

Krupinski J, Kaluza J, Kumar P, Wang M, Kumar S. (1993) Prognostic value of blood vessel density in ischaemic stroke. Lancet 342:742

Kubo M, Hata J, Ninomiya T, Matsuda K, Yonemoto K, Nakano T, Matsushita T, Yamazaki K, Ohnishi Y, Saito S, Kitazono T, Ibayashi S, Sueishi K, Iida M, Nakamura Y, Kiyohara Y. (2007) A nonsynonymous SNP in PRKCH (protein kinase C eta) increases the risk of cerebral infarction. Nat Genet 39:212-217

Lakhan SE, Kirchgessner A, Hofer M. (2009) Inflammatory mechanisms in ischemic stroke: therapeutic approaches. J Transl Med 7:97

Lanktree MB, Dichgans M, Hegele RA. (2010) Advances in genomic analysis of stroke: what have we learned and where are we headed? Stroke 41:825-832

Leker RR, Soldner F, Velasco I, Gavin DK, Androutsellis-Theotokis A, McKay RD. (2007) Long-lasting regeneration after ischemia in the cerebral cortex. Stroke 38:153-161

Levy E, Carman MD, Fernandez-Madrid IJ, Power MD, Lieberburg I, van Duinen SG, Bots GT, Luyendijk W, Frangione B. (1990) Mutation of the Alzheimer's disease amyloid gene in hereditary cerebral hemorrhage, Dutch type. Science 248:1124-1126

Lisabeth LD, Smith MA, Brown DL, Uchino K, Morgenstern LB. (2005) Family history and stroke outcome in a bi-ethnic, population-based stroke surveillance study. BMC Neurol 5:20

Liu J, Sun K, Bai Y, Zhang W, Wang X, Wang Y, Wang H, Chen J, Song X, Xin Y, Liu Z, Hui R. (2009) Association of three-gene interaction among MTHFR, ALOX5AP and NOTCH3 with thrombotic stroke: a multicenter case-control study. Hum Genet 125:649-656

Lloyd-Jones D, Adams R, Carnethon M, De Simone G, Ferguson TB, Flegal K, Ford E, Furie K, Go A, Greenlund K, Haase N, Hailpern S, Ho M, Howard V, Kissela B, Kittner S, Lackland D, Lisabeth L, Marelli A, McDermott M, Meigs J, Mozaffarian D, Nichol G, O'Donnell C, Roger V, Rosamond W, Sacco R, Sorlie P, Stafford R, Steinberger J, Thom T, Wasserthiel-Smoller S, Wong N, Wylie-Rosett J, Hong Y, Subcommittee AHASCaSS. (2009) Heart disease and stroke statistics--2009 update: a report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Circulation 119:e21-181

MacLellan CL, Silasi G, Auriat AM, Colbourne F. (2010) Rodent models of intracerebral hemorrhage. Stroke 41:S95-98

Maguire J, Thakkinstian A, Levi C, Lincz L, Bisset L, Sturm J, Scott R, Whyte S, Attia J. (2010) Impact of COX-2 rs5275 and rs20417 and GPIIIa rs5918 Polymorphisms on 90-Day Ischemic Stroke Functional Outcome: A Novel Finding. J Stroke Cerebrovasc Dis

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM. (2009) Finding the missing heritability of complex diseases. Nature 461:747-753

Marchini J, Donnelly P, Cardon LR. (2005) Genome-wide strategies for detecting multiple loci that influence complex diseases. Nat Genet 37:413-417

Markus H. (2003) Stroke Genetics. Oxford: Oxford University Press

Markus H. (2004) Genes for stroke. J Neurol Neurosurg Psychiatry 75:1229-1231

Martínez-Fernández E, Gil-Peralta A, García-Lozano R, Chinchón I, Aguilera I, Fernández-López O, Arenas J, Campos Y, Bautista J. (2001) Mitochondrial disease and stroke. Stroke 32:2507-2510

Martínez-González NA, Sudlow CL. (2006) Effects of apolipoprotein E genotype on outcome after ischaemic stroke, intracerebral haemorrhage and subarachnoid haemorrhage. J Neurol Neurosurg Psychiatry 77:1329-1335

Matarín M, Brown WM, Scholz S, Simón-Sánchez J, Fung H, Hernandez D, Gibbs JR, Vrieze FWD, Crews C, Britton A, Langefeld CD, Brott TG, Brown Jr RD, Worrall BB, Frankel M, Silliman S, Case LD, Singleton A, Hardy JA, Rich SS, Meschia JF. (2007) A genome-wide genotyping

study in patients with ischaemic stroke: initial analysis and data release. Lancet Neurology 6:414-420

Matarin M, Simon-Sanchez J, Fung HC, Scholz S, Gibbs JR, Hernandez DG, Crews C, Britton A, De Vrieze FW, Brott TG, Brown RD, Jr., Worrall BB, Silliman S, Case LD, Hardy JA, Rich SS, Meschia JF, Singleton AB. (2008) Structural genomic variation in ischemic stroke. Neurogenetics 9:101-108

Matarin M, Singleton A, Hardy J, Meschia J. (2010) The genetics of ischaemic stroke. J Intern Med 267:139-155

Mayeux R. (2005) Mapping the new frontier: complex genetic disorders. J Clin Invest 115:1404-1407

McCarron MO, Muir KW, Nicoll JA, Stewart J, Currie Y, Brown K, Bone I. (2000) Prospective study of apolipoprotein E genotype and functional outcome following ischemic stroke. Arch Neurol 57:1480-1484

McCarron MO, Muir KW, Weir CJ, Dyker AG, Bone I, Nicoll JA, Lees KR. (1998) The apolipoprotein E epsilon4 allele and outcome in cerebrovascular disease. Stroke 29:1882-1887

Megherbi SE, Milan C, Minier D, Couvreur G, Osseby GV, Tilling K, Di Carlo A, Inzitari D, Wolfe CD, Moreau T, Giroud M. (2003) Association between diabetes and stroke subtype on survival and functional outcome 3 months after stroke: data from the European BIOMED Stroke Project. Stroke 34:688-694

Mehta SL, Manhas N, Raghubir R. (2007) Molecular targets in cerebral ischemia for developing novel therapeutics. Brain Res Rev 54:34-66

Meschia JF. (2004) Clinically translated ischemic stroke genomics. Stroke 35:2735-2739

Mitsios N, Gaffney J, Kumar P, Krupinski J, Kumar S, Slevin M. (2006) Pathophysiology of acute ischaemic stroke: an analysis of common signalling mechanisms and identification of new molecular targets. Pathobiology 73:159-175

Mommersteeg MT, Hoogaars WM, Prall OW, de Gier-de Vries C, Wiese C, Clout DE, Papaioannou VE, Brown NA, Harvey RP, Moorman AF, Christoffels VM. (2007) Molecular pathway for the localized formation of the sinoatrial node. Circ Res 100:354-362

Moore JH. (2003) The ubiquitous nature of epistasis in determining susceptibility to common human diseases. Hum Hered 56:73-82

Moore JH, Williams SM. (2005) Traversing the conceptual divide between biological and statistical epistasis: systems biology and a more modern synthesis. Bioessays 27:637-646

Munafò MR, Flint J. (2004) Meta-analysis of genetic association studies. Trends Genet 20:439-444

Nakatomi H, Kuriu T, Okabe S, Yamamoto S, Hatano O, Kawahara N, Tamura A, Kirino T, Nakafuku M. (2002) Regeneration of hippocampal pyramidal neurons after ischemic brain injury by recruitment of endogenous neural progenitors. Cell 110:429-441

Nilsson-Ardnor S, Janunger T, Wiklund PG, Lackovic K, Nilsson AK, Lindgren P, Escher SA, Stegmayr B, Asplund K, Holmberg D. (2007) Genome-wide linkage scan of common stroke in families from northern Sweden. Stroke 38:34-40

NINDS. (2004) Stroke: Hope Through Research. NIH Publication 99-2222

Orr N, Chanock S. (2008) Common genetic variation and human disease. Adv Genet 62:1-32

Palmer LJ, Cardon LR. (2005) Shaking the tree: mapping complex disease genes with linkage disequilibrium. Lancet 366:1223-1234

Paternoster L, Chen W, Sudlow CL. (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 40:2020-2026

Pearson TA, Manolio TA. (2008) How to interpret a genome-wide association study. JAMA 299:1335-1344

Peddareddygari LR, Dutra AV, Levenstien MA, Sen S, Grewal RP. (2009) An analysis of methylenetetrahydrofolate reductase and glutathione S-transferase omega-1 genes as modifiers of the cerebral response to ischemia. BMC Neurol 9:37

Pereira TV, Rudnicki M, Franco RF, Pereira AC, Krieger JE. (2007) Effect of the G-308A polymorphism of the tumor necrosis factor alpha gene on the risk of ischemic heart disease and ischemic stroke: a meta-analysis. Am Heart J 153:821-830

Peters N, Opherk C, Bergmann T, Castro M, Herzog J, Dichgans M. (2005) Spectrum of mutations in biopsy-proven CADASIL: implications for diagnostic strategies. Arch Neurol 62:1091-1094

Polychronopoulos P, Gioldasis G, Ellul J, Metallinos IC, Lekka NP, Paschalis C, Papapetropoulos T. (2002) Family history of stroke in stroke types and subtypes. J Neurol Sci 195:117-122

Revesz T, Holton JL, Lashley T, Plant G, Rostagno A, Ghiso J, Frangione B. (2002) Sporadic and familial cerebral amyloid angiopathies. Brain Pathol 12:343-357

Ritchie MD. (2011) Using biological knowledge to uncover the mystery in the search for epistasis in genome-wide association studies. Ann Hum Genet 75:172-182

Ritchie MD, Hahn LW, Moore JH. (2003) Power of multifactor dimensionality reduction for detecting gene-gene interactions in the presence of genotyping error, missing data, phenocopy, and genetic heterogeneity. Genet Epidemiol 24:150-157

Ritchie MD, Hahn LW, Roodi N, Bailey LR, Dupont WD, Parl FF, Moore JH. (2001) Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. Am J Hum Genet 69:138-147

Saenger AK, Christenson RH. (2010) Stroke biomarkers: progress and challenges for diagnosis, prognosis, differentiation, and treatment. Clin Chem 56:21-33

Sahni R, Weinberger J. (2007) Management of intracerebral hemorrhage. Vasc Health Risk Manag 3:701-709

Sarzynska-Dlugosz I, Gromadzka G, Baranska-Gieruszczak M, Ciesielska A, Czlonkowska A. (2007) APOE does not predict poor outcome 1 year after ischemic stroke. Neurol Res 29:64-69

Schäbitz WR, Steigleder T, Cooper-Kuhn CM, Schwab S, Sommer C, Schneider A, Kuhn HG. (2007) Intravenous brain-derived neurotrophic factor enhances poststroke sensorimotor recovery and stimulates neurogenesis. Stroke 38:2165-2172

Schulz UG, Flossmann E, Rothwell PM. (2004) Heritability of ischemic stroke in relation to age, vascular risk factors, and subtypes of incident stroke in population-based studies. Stroke 35:819-824

Seyedabadi M, Goodchild AK, Pilowsky PM. (2002) Site-specific effects of apelin-13 in the rat medulla oblongata on arterial pressure and respiration. Auton Neurosci 101:32-38

Sharma P, Carter ND, Barley J, Lunt R, Seymour CA, Brown MM. (1994) Polymorphisms in the gene encoding angiotensin 1-converting enzyme and relationship to its post-translational product in cerebral infarction. J Hum Hypertens 8:633-634

Shen CD, Zhang WL, Sun K, Wang YB, Zhen YS, Hui RT. (2007) Interaction of genetic risk factors confers higher risk for thrombotic stroke in male Chinese: a multicenter case-control study. Ann Hum Genet 71:620-629

Shendure J, Ji H. (2008) Next-generation DNA sequencing. Nat Biotechnol 26:1135-1145

Slevin M, Kumar P, Gaffney J, Kumar S, Krupinski J. (2006) Can angiogenesis be exploited to improve stroke outcome? Mechanisms and therapeutic potential. Clin Sci (Lond) 111:171-183

Small DL, Buchan AM. (2000) Animal models. Br Med Bull 56:307-317

Spanbroek R, Grabner R, Lotzer K, Hildner M, Urbach A, Ruhling K, Moos MP, Kaiser B, Cohnert TU, Wahlers T, Zieske A, Plenz G, Robenek H, Salbach P, Kuhn H, Radmark O, Samuelsson B, Habenicht AJ. (2003) Expanding expression of the 5-lipoxygenase pathway within the arterial wall during human atherogenesis. Proc Natl Acad Sci U S A 100:1238-1243

Steinberg MH, Adewoye AH. (2006) Modifier genes and sickle cell anemia. Curr Opin Hematol 13:131-136

Sudlow C, MartÃ-nez GonzÃ¡lez NA, Kim J, Clark C. (2006) Does apolipoprotein E genotype influence the risk of ischemic stroke, intracerebral hemorrhage, or subarachnoid hemorrhage? Systematic review and meta-analyses of 31 studies among 5961 cases and 17,965 controls. Stroke 37:364-370

Switzer JA, Hess DC, Nichols FT, Adams RJ. (2006) Pathophysiology and treatment of stroke in sickle-cell disease: present and future. Lancet Neurol 5:501-512

Tilling K, Sterne JA, Rudd AG, Glass TA, Wityk RJ, Wolfe CD. (2001) A new method for predicting recovery after stroke. Stroke 32:2867-2873

Tiret L, Rigat B, Visvikis S, Breda C, Corvol P, Cambien F, Soubrier F. (1992) Evidence, from combined segregation and linkage analysis, that a variant of the angiotensin I-converting enzyme (ACE) gene controls plasma ACE levels. Am J Hum Genet 51:197-205

Treger I, Froom P, Ring H, Friedman G. (2003) Association between apolipoprotein E4 and rehabilitation outcome in hospitalized ischemic stroke patients. Arch Phys Med Rehabil 84:973-976

Truelsen T, Ekman M, Boysen G. (2005) Cost of stroke in Europe. Eur J Neurol 12 Suppl 1:78-84

Tsuji K, Aoki T, Tejima E, Arai K, Lee SR, Atochin DN, Huang PL, Wang X, Montaner J, Lo EH. (2005) Tissue plasminogen activator promotes matrix metalloproteinase-9 upregulation after focal cerebral ischemia. Stroke 36:1954-1959

Tyler AL, Asselbergs FW, Williams SM, Moore JH. (2009) Shadows of complexity: what biological networks reveal about epistasis and pleiotropy. Bioessays 31:220-227

van Swieten JC, Koudstaal PJ, Visser MC, Schouten HJ, van Gijn J. (1988) Interobserver agreement for the assessment of handicap in stroke patients. Stroke 19:604-607

Wald DS, Law M, Morris JK. (2002) Homocysteine and cardiovascular disease: evidence on causality from a meta-analysis. BMJ 325:1202

Wang X, Cheng S, Brophy VH, Erlich HA, Mannhalter C, Berger K, Lalouschek W, Browner WS, Shi Y, Ringelstein EB, Kessler C, Luedemann J, Lindpaintner K, Liu L, Ridker PM, Zee RY, Cook NR, Consortium RMSSSNP. (2009) A meta-analysis of candidate gene polymorphisms and ischemic stroke in 6 study populations: association of lymphotoxin-alpha in nonhypertensive patients. Stroke 40:683-695

Warlow C, Sudlow C, Dennis M, Wardlaw J, Sandercock P. (2003) Stroke. Lancet 362:1211-1224

Weimar C, Roth M, Willig V, Kostopoulos P, Benemann J, Diener HC. (2006) Development and validation of a prognostic model to predict recovery following intracerebral hemorrhage. J Neurol 253:788-793

Weimar C, Ziegler A, Konig IR, Diener HC. (2002) Predicting functional outcome and survival after acute ischemic stroke. J Neurol 249:888-895

Weisscher N, Vermeulen M, Roos YB, De Haan RJ. (2008) What should be defined as good outcome in stroke trials; a modified Rankin score of 0-1 or 0-2? J Neurol 255:867-874

Whisnant JP. (1997) Modeling of risk factors for ischemic stroke. The Willis Lecture. Stroke 28:1840-1844

WHO. (1988) The World Health Organization MONICA Project (monitoring trends and determinants in cardiovascular disease): a major international collaboration. WHO MONICA Project Principal Investigators. J Clin Epidemiol 41:105-114

Wong AA, Davis JP, Schluter PJ, Henderson RD, O'Sullivan JD, Read SJ. (2005) The effect of admission physiological variables on 30 day outcome after stroke. J Clin Neurosci 12:905-910

Wu TC, Grotta JC. (2010) Stroke treatment and prevention: five new things. Neurology 75:S16-21

Xi G, Keep RF, Hoff JT. (2006) Mechanisms of brain injury after intracerebral haemorrhage. Lancet Neurol 5:53-63

Yamada Y, Fuku N, Tanaka M, Aoyagi Y, Sawabe M, Metoki N, Yoshida H, Satoh K, Kato K, Watanabe S, Nozawa Y, Hasegawa A, Kojima T. (2009) Identification of CELSR1 as a susceptibility gene for ischemic stroke in Japanese individuals by a genome-wide association study. Atherosclerosis 207:144-149

Zintzaras E, Rodopoulou P, Sakellaridis N. (2009) Variants of the arachidonate 5-lipoxygenase-activating protein (ALOX5AP) gene and risk of stroke: a HuGE gene-disease association review and meta-analysis. Am J Epidemiol 169:523-532

Chapter 2. Matrix metalloproteinase genes

Manso H, Krug T, Sobral J, Albergaria I, Gaspar G, Ferro JM, Oliveira SA, Vicente AM (2010) Variants of the Matrix Metalloproteinase-2 but not the Matrix Metalloproteinase-9 genes significantly influence functional outcome after stroke. BMC Med Genet, 11:40.

Abstract

Background – Multiple lines of evidence suggest that genetic factors contribute to stroke recovery. The matrix metalloproteinases-2 (MMP-2) and -9 (MMP-9) are modulators of extracellular matrix components, with important regulatory functions in the Central Nervous System (CNS). Shortly after stroke, MMP-2 and MMP-9 have mainly damaging effects for brain tissue. However, MMPs also have a beneficial activity in angiogenesis and neurovascular remodeling during the delayed neuroinflammatory response phase, thus possibly contributing to stroke functional recovery.

Methods – In the present study, the role of *MMP2* and *MMP9* genetic variants in stroke recovery was investigated in 546 stroke patients. Functional outcome was assessed three months after a stroke episode using the modified Rankin Scale (mRS), and patients were classified in two groups: good recovery (mRS≤1) or poor recovery (mRS>1). Haplotype tagging single nucleotide polymorphisms (SNPs) in the *MMP2* (N=21) and *MMP9* (N=4) genes were genotyped and tested for association with stroke outcome, adjusting for significant non-genetic clinical variables.

Results – Six SNPs in the *MMP2* gene were significantly associated with stroke outcome ($0.002<P<0.042$), two of which survived the Bonferroni correction for multiple testing. In the subset of ischemic stroke patients, association of five of these SNPs remained positive ($0.004<P<0.031$). No significant associations were found for the *MMP9* gene.

Conclusions – The results presented strongly indicate that *MMP2* genetic variants are an important mediator of functional outcome after stroke.

Background

While remaining one of the most common causes of death worldwide, stroke is also a leading cause of significant disability: after a first stroke event, 50-70% of stroke patients regain functional independence, but 15-30% are permanently disabled and 20% require institutional care at 3 months after onset [Asplund et al. 1998]. Clinical and demographic factors can influence stroke outcome. In addition, genetic factors are likely to have an impact in stroke recovery processes and outcome: family history of stroke is associated with stroke outcome [Jood et al. 2005; Lisabeth et al. 2005] and many animal models of stroke implicate genes that regulate angiogenesis, neuronal regeneration and proliferation, and neuroinflammation, in stroke recovery [McColl et al. 2007; Nygren et al. 2006; Sun et al. 2003; Svedin et al. 2007].

Several lines of evidence suggest that matrix metalloproteinases (MMPs) are fundamental players in stroke recovery. These molecules belong to a family of zinc-dependent endopeptidases that modulate extracellular matrix (ECM) components in many Central Nervous System (CNS) developmental and regenerative processes such as neurogenesis, axonal growth and regeneration, and myelin formation. The expression and activity of MMPs is tightly regulated. Most MMPs require proteolytic processing by proteases or other MMPs to become activated, and can be inhibited by tissue inhibitors of metalloproteinases (TIMPs). Dysregulated MMP activity will lead to uncontrolled degradation of ECM and basal lamina proteins, with serious harmful effects for the blood-brain barrier (BBB) integrity and neuroinflammatory or neurotoxic consequences [Candelario-Jalil et al. 2009; Rosell and Lo 2008]. Such dysregulation of MMPs is known to occur after stroke, leading to a degradation of the neurovascular matrix, disrupting cell-matrix homeostasis and weakening the BBB, and thus contributing to cell death, neurotoxicity, edema and hemorrhage [Gu et al. 2005; Rosell and Lo 2008]. The variation profiles of MMPs in blood after a stroke event [Horstmann et al. 2003; Horstmann et al. 2006] suggest that these molecules can eventually be used as biomarkers for brain damage and neurological outcome, while their contribution to tissue destruction renders MMPs inhibitors potentially interesting therapeutic targets for stroke.

Emerging studies, however, indicate that MMPs may also have a beneficial activity in angiogenesis and neurovascular remodeling during the delayed neuroinflammatory response phase after stroke, possibly contributing to stroke functional recovery [Rosell and Lo 2008].

While inhibition of MMP activity has consistently been demonstrated to be effective in reducing edema, infarct size and hemorrhagic transformation, some studies suggest the existence of a time window for these beneficial effects to take place [Sood et al. 2008; Zhao et al. 2006].

In the present study we tested the impact of genetic variants in *MMP2* and *MMP9* in stroke recovery, in a population sample of 546 patients evaluated for stroke outcome at three months after the stroke event.

Methods

Participants in the present study were recruited in the context of a wider research project to evaluate stroke risk factors in a Portuguese population sample, which enrolled first-ever stroke patients under 65 years of age through Neurology and Internal Medicine Departments of several hospitals in Portugal. Stroke was defined as a focal neurological deficit of sudden or rapid onset lasting more than 24 hours, and classified into ischemic or intracerebral hemorrhage based on brain imaging (computed tomography and/or magnetic resonance imaging). The diagnosis of stroke was confirmed by a neurologist. Demographic characteristics (age and gender), information on previous vascular risk factors and comorbid conditions (diabetes mellitus, hypertension, cardiac disease, dyslipidemia, obesity), life-style risk factors (smoking, alcohol consumption, physical inactivity and others), and detailed clinical data during hospitalization, including neurological symptoms, complications and interventions, were collected for the majority of patients. Occurrence of aphasia, neglect, paresis, gaze paresis, dysphagia, permanent altered consciousness, urinary incontinence and medical and neurological complications were clinical parameters indicative of stroke severity. Stroke outcome at discharge and at three months was assessed, by direct interview, using the modified Rankin Scale (mRS).

For the present study, 568 patients with relevant clinical data and a DNA sample were available. Eight patients had a second stroke event after enrolment, affecting patient recovery, and were thus excluded. Of the remaining 560, 14 did not return after discharge for the three months evaluation, and therefore only 546 patients were included in the analysis. Patients were classified in two groups, according to their mRS at three months: patients with mRS≤1 were assigned to the “good recovery” group and patients with mRS>1 were assigned to the

"poor recovery" group (handicapped patients). 276 individuals were included in the good recovery group (63.0% males and 37.0% females) and 270 in the poor recovery group (64.4% males, 35.6% females). The poor recovery group included 12 patients who died before the three months evaluation (seven of them before hospital discharge, and five others after discharge). Genetic power calculations were performed using the CaTS software [Skol et al. 2006].

The study was approved by the Ethics Committee of Instituto Nacional de Saúde Dr. Ricardo Jorge and other hospitals involved, subjects gave informed consent and procedures followed were in accordance with institutional guidelines.

Single nucleotide polymorphisms (SNPs) within the *MMP2* and *MMP9* genes and up to 5kb of the flanking regions were selected using the Haploview software (v4.0) [Barrett et al. 2005], based on their tagging potential (HapMap Release 21/phase II July 2006). 4 SNPs in *MMP9* and 20 SNPs in *MMP2* were genotyped using the Sequenom iPLEX assays with allele detection by mass spectroscopy, using Sequenom MassARRAY technology (Sequenom, San Diego, USA) and following the manufacturer's protocol. Primer sequences were designed using Sequenom's MassARRAY Assay Design 3.0 software. 1 SNP in *MMP2* was genotyped using TaqMan® Pre-Designed SNP Genotyping Assays, in an ABI PRISM 7900HT Sequence Detector System (Applied Biosystems, Foster City, USA). Extensive quality control was performed using eight HapMap individuals, duplicated samples within and across genotyping plates, Mendelian segregation in three pedigrees and no-template samples. Call rates <90% and deviation from Hardy-Weinberg equilibrium led to SNP exclusion from the analysis. 2 SNPs in *MMP9* failed quality control and were substituted. In total, 21 *MMP2* SNPs and 4 *MMP9* SNPs were analyzed.

The effect of discrete and continuous non-genetic variables on stroke outcome at three months was determined using the Pearson's χ^2 test and Mann-Whitney test, respectively. These included age, gender, stroke risk factors as well as data on clinical variables collected during hospitalization (like occurrence of paresis, aphasia and medical complications). Variables with a $P<0.25$ in univariate analysis or of particular clinical relevance were included in a logistic regression model using forward selection [Hosmer and Lemeshow 2000] and were maintained in the model if they were associated at a $P\leq0.05$ level with stroke outcome. Logistic regression analyses were then used to determine the effect of each genetic variable on stroke outcome after adjustment for those significant non-genetic variables. Odds ratio (OR) and 95% confidence intervals (95% CI) were computed for the log-additive model.

Univariate and logistic regression analyses were performed using MASS and SNPassoc packages of the R software [R: A language and Environment for Statistical Computing 2004] (v2.6.0). The Gabriel et al. (2002) [Gabriel et al. 2002] default method of the Haploview software [Barrett et al. 2005] (v4.0) was used to determine haplotype blocks in the *MMP2* and *MMP9* genes. Since recovery processes may be regulated differently in ischemic and hemorrhagic stroke patients, we performed the same analyses in the subset of ischemic stroke patients. The small number of hemorrhagic stroke patients (N=105) precluded the independent analysis of this subset.

Significant associations in individual SNP analysis were corrected for multiple testing using the Bonferroni method. The alternative SNPSpD approach, based on the spectral decomposition (SpD) of matrices of pairwise linkage disequilibrium (LD) between SNPs was also applied [Nyholt 2004]. Since some of the 21 SNPs genotyped in the *MMP2* gene are in LD with each other in our sample, we used the SNPSpD approach to estimate the effective number of independent SNPs in our sample for multiple testing corrections.

Results

Clinical and demographic characteristics of the population sample are presented in Table 1. Univariate analysis showed that type of stroke and six clinical features indicative of stroke severity – occurrence of aphasia, urinary incontinence, paresis, altered consciousness, medical and neurological complications during hospitalization – were significant predictors of poor outcome. Sex ratio, age, and stroke risk factors were similar between the poor and good recovery groups, and approximately the same proportion of patients was being treated for hypertension in either group (34.0% and 34.6% in the good and poor recovery groups, respectively). Assuming an additive genetic model and phenotype allele frequency of 30%, our sample was 82% powered to detect a genotype relative risk of 1.5 with a type I error of 5%.

Of 21 *MMP2* SNPs, six were associated with stroke outcome at three months under a log-additive model ($0.002<P<0.042$) after adjusting for the significant covariates in a multivariate model: history of hypertension, type of stroke, occurrence of aphasia, paresis, altered consciousness and medical complications during hospitalization (Table 2; see Supplementary table 1). History of hypertension, although not associated in the univariate

analysis, became significant in the multivariate model before inclusion of genetic variants, and was therefore included in the final regression model. SNPs rs2241145 and rs1992116 remained significantly associated with stroke outcome after Bonferroni correction for multiple testing (OR[95%CI]=1.66[1.20-2.30], $_{corrected}P$=0.044, and OR[95%CI]=1.67[1.20-2.31], $_{corrected}P$=0.039, respectively). Two haplotypes (one of which rare) were nominally associated with stroke outcome at three months (Table 3, Figure 1A; see Supplementary table 2).

The hypothesis that the recovery processes after ischemic and hemorrhagic stroke may be different and regulated by different sets of genes [Mehta et al. 2007; Xi et al. 2006] led us to analyze the ischemic stroke subset independently. The haemorrhagic subset was too small for independent analysis (N=105). In the ischemic stroke sample, five out of the previously associated SNPs in the *MMP2* gene remained significantly associated with stroke outcome at three months under a log-additive model (0.004<*P*<0.031), after adjusting for the same significant covariates (excluding type of stroke) (Table 2; see Supplementary table 1). ORs for these SNPs in this subset were similar to the overall study sample. None of the SNPs remained significant after Bonferroni correction for multiple testing. However, when the SNPSpD method was used, taking into account regional LD patterns and therefore the number of SNPs which are effectively independent, the two SNPs that survived Bonferroni correction in the whole sample remained significant for the ischemic stroke subset (rs2241145 and rs1992116) (see Supplementary table 1). Four additional *MMP2* SNPs were nominally associated with ischemic stroke outcome at three months (0.016<*P*<0.041, Table 2). Only one haplotype in *MMP2* was also associated (Table 3; see Supplementary table 2).

OR analysis indicates that, for the majority of significantly associated SNPs (including rs2241145 and rs1992116), carriers of the minor allele (less frequent allele) are significant predictors of poor outcome (OR>1); only for rs243842 in the whole population sample, and for rs857403 and rs183112 in the ischemic subset, carriers of the minor allele show an improved chance of good recovery from stroke (OR<1).

In the *MMP9* gene, one rare haplotype was associated with stroke outcome in the overall population sample (*P*=0.007, Table 3, Figure 1B; see Supplementary table 2), but no independent association was found for any of the four tested SNPs (see Supplementary table 1). No SNP or haplotype in the *MMP9* gene was associated with stroke outcome at three months in the ischemic subset (see Supplementary tables 1 and 2).

None of the tested SNPs were associated with hypertension, indicating that the MMP-2 effect on recovery was not mediated by its role on vascular structure (data not shown).

Two of the *MMP2* SNPs (rs1053605 and rs243849) are located in exonic regions of the *MMP2* gene (exons 5 and 7, respectively), two SNPs (rs243866 and rs243865) are located upstream of the gene, and six SNPs are intronic (Figure 1A). Both nucleotide transitions in the exonic SNPs are silent. To investigate possible functional consequences for gene transcription of the two upstream SNPs (rs243866 and rs243865) and the two intronic SNPs that survived correction for multiple testing (rs2241145 and rs1992116), we conducted a bioinformatics search for putative transcription factor binding sites. The A allele of the upstream SNP rs243866 lies in the core of a sequence with high similarity to the matrix for two binding factors, the IPF1 (insulin promoter factor 1), and the POU5F1 (POU domain class 5 transcription factor 1). Both proteins are transcription activators. Since the AA and AG genotypes are more frequent in the poor recovery group, we can hypothesize that the presence of the A allele may lead to an increased transcription of the *MMP2* gene, and thus explain the negative impact on stroke recovery observed in this population sample. The presence of the T allele in the upstream rs243865 SNP forms a sequence with high similarity to the matrix for the PLZF binding factor (promyelocytic leukemia zinc finger protein), while the sequence containing the C allele has a stronger similarity with the matrix for the VDR/RXR (vitamin D hormone receptor/retinoid X receptor) heterodimer. However, both transcription factors act as repressors, and therefore these findings are more difficult to interpret. The rs2241145 and rs1992116 intronic SNPs did not contain sequences for any known putative transcription factor binding sites.

Discussion

In the present study we show that *MMP2* gene variants are strongly associated with patient's functional disability at three months after stroke onset, in a large Portuguese population sample. Given the possible genetic heterogeneity in recovery processes after hemorrhagic and ischemic stroke [Mehta et al. 2007; Xi et al. 2006], we also analyzed the association of this gene with stroke outcome in the restricted subgroup of ischemic stroke patients. All but one *MMP2* gene variants associated with stroke in the overall population sample remained associated with ischemic stroke in this smaller subset. Additional markers were associated only in this subset, possibly reflecting the increased genetic homogeneity of the ischemic group in terms of recovery processes. Associated SNPs in the ischemic subset

did not, however, withstand Bonferroni correction for multiple testing. This could reflect the reduction in power due to the smaller sample size in the restricted analysis and/or the overcorrection for the false positive rate that is the main frequent criticism for this method. In fact, the alternative SNPSpD approach [Nyholt 2004], which takes into account LD patterns between genotyped SNPs in the tested population, may be more appropriate since the 21 genotyped *MMP2* SNPs are not independent; with this approach, the significance of association of two specific SNPs with stroke, in the ischemic subset or in the overall population sample, was retained after multiple testing correction. The association results after multiple testing correction, using the stringent Bonferroni method or the SNPSpD approach, strongly support a role for *MMP2* in stroke recovery. Validation through replication in a larger sample set by other groups is now advisable.

A limitation of the present study was the lack of availability of the National Institute of Health Stroke Scale (NIHSS) for these patients. To control for the effect of the severity of stroke in patients' outcome, we performed a logistic regression analysis using, as covariates, individual clinical variables associated with stroke clinical severity in our sample. Each selected variable was entered in the logistic regression model to identify those behaving as clinical predictors of stroke outcome. While this approach may not be as comprehensive as a widely used severity scale, it allowed us to include in the analysis parameters that reflect the severity of the event and, to a certain extent, patient's status at baseline.

While subject of controversy, the cut-off for the good and poor recovery groups was set between 1 and 2 because we chose to focus on a non-handicaped recovery group. According to Weisscher et al. (2008) [Weisscher et al. 2008], there is a clear lag on performance of outdoor activities between mRS 1 and 2, while between mRS 2 and 3 the major difference is the ability to perform complex activities of daily life, and thus a more clearly defined good outcome is given by setting the cut-off between mRS 1 and 2.

Multiple studies in animal models and humans have shown that the actions of MMPs contribute to BBB disruption and brain cell death, early after a stroke event. These damaging processes can be inhibited by MMP inhibitors, leading to reductions in infarct volume and significant improvements in behavioural scores compared with controls [Gu et al. 2005]. However, fitting with their role in development and regeneration, a beneficial influence of MMPs in the recovery processes that occur in later stages after a stroke event, including angiogenesis, remyelination, neural migration and general recovery of the neurovascular unit has been shown [Girolamo et al. 2004; Hsu et al. 2006; Sood et al. 2008; Zhao et al. 2006]. At present, we cannot dissect whether gene variation in *MMP2* is more important for the

damaging effects in the earlier stages after stroke, or to the beneficial delayed responses, or both. Functional studies will be required to answer this question. However, the present findings may have important implications. On one hand it challenges the usefulness of MMP inhibitors for the treatment of stroke, not only because the time window of usefulness is likely limited, but also because it may depend on the individual's *MMP2* genotype. On the other hand, and given that MMP-2 has also been suggested to influence the risk of hemorrhagic transformation upon recombinant tissue plasminogen activator (tPA) therapy [Liu et al. 2006], it is a plausible hypothesis that treatment outcome may also be associated with *MMP2* gene variants. Further work needs to be carried out to elucidate these questions.

Conclusions

The present study further reinforces the contribution of MMPs for stroke recovery by showing that specific *MMP2*, but not *MMP9*, gene variants influence stroke outcome. Replication of these associations in larger population samples, together with approaches that integrate evidence from multiple levels, including gene expression and functional analysis, will contribute for the validation of these results. Together with previous observations, the study leads to the hypothesis that individual variation in the *MMP2* gene may influence stroke treatment outcome.

Acknowledgements

The authors are grateful to all study participants and their families. The study subjects in this study were recruited in the context of an earlier study designated “Acidentes Vasculares Cerebrais antes dos 65 anos”, funded by Fundação para a Ciência e a Tecnologia (FCT) (PECS/T/SAU/179/95). The authors wish to thank Dr. Marinho Falcão and his team at Instituto Nacional de Saúde Dr. Ricardo Jorge, and all the clinicians that recruited study subjects from the following hospitals: H. S.João, H. Évora, H. Funchal, H. Marmeleiros, H. S.Bento, H. S.José, H. S.Marcos, H. Garcia d'Orta, H. Faro, H. Coimbra, H. Vila Nova de Gaia, H. Aveiro, SAMS, H. Capuchos and H. Sto.António. The authors also wish to thank the technical assistance provided by the Genotyping Unit at Instituto Gulbenkian de Ciência.

This work was supported in part by the Marie Curie International Reintegration Grant 513760 (SAO), the Marie Curie Intra-European Fellowship 024563 (SAO), the FCT grant PTDC/SAU-GMG/64426/2006, and fellowships from FCT (HM, TK) and the Portuguese Instituto do Emprego e Formação Profissional (TK).

References

Asplund K, Stegmayr B, Peltonen M. (1998) From the twentieth to the twenty-first century: a public health perspective on stroke. In: *Cerebrovascular Disease Pathophysiology, Diagnosis, and Management.* (Ginsberg MD, Bogousslavsky J, eds), Oxford, UK: Blackwell Science, pp 901-918

Barrett JC, Fry B, Maller J, Daly MJ. (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21:263-265

Candelario-Jalil E, Yang Y, Rosenberg GA. (2009) Diverse roles of matrix metalloproteinases and tissue inhibitors of metalloproteinases in neuroinflammation and cerebral ischemia. Neuroscience 158:983-994

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D. (2002) The structure of haplotype blocks in the human genome. Science 296:2225-2229

Girolamo F, Virgintino D, Errede M, Capobianco C, Bernardini N, Bertossi M, Roncali L. (2004) Involvement of metalloprotease-2 in the development of human brain microvessels. Histochem Cell Biol 122:261-270

Gu Z, Cui J, Brown S, Fridman R, Mobashery S, Strongin AY, Lipton SA. (2005) A highly specific inhibitor of matrix metalloproteinase-9 rescues laminin from proteolysis and neurons from apoptosis in transient focal cerebral ischemia. J Neurosci 25:6401-6408

Horstmann S, Kalb P, Koziol J, Gardner H, Wagner S. (2003) Profiles of matrix metalloproteinases, their inhibitors, and laminin in stroke patients: influence of different therapies. Stroke 34:2165-2170

Horstmann S, Su Y, Koziol J, Meyding-Lamade U, Nagel S, Wagner S. (2006) MMP-2 and MMP-9 levels in peripheral blood after subarachnoid hemorrhage. J Neurol Sci 251:82-86

Hosmer DW, Lemeshow S. (2000) Applied Logistic Regression. New York: John Wiley & Sons, Inc.

Hsu JY, McKeon R, Goussev S, Werb Z, Lee JU, Trivedi A, Noble-Haeusslein LJ. (2006) Matrix metalloproteinase-2 facilitates wound healing events that promote functional recovery after spinal cord injury. J Neurosci 26:9841-9850

Jood K, Ladenvall C, Rosengren A, Blomstrand C, Jern C. (2005) Family History in Ischemic Stroke Before 70 Years of Age. The Sahlgrenska Academy Study on Ischemic Stroke. Stroke 36:1383-1387

Lisabeth LD, Smith MA, Brown DL, Uchino K, Morgenstern LB. (2005) Family history and stroke outcome in a bi-ethnic, population-based stroke surveillance study. BMC Neurol 5:20

Liu XS, Zhang ZG, Zhang L, Morris DC, Kapke A, Lu M, Chopp M. (2006) Atorvastatin downregulates tissue plasminogen activator-aggravated genes mediating coagulation and vascular permeability in single cerebral endothelial cells captured by laser microdissection. J Cereb Blood Flow Metab 26:787-796

McColl BW, McGregor AL, Wong A, Harris JD, Amalfitano A, Magnoni S, Baker AH, Dickson G, Horsburgh K. (2007) APOE epsilon3 gene transfer attenuates brain damage after experimental stroke. J Cereb Blood Flow Metab 27:477-487

Mehta SL, Manhas N, Raghubir R. (2007) Molecular targets in cerebral ischemia for developing novel therapeutics. Brain Res Rev 54:34-66

Nygren J, Kokaia M, Wieloch T. (2006) Decreased expression of brain-derived neurotrophic factor in BDNF(+/-) mice is associated with enhanced recovery of motor performance and increased neuroblast number following experimental stroke. J Neurosci Res 84:626-631

Nyholt DR. (2004) A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. Am. J. Hum. Genet. 74:765-769

R: A language and Environment for Statistical Computing. (2004). Vienna

Rosell A, Lo EH. (2008) Multiphasic roles for matrix metalloproteinases after stroke. Curr Opin Pharmacol 8:82-89

Skol AD, Scott LJ, Abecasis GR, Boehnke M. (2006) Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. Nature Genetics 38:209-213

Sood RR, Taheri S, Candelario-Jalil E, Estrada EY, Rosenberg GA. (2008) Early beneficial effect of matrix metalloproteinase inhibition on blood-brain barrier permeability as measured by magnetic resonance imaging countered by impaired long-term recovery after stroke in rat brain. J Cereb Blood Flow Metab 28:431-438

Sun Y, Jin K, Xie L, Childs J, Mao XO, Logvinova A, Greenberg DA. (2003) VEGF-induced neuroprotection, neurogenesis, and angiogenesis after focal cerebral ischemia. J Clin Invest 111:1843-1851

Svedin P, Hagberg H, Savman K, Zhu C, Mallard C. (2007) Matrix metalloproteinase-9 gene knock-out protects the immature brain after cerebral hypoxia-ischemia. J Neurosci 27:1511-1518

Weisscher N, Vermeulen M, Roos YB, De Haan RJ. (2008) What should be defined as good outcome in stroke trials; a modified Rankin score of 0-1 or 0-2? J Neurol 255:867-874

Xi G, Keep RF, Hoff JT. (2006) Mechanisms of brain injury after intracerebral haemorrhage. Lancet Neurol 5:53-63

Zhao BQ, Wang S, Kim HY, Storrie H, Rosen BR, Mooney DJ, Wang X, Lo EH. (2006) Role of matrix metalloproteinases in delayed cortical responses after stroke. Nat Med 12:441-445

Figures

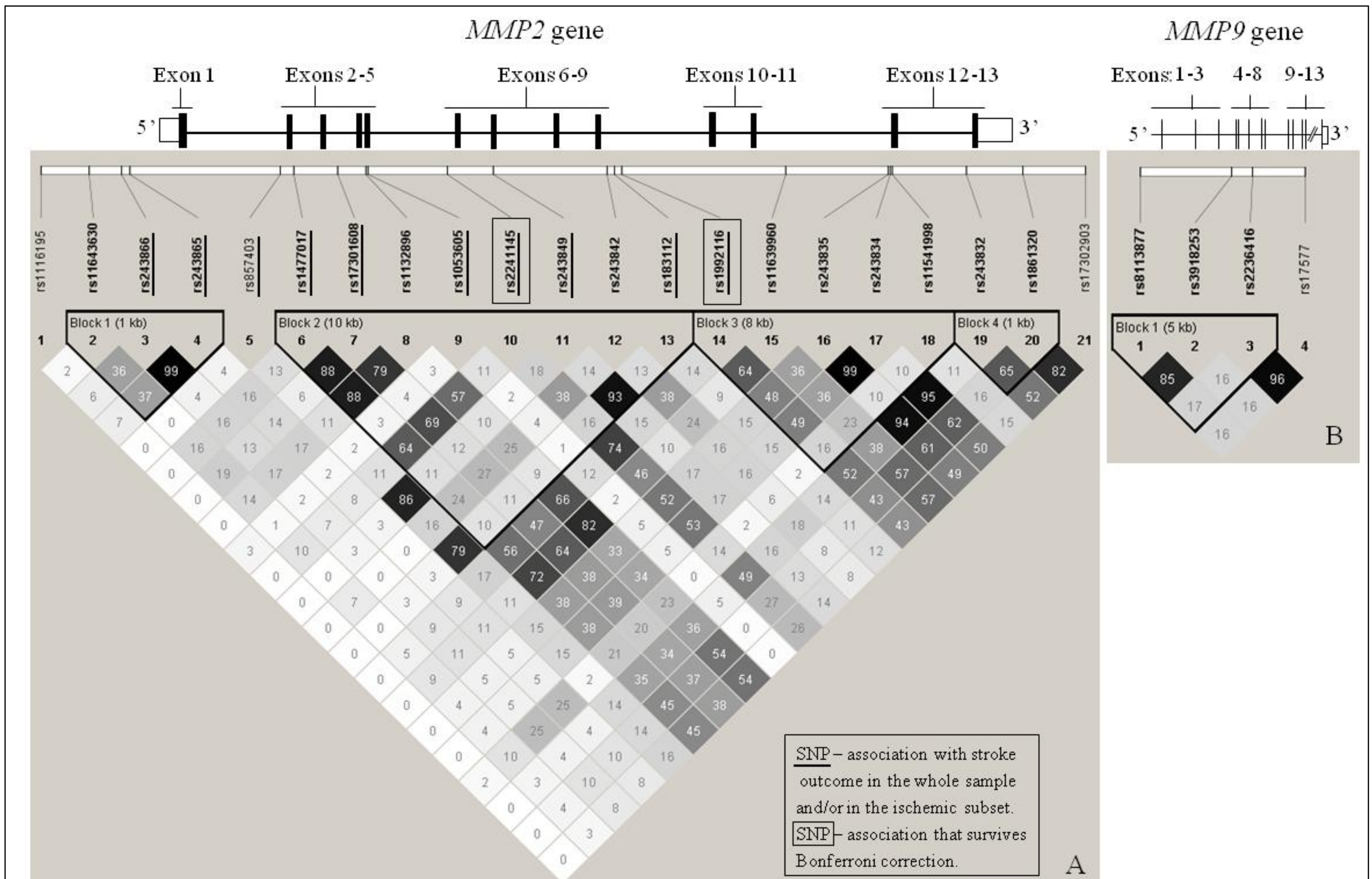

Figure 1 – Schematic diagrams of the *MMP2* (A) and *MMP9* (B) genes showing the location of the 13 exons (black boxes), the 5' and 3' untranslated regions (white boxes) and the pairwise r^2 plots for the 21 genotyped SNPs in *MMP2* and 4 genotyped SNPs in *MMP9*, in our population sample. Markers associated with three months outcome are indicated. Linkage disequilibrium blocks were generated using the Gabriel et al. [2002] method.

Tables

Table 1 – Demographic and clinical characteristics of stroke patients.

Characteristic	Good Recovery (mRS≤1)	Poor Recovery (mRS>1)	*P* *
Age and Gender			
Age, mean±SD (yrs)	50.8±9	52.5±8.5	0.028
Gender (male), n/N (%)	174/276 (63.0)	174/270 (64.4)	0.734
Past History, n/N (%)			
Hypertension	159/241 (66.0)	143/240 (59.6)	0.147
Diabetes	36/259 (13.9)	47/246 (19.1)	0.115
Cardiac Disease	37/264 (14.0)	43/257 (16.7)	0.390
Stroke type, n/N (%)			7.05×10^{-5}
Ischemic stroke	238/276 (86.2)	193/270 (71.5)	–
Hemorrhagic stroke	33/276 (12.0)	72/270 (26.7)	–
Unknow type of stroke	5/276 (1.8)	5/270 (1.9)	–
Stroke Features, n/N (%)			
Aphasia	53/258 (20.5)	98/250 (39.2)	4.23×10^{-6}
Neglect	11/266 (4.1)	19/240 (7.9)	0.072
Dysphagia	15/270 (5.6)	25/251 (10.0)	0.059
Urinary Incontinence	5/272 (1.8)	15/251 (6.0)	0.014
Paresis	203/273 (74.4)	244/269 (90.7)	5.59×10^{-7}
Altered consciousness	21/275 (7.6)	59/265 (22.3)	1.72×10^{-6}
Medical complications	18/265 (6.8)	82/254 (32.3)	1.83×10^{-13}
Neurologic complications	14/274 (5.1)	39/267 (14.6)	2.03×10^{-4}

SD – standard deviation, yrs – years.

*Mann-Whitney test or Pearson's χ^2 test.

Table 2 – Genotype frequency distribution and association with stroke outcome at three months for *MMP2* SNPs.

SNP	Genotype	Whole sample*				Ischemic subset†			
		Genotype frequency		OR [95% CI]	*P*	Genotype frequency		OR [95% CI]	*P*
		Good recovery, n (%)	Poor recovery, n (%)			Good recovery, n (%)	Poor recovery, n (%)		
rs243866									
	G/G	142 (67.6)	117 (57.9)	1.67 [1.10-2.52]	**0.014**	125 (67.6)	83 (56.8)	1.78 [1.13-2.80]	**0.013**
	A/G	66 (31.4)	76 (37.6)			59 (31.9)	55 (37.7)		
	A/A	2 (1.0)	9 (4.5)			1 (0.5)	8 (5.5)		
rs243865									
	C/C	141 (67.8)	117 (57.9)	1.65 [1.09-2.50]	**0.016**	124 (67.8)	83 (56.8)	1.76 [1.12-2.78]	**0.014**
	C/T	65 (31.2)	76 (37.6)			58 (31.7)	55 (37.7)		
	T/T	2 (1.0)	9 (4.5)			1 (0.5)	8 (5.5)		
rs857403									
	A/A	124 (59.3)	138 (68.3)	0.71 [0.48-1.06]	0.091	105 (57.1)	103 (70.5)	0.62 [0.40-0.97]	**0.035**
	T/A	75 (35.9)	56 (27.7)			70 (38)	37 (25.3)		
	T/T	10 (4.8)	8 (4.0)			9 (4.9)	6 (4.1)		
rs1477017									
	A/A	100 (47.6)	81 (40.1)	1.42 [1.01-2.00]	**0.042**	86 (46.5)	55 (37.7)	1.51 [1.04-2.20]	**0.031**
	G/A	91 (43.3)	98 (48.5)			82 (44.3)	72 (49.3)		
	G/G	19 (9.0)	23 (11.4)			17 (9.2)	19 (13)		
rs17301608									
	C/C	94 (45.0)	75 (37.3)	1.40 [1.00-1.95]	0.051	80 (43.5)	50 (34.5)	1.47 [1.01-2.12]	**0.041**
	C/T	94 (45.0)	100 (49.8)			85 (46.2)	73 (50.3)		
	T/T	21 (10.0)	26 (12.9)			19 (10.3)	22 (15.2)		
rs1053605									
	C/C	188 (89.5)	170 (84.2)	2.02 [1.09-3.75]	**0.023**	166 (89.7)	127 (87.0)	1.82 [0.93-3.58]	0.082
	C/T	22 (10.5)	28 (13.9)			19 (10.3)	16 (11.0)		
	T/T	0 (0.0)	4 (2.0)			0 (0.0)	3 (2.0)		
rs2241145									
	G/G	79 (37.8)	56 (27.9)	1.66 [1.20-2.30]	**0.002‡**	68 (37.0)	39 (26.9)	1.67 [1.17-2.40]	**0.004**
	G/C	100 (47.8)	101 (50.2)			88 (47.8)	72 (49.7)		
	C/C	30 (14.4)	44 (21.9)			28 (15.2)	34 (23.4)		
rs243849									
	C/C	131 (62.7)	143 (71.5)	0.70 [0.46-1.07]	0.095	112 (60.9)	108 (75)	0.59 [0.36-0.96]	**0.031**
	T/C	70 (33.5)	52 (26.0)			65 (35.3)	33 (22.9)		
	T/T	8 (3.8)	5 (2.5)			7 (3.8)	3 (2.1)		
rs183112									
	G/G	134 (64.1)	145 (72.9)	0.66 [0.43-1.03]	0.067	115 (62.5)	110 (76.9)	0.54 [0.32-0.90]	**0.016**
	A/G	70 (33.5)	51 (25.6)			65 (35.3)	32 (22.4)		
	A/A	5 (2.4)	3 (1.5)			4 (2.2)	1 (0.7)		
rs1992116									
	G/G	87 (41.6)	65 (32.3)	1.67 [1.20-2.31]	**0.002‡**	76 (41.3)	48 (33.1)	1.68 [1.17-2.42]	**0.004**
	A/G	97 (46.4)	94 (46.8)			86 (46.7)	67 (46.2)		
	A/A	25 (12.0)	42 (20.9)			22 (12.0)	30 (20.7)		

95%CI – 95% Confidence Interval.

Results were adjusted for significant covariates; Odds Ratio (OR)>1 indicates increased probability of poor recovery for the carriers of the minor allele; only associated SNPs are shown.

*OR [95%CI] and *P* for the log-additive genetic model after adjustment for significant covariates (history of hypertension, type of stroke, and occurrence of aphasia, paresis, altered consciousness and complications during hospitalization).

†OR [95%CI] and *P* for the log-additive genetic model after adjustment for significant covariates (history of hypertension, and occurrence of aphasia, paresis, altered consciousness and complications during hospitalization).

‡Significant result after Bonferroni correction.

Table 3 – Haplotype frequency distribution of the *MMP2* and *MMP9* genes, and association with stroke outcome.

Gene	Haplotypes		Whole sample					Ischemic subset				
			Haplotype frequency	Good recovery (%)	Poor recovery (%)	χ^2	*P*	Haplotype frequency	Good recovery (%)	Poor recovery (%)	χ^2	*P*
MMP2	rs11643630-rs243866-rs243865	TAT	0.200	17.2	22.7	5.150	**0.023**	0.198	16.8	23.5	6.125	**0.013**
MMP2	rs1477017-rs17301608-rs1132896-rs1053605-rs2241145-rs243849-rs243842-rs183112	ACGCGTTG	0.011	1.8	0.4	4.776	**0.029**	0.014	2.1	0.6	3.372	0.066
MMP9	rs8113877-rs3918253-rs2236416	TCA	0.034	4.8	1.9	7.403	**0.007**	0.038	4.9	2.4	3.680	0.055

Only haplotypes with significant association results are presented.

Supplementary tables

Supplementary table 1 - Association analysis results for *MMP2* and *MMP9* SNPs and stroke outcome.

gene	marker	position*	location*	MA	Whole sample							Ischemic Subset						
					HWE *P*	**missing genotypes (%)**	**MAF good recovery**	**MAF poor recovery**	***P*** †	***P*** ‡	***P*** §	**HWE *P***	**missing genotypes (%)**	**MAF good recovery**	**MAF poor recovery**	***P*** †	***P*** ‡	***P*** §
MMP2	rs1116195	55508873	upstream	A	0.177	0.01	0.379	0.403	0.927	NS	NS	0.610	0.01	0.370	0.401	0.844	NS	NS
	rs11643630	55510459	upstream	T	0.207	0.02	0.377	0.423	0.050	NS	NS	0.157	0.01	0.382	0.417	0.120	NS	NS
	rs243866	55511537	upstream	A	0.082	0.00	0.174	0.229	**0.014**	NS	NS	0.133	0.00	0.170	0.237	**0.013**	NS	NS
	rs243865	55511806	upstream	T	0.082	0.00	0.173	0.230	**0.016**	NS	NS	0.133	0.01	0.169	0.238	**0.014**	NS	NS
	rs857403	55516708	intron 1-2	T	0.614	0.00	0.242	0.193	0.091	NS	NS	0.671	0.01	0.249	0.180	**0.035**	NS	NS
	rs1477017	55517162	intron 2-3	G	0.435	0.00	0.304	0.344	**0.042**	NS	NS	0.664	0.00	0.313	0.355	**0.031**	NS	NS
	rs17301608	55518610	intron 3-4	T	0.567	0.00	0.324	0.361	0.051	NS	NS	0.833	0.01	0.335	0.375	**0.041**	NS	NS
	rs1132896	55519535	exon 5	C	0.684	0.01	0.288	0.314	0.071	NS	NS	1.000	0.01	0.294	0.323	0.055	NS	NS
	rs1053605	55519607	exon 5	T	0.216	0.00	0.063	0.087	**0.023**	NS	NS	0.255	0.00	0.065	0.078	0.082	NS	NS
	rs2241145	55522200	intron 5-6	C	0.598	0.01	0.398	0.450	**0.002**	**0.044**	**0.021**	0.844	0.01	0.409	0.455	**0.004**	NS	**0.044**
	rs243849	55523705	exon 7	T	0.892	0.01	0.218	0.177	0.095	NS	NS	1.000	0.01	0.222	0.157	**0.031**	NS	NS
	rs243842	55527422	intron 9-10	C	0.855	0.00	0.387	0.365	0.058	NS	NS	1.000	0.01	0.373	0.389	0.180	NS	NS
	rs183112	55527682	intron 9-10	A	0.313	0.01	0.200	0.165	0.067	NS	NS	0.128	0.01	0.200	0.139	**0.016**	NS	NS
	rs1992116	55527891	intron 9-10	A	0.857	0.00	0.358	0.418	**0.002**	**0.038**	**0.018**	0.476	0.01	0.363	0.411	**0.004**	NS	**0.042**
	rs11639960	55533270	intron 11-12	G	0.919	0.01	0.296	0.311	0.389	NS	NS	0.645	0.01	0.300	0.304	0.355	NS	NS
	rs243835	55536622	intron 11-12	T	0.103	0.00	0.498	0.487	0.124	NS	NS	0.847	0.00	0.492	0.479	0.179	NS	NS
	rs243834	55536687	intron 11-12 (splice site)	G	0.122	0.01	0.500	0.485	0.095	NS	NS	1.000	0.01	0.494	0.474	0.134	NS	NS
	rs11541998	55536763	exon 12	G	0.339	0.02	0.089	0.109	0.282	NS	NS	0.241	0.02	0.084	0.105	0.147	NS	NS
	rs243832	55539191	intron 12-13	G	0.297	0.03	0.494	0.494	0.099	NS	NS	0.696	0.03	0.500	0.489	0.142	NS	NS
	rs1861320	55541040	downstream	T	0.928	0.00	0.387	0.398	0.414	NS	NS	0.545	0.01	0.390	0.401	0.467	NS	NS
	rs17302903	55543073	downstream	C	0.928	0.02	0.388	0.413	0.291	NS	NS	0.615	0.02	0.398	0.415	0.388	NS	NS
MMP9	rs8113877	44635045	upstream	G	0.784	0.02	0.359	0.401	0.261	NS	NS	0.599	0.02	0.356	0.371	0.376	NS	NS
	rs3918253	44639511	intron 3-4	C	0.860	0.00	0.411	0.420	0.792	NS	NS	0.764	0.00	0.405	0.399	0.966	NS	NS
	rs2236416	44640575	intron 6-7	G	0.237	0.02	0.090	0.114	0.326	NS	NS	0.404	0.02	0.081	0.111	0.172	NS	NS
	rs17577	44643111	exon 12	A	0.181	0.00	0.098	0.115	0.515	NS	NS	0.284	0.00	0.088	0.111	0.318	NS	NS

HWE – Hardy-Weinberg equilibrium, MA – minor allele, MAF – minor allele frequency, NS – non-significant.

*According to Ensembl Release 56 – September 2009.

†*P* for the log-additive genetic model after adjustment for significant covariates.

‡*P* after Bonferroni correction.

§*P* after SNPSpD approach.

Supplementary table 2 - Association analysis results for *MMP2* and *MMP9* haplotypes and stroke outcome.

Gene	Haplotypes		Whole sample					Ischemic subset				
			Haplotype frequency	**Good recovery (%)**	**Poor recovery (%)**	χ^2	***P***	**Haplotype frequency**	**Good recovery (%)**	**Poor recovery (%)**	χ^2	***P***
MMP2	haplotype block 1:	GGC	0.599	62.2	57.6	2.396	0.122	0.599	61.5	57.9	1.148	0.284
	rs11643630-rs243866-	TGC	0.200	20.6	19.5	0.204	0.652	0.200	21.5	18.2	1.381	0.240
	rs243865	TAT	0.200	17.2	22.7	5.150	**0.023**	0.198	16.8	23.5	6.125	**0.013**
MMP2	haplotype block 2:	ACGCGCCG	0.366	37.7	35.4	0.624	0.430	0.366	36.3	37.0	0.046	0.830
	rs1477017-rs17301608-	GTCCCCTG	0.293	28.1	30.6	0.829	0.362	0.297	28.4	31.3	0.907	0.341
	rs1132896-rs1053605-	ACGCGTTA	0.187	20.1	17.3	1.363	0.243	0.177	19.9	14.9	3.679	0.055
	rs2241145-rs243849-	ACGTCCTG	0.074	6.2	8.6	2.241	0.134	0.070	6.3	7.8	0.714	0.398
	rs243842-rs183112	ATGCCCTG	0.021	2.4	1.9	0.368	0.544	0.028	2.9	2.6	0.097	0.756
		GTGCCCTG	0.017	1.3	2.2	1.444	0.230	0.016	1.5	1.8	0.156	0.693
		ACGCGCTG	0.013	0.8	1.9	2.613	0.106	-	-	-	-	-
		ACGCGTTG	0.011	1.8	0.4	4.776	**0.029**	0.014	2.1	0.6	3.372	0.066
		GTGCCCCG	-	-	-	-	-	0.011	0.6	1.6	1.752	0.186
MMP2	haplotype block 3:	GATGC	0.471	49.2	45.1	1.815	0.178	0.464	48.4	44.0	1.646	0.200
	rs1992116-rs11639960-	AGCAC	0.191	19.6	18.5	0.181	0.671	0.196	20.4	18.6	0.469	0.493
	rs243835-rs243834-	GACAC	0.132	14.0	12.3	0.763	0.383	0.143	14.4	14.2	0.008	0.931
	rs11541998	AGCAG	0.098	8.9	10.6	0.939	0.332	0.092	8.4	10.1	0.767	0.381
		AACAC	0.080	6.9	9.2	2.030	0.154	0.078	6.9	8.9	1.097	0.295
		AGTGC	0.011	0.5	1.7	3.215	0.073	0.012	0.6	1.9	2.885	0.089
MMP2	haplotype block 4:	GG	0.503	50.5	50.0	0.034	0.854	0.496	50.0	49.2	0.048	0.826
	rs243832-rs1861320	CT	0.392	38.5	39.9	0.215	0.643	0.395	38.8	40.3	0.196	0.658
		CG	0.104	10.7	10.1	0.118	0.731	0.108	11.0	10.5	0.060	0.806
MMP9	haplotype block 1:	TTA	0.582	58.5	57.8	0.056	0.813	0.595	59.0	60.2	0.111	0.739
	rs8113877-rs3918253-	GCA	0.280	27.4	28.7	0.232	0.630	0.271	27.7	26.3	0.196	0.658
	rs2236416	GCG	0.100	8.7	11.2	1.889	0.169	0.093	8.1	10.8	1.893	0.169
		TCA	0.034	4.8	1.9	7.403	**0.007**	0.038	4.9	2.4	3.680	0.055

Chapter 3. Inflammatory genes

Manso H, Krug T, Sobral J, Albergaria I, Gaspar G, Ferro JM, Oliveira SA, Vicente AM (2011) Variants in the inflammatory *IL6* and *MPO* genes modulate stroke susceptibility through main effects and gene-gene interactions. J Cereb Blood Flow Metab [Epub ahead of print].

Abstract

There is substantial evidence that inflammation within the CNS contributes to stroke risk and recovery. Inflammatory conditions increase stroke risk, and the inflammatory response is of major importance in recovery and healing processes after stroke. We investigated the role of inflammatory genes *IL1B*, *IL6*, *MPO* and *TNF* in stroke susceptibility and recovery in a population sample of 672 patients and 530 controls, adjusting for demographic, clinical and life-style risk factors and/or stroke severity parameters. We also considered the likely complexity of inflammatory mechanisms in stroke, by assessing the combined effects of multiple genes. Two *IL6* and one *MPO* SNPs were significantly associated with stroke risk ($0.022 < {}_{corrected}P < 0.042$), highlighting gene variants of low to moderate effect in stroke risk. An epistatic interaction between the *IL6* and *MPO* genes was also identified in association with stroke susceptibility (P=0.031 after 1000 permutations). In a subset of 546 patients, one *IL6* haplotype was associated with stroke outcome at three months (${}_{corrected}P$=0.024), an intriguing finding warranting further validation. Our findings support the association of the *IL6* gene and present novel evidence for the involvement of *MPO* in stroke susceptibility, suggesting a modulation of stroke risk by main gene effects, clinical and life-style factors and gene-gene interactions.

Keywords: Cerebrovascular disease, Genetics, Inflammation, Regeneration and recovery, Risk Factors.

Introduction

The brain was once regarded as an "immune privileged" organ, neither susceptible to inflammation nor affected by systemic inflammatory responses. This view has, however, completely changed, and the brain is nowadays known to exhibit key features of inflammation, such as synthesis of cytokines and glial activation, and to intervene in the regulation of systemic inflammation and in acute phase response after brain injury [reviewed in Lucas et al. 2006]. There is also substantial evidence that inflammation within the Central Nervous System (CNS) plays a role in many brain disorders including stroke, a major cause of death and significant disability in Western countries.

Stroke pathophysiology is likely regulated by a combination of environmental/life-style and unclear genetic risk factors. Increasingly, research studies are suggesting that inflammation significantly contributes to stroke risk, progression and outcome [Rodríguez-Yáñez and Castillo 2008; Wang et al. 2007]. For instance, known clinical risk factors for stroke, like atherosclerosis, diabetes, obesity, hypertension, and peripheral infection, are associated with an elevated systemic inflammatory profile [Bastard et al. 2006; Hansson and Libby 2006; Moutsopoulos and Madianos 2006]. Atherosclerosis, in particular, is an inflammatory disease and a major contributor to stroke, either through thromboembolism, which results from the rupture of atherosclerotic plaques, or indirectly through cardioembolism [Hansson and Libby 2006].

Inflammation is equally of major importance in the acute phase of stroke and in the recovery process. It is known that the inflammatory response that follows ischemic or hemorrhagic stroke contributes to exacerbate the initial injury, but that neuroprotective and regenerative molecules are secreted at different stages after a stroke event [Correale and Villa 2004; Lakhan et al. 2009; Lucas et al. 2006]. Clearly, inflammation in CNS injury in general, and in stroke in particular, cannot be classified straightforwardly as harmful. Although there are many inflammatory mediators with detrimental effects, some can be beneficial and others may have dual roles, suggesting a complex orchestration in the acute and recovery phases after stroke [Lucas et al. 2006].

An inflammatory process is thus implicated in pathological conditions that increase stroke risk, in the injury mechanisms upon stroke and in the recovery pathways

that mediate stroke outcome. It is likely that variants of genes encoding inflammatory molecules will influence not only individual stroke risk, but also the extension of the injury and the recovery process, and a number of studies have assessed this hypothesis. For instance, the tumor necrosis factor (*TNF*) gene has been associated with subarachnoid hemorrhage [Yamada et al. 2006], and polymorphisms in the interleukin 1 beta (*IL1B*) and interleukin 6 (*IL6*) genes have been associated with ischemic stroke and with ischemic stroke and intracerebral hemorrhage, respectively [Bis et al. 2008; Yamada et al. 2006]. However, conflicting results have been obtained for the *IL6* gene [Tso et al. 2007]. Other lines of evidence show that inflammatory molecules influence the extension of injury and the recovery process. IL-1 beta and TNF-alpha are known to be released by neurons and endothelial cells in response to ischemia, initiating an inflammatory response and inducing IL-6 and IL-8, with deleterious consequences [Rodríguez-Yáñez and Castillo 2008]. The myeloperoxidase (*MPO)* gene is another intriguing candidate, as the encoded enzyme catalyses the formation of MPO-derived reactive species that may contribute to atherosclerosis progression and destabilization of atherosclerotic plaques [reviewed in Schindhelm et al. 2009]. Reinforcing the hypothesis of a role in stroke, *MPO* polymorphisms have been associated with the size of the brain infarct and functional outcome [Hoy et al. 2003].

In this study we tested the genetic association of major inflammatory players *IL1B* (2q14), *IL6* (7p21), *TNF* (6p21.3) and *MPO* (17q23.1) with stroke susceptibility and stroke outcome at three months. The apparent complexity of the inflammatory mechanisms in stroke, and the multiplicity of players involved suggest a concerted process, in which implicated molecules interact to tightly regulate each other. Still, non-additive interactions or epistasis are generally overlooked in genetic studies. Epistasis is a plausible explanation for the lack of replication across different populations in candidate genes studies or in genome-wide association studies (GWAS), where it is particularly difficult to assess due to the large dimension of the data [Lanktree et al. 2010; Moore 2003]. We therefore examined both independent gene effects and the occurrence of gene-gene interactions among the tested inflammatory genes in stroke risk and stroke recovery.

Materials and Methods

Study population

The study population included 672 first-ever stroke patients, recruited through Neurology and Internal Medicine Departments of several hospitals in Portugal. Stroke definition and the protocol for clinical assessment of patients were previously described [Krug et al. 2010; Manso et al. 2010]. 530 healthy controls with no clinical history of stroke were also enrolled. Since stroke is a late-onset disease, we included older healthy individuals to reduce the probability of misclassification as controls. Information on clinical and life-style risk factors, matching the data available for patients, was obtained by direct interview of control subjects. A subset of 546 patients was included in the outcome analysis. These patients were classified in two groups based on the modified Rankin Scale (mRS) at three months: patients with mRS≤1 were scored as "good recovery" and with mRS>1 were scored as "poor recovery" as previously described [Manso et al. 2010].

The study was approved by the Ethics Committee of the Portuguese Dr. Ricardo Jorge National Institute of Health and other hospitals involved, and participants gave their informed consent.

SNP genotyping

To tag the genetic variation in the *IL1B*, *IL6*, *MPO* and *TNF* gene regions, single nucleotide polymorphisms (SNPs), located within and up to 5kb upstream and downstream of those genes, were selected using the H-clust method [Rinaldo et al. 2005] (HapMap Release 21/phase II July 2006). 3 SNPs in *IL1B*, 6 in *IL6*, 2 in *MPO*, and 3 in *TNF* were genotyped using Sequenom iPLEX assays with allele detection by mass spectroscopy, using Sequenom MassARRAY technology (Sequenom, San Diego, California) and following the manufacturer's protocol. Primer sequences were designed using Sequenom's MassARRAY Assay Design 3.0 software. Quality control analyses were performed based on the genotyping of eight HapMap individuals, duplicated samples within and across genotyping plates, Mendelian segregation in three pedigrees

and no-template samples. For each SNP, call rate <90% and deviation from Hardy-Weinberg equilibrium (HWE; $P<0.05$) were checked.

Statistical analysis

To identify potential confounders, univariate analyses were performed comparing demographic and clinical and life-style risk factors between patients and controls, using the Pearson's χ^2 test and the Mann-Whitney test for discrete and continuous variables, respectively. Variables with a $P<0.25$ in univariate analysis (Table 1) or of particular clinical relevance were included in a logistic regression model using forward selection [Hosmer and Lemeshow 2000] and were maintained in the model if they were associated with stroke susceptibility at a $P\leq0.05$ level. The selected covariates were not correlated ($-0.5<$interaction $i<0.5$). Logistic regression analyses were then used to determine the effect of each genetic variable on stroke susceptibility after adjustment for the significant covariates. Odds ratio (OR) and 95% confidence intervals (95% CI) were computed for the log-additive model. A similar procedure was followed for the analysis of stroke outcome. Demographic and clinical data reflecting the severity of stroke was compared between patients with poor (mRS>1) and good (mRS≤1) outcome at three months to identify potential confounders (Table 2). Logistic regression analyses were then used to determine the effect of each genetic variable on patient's outcome after adjustment for the significant covariates. OR and 95% CI were also computed for the log-additive model.

Univariate and logistic regression analyses were performed using MASS and SNPassoc packages of the R software [R: A language and Environment for Statistical Computing 2004] (v2.6.0). Haplotype blocks in the four genes were determined using the default method [Gabriel et al. 2002] of the Haploview software [Barrett et al. 2005] (v4.0) and haplotype-based association analyses were performed. This algorithm may select different haplotype blocks for the entire sample and the outcome study subset, as occurred specifically for the *IL6* gene, resulting in different haplotypes being tested for association with stroke susceptibility and outcome. Bonferroni correction for multiple testing was used to correct significant associations in individual SNP analysis, as well as haplotype-based association analysis.

Testing for genetic interactions in association with stroke susceptibility and outcome was performed using the multifactor-dimensionality reduction (MDR) method

[Ritchie et al. 2001] (v2.0, beta 7.2), a nonparametric and genetic model-free approach. Briefly, by pooling multilocus genotypes into high and low risk groups, the MDR reduces the dimensionality of the data from *N* dimensions to one dimension. The new multilocus genotype attribute is then tested for its ability to classify and predict disease status, or good/poor outcome at three months. False-positive results due to multiple testing are reduced through combination of the cross-validation strategy and permutation testing [Moore 2003; Ritchie et al. 2001]. Since the MDR method does not accept missing data, we imputed missing genotypes for each SNP using the PLINK software (http://pngu.mgh.harvard.edu/purcell/plink/) [Purcell et al. 2007] and used CEU HapMap genotype data as reference. Best models for each group of genes were chosen from among the best 2, 3 and 4-*loci* models, based on the testing balanced accuracy (TBA) and the cross-validation consistency (CVC) of 10 cross-validation intervals. TBA measures how often individuals are correctly classified in relation to disease or outcome status and CVC measures the number of times the MDR found the same set of loci across the cross validation subsets. A model with a TBA>0.6 is almost always statistically significant, while a TBA>0.55 is considered interesting; TBA=0.5 is random (http://compgen.blogspot.com/2006/12/mdr-101-part-4-results.html). The statistical significance of the best models was calculated after 1000 permutations using the MDR Permutation Tool (v1.0, beta 2). OR-based MDR (OR-MDR) (v1.3-1) was used to determine the OR and 95% CI for each genotype combination as a quantitative measure of disease risk [Chung et al. 2007]. To determine the gain in information about disease or outcome status by combining two variables together over that provided by the independent analysis of these variables, the MDR uses entropy measures [Jakulin and Bratko 2003]. Evidence for a synergistic interaction occurs when the combination of two or more SNPs gives a positive information gain (IG). If the IG is negative, there is evidence for redundancy or correlation between SNPs; and if IG=0, the SNPs have independent effects. Entropy-based interaction dendrograms are used for interpreting epistasis models [Moore et al. 2006].

Results

The demographic and clinical characteristics of our population sample are presented in Table 1. Univariate analysis showed that four stroke risk factors – hypertension, diabetes, smoking and alcohol consumption – were, as expected, significantly more frequent in patients than in control individuals. Gender and age were also significantly different between these two groups. During sample collection, the incidence of stroke was higher in males than females, as expected in this age range; thus, the male/female ratio was higher in the patients' group. Since stroke is a late-onset disease, we selected controls with a higher mean age than patients to reduce the probability of mis-classification as "stroke free". Multivariate analyses were performed after univariate analyses and non-genetic confounders were identified. Adjusting for these covariates was carried out in the final logistic regression model, which also included genetic markers.

All SNPs tested were in HWE in controls and met quality control criteria, and were thus further analyzed. Two contiguous SNPs in the *IL6* gene were associated with stroke susceptibility under a log-additive model (rs2069837: P=0.005, OR[95%CI]=0.66[0.50-0.89]; rs2069861: P=0.007, OR[95%CI]=1.74[1.15-2.63]) (Figure 1, Table 3, supplementary table 1), after adjusting for covariates significant in the multivariate analysis model – gender, hypertension, diabetes and smoking status. These associations with stroke susceptibility remained significant after Bonferroni correction for multiple testing (rs2069837: $_{corrected}P$=0.032; rs2069861: $_{corrected}P$=0.042). A three-marker haplotype containing the two SNPs individually associated with stroke susceptibility and a third SNP (rs10242595) contiguous to rs2069861 conferred an increased risk of stroke (A[rs2069837]–T[rs2069861]–G[rs10242595], P=0.014) (supplementary table 2). This association did not withstand a Bonferroni correction, although this method may be overconservative because these SNPs are not fully independent. Overall, these results highlight a region in the *IL6* gene as a likely susceptibility *locus*, with contiguous *tag* SNPs associated with stroke susceptibility.

One SNP in the *MPO* gene was significantly associated with stroke susceptibility (rs8178406: P=0.011, OR[95%CI]=0.78[0.65-0.95]), and this association survived Bonferroni correction ($_{corrected}P$=0.022) (table 3). Interestingly, restricting the analysis to ischemic patients showed a more significant association ($_{corrected}P$=0.006).

We did not attempt the independent analysis of hemorrhagic patients, as these were too few for adequate statistical power. None of the tested SNPs in *IL1B* and *TNF* were associated with stroke susceptibility in this sample (supplementary table 1).

Because gene interactions may have an important impact on complex phenotypes, including human disease susceptibility, we investigated the existence of genetic interactions that could contribute to stroke risk, using the MDR method. The most significant model for interaction was a two-marker combination between rs10242595 in the *IL6* gene and rs8178406 in the *MPO* gene (Table 4). This model shows a moderately increased TBA of 0.556, thus correctly classifying 55.6% of the individuals tested (*P*=0.031, based on 1000-fold permutations), but a high CVC of 9/10, i.e., the model was selected 9 times out of 10 cross validation subsets. The global OR for this model was 1.69 [95%CI=1.31-2.19]. Two genotype combinations of these SNPs contributed to stroke: AA(rs10242595)-CC(rs8178406) and GA(rs10242595)-TT(rs8178406) (OR[95%CI]=2.80[1.17-7.53] and 1.53[1.22-1.99], respectively) (Table 5). The *IL6* rs10242595 SNP is part of the three-marker haplotype associated with stroke susceptibility (supplementary table 2), while the *MPO* rs8178406 SNP was independently associated with stroke (Table 3). Interpretation of the genetic effects identified by the MDR is possible using the interaction dendrogram shown in Figure 2. Interaction between rs10242595 and rs8178406 shows a positive information gain, indicating a nonlinear, synergistic relationship between the *IL6* (rs10242595) and the *MPO* (rs8178406) genes (i.e. epistasis). Our finding thus suggests that an interaction between two genetic variants in the *IL6* and *MPO* genes contributes to stroke susceptibility, warranting confirmation in independent populations. Additional models were not significant, although there are trends possibly suggesting that interactions between *IL6* (rs10242595) and *TNF* (rs909253), as well as *MPO* (rs2071590) and *TNF* (rs8178406) (permuted *P*=0.054 and *P*=0.060, respectively) may contribute to stroke susceptibility (Table 4).

The impact of *IL1B*, *IL6*, *MPO* and *TNF* genetic variants in patient's outcome at three months was investigated in the subset of 546 patients for whom clinical information during hospitalization and at three months was available. The demographic and clinical characteristics of our population sample are presented in Table 2. Occurrence of aphasia, urinary incontinence, paresis, altered consciousness and medical and neurological complications during hospitalization, which reflect stroke severity, were identified in univariate analysis as significant predictors of poor outcome. One

SNP in the *IL6* gene was associated with stroke outcome at three months (rs1800795: *P*=0.011, OR[95%CI]=1.52[1.10-2.12]) (supplementary table 3) after adjusting for covariates significant in the multivariate model – type of stroke, history of hypertension, and occurrence of aphasia, paresis, altered consciousness and complications during hospitalization, but did not remain significant after Bonferroni correction ($_{corrected}P$=0.066). One two-marker haplotype containing this SNP was associated with an increased probability of good recovery at three months (G[rs1800795]–A[rs2069837], *P*=0.008), and this haplotypic association survived Bonferroni correction ($_{corrected}P$=0.048) (supplementary table 4). We found no evidence for an association of *IL1B*, *MPO* or *TNF* with stroke outcome at three months (supplementary tables 3 and 4). No significant interaction model was found for stroke outcome at three months (supplementary table 5).

Discussion

The objective of the present study was to investigate the role of selected inflammatory genes in stroke susceptibility and recovery. We found evidence for a main effect of the *IL6* and *MPO* genes in stroke risk, with specific polymorphisms significantly associated with stroke susceptibility, after adjustment for confounding demographic, clinical or life-style risk factors. We also report an epistatic gene interaction effect between *IL6* and *MPO* in stroke susceptibility. Our genetic findings thus support previous evidence from other research areas for a role of inflammatory molecules in stroke.

Association analysis of the *IL6* gene showed that two SNPs survived Bonferroni correction, highlighting a region in the *IL6* gene that is likely to harbour risk variants of moderate to low effect size. The associated SNPs are contiguous to the *IL6* SNP (rs1800795) that has been widely tested in multiple population sets, but are not in linkage disequilibrium with this functional polymorphism in our sample. We did not replicate the association with this SNP, suggesting that the present results are signaling a different causative variant in the *IL6* gene, but still reinforce a role of the *IL6* gene in stroke susceptibility. Accordingly, several previous studies have failed to confirm the association of rs1800795 with stroke, while others showed heterogeneity regarding the

associated allele or genotype [Tso et al. 2007]. These conflicting results may be due to allelic or genetic heterogeneity and/or limitations in study designs, or reflect true differences in stroke etiology between populations. Our study also provided novel evidence for the association of the *MPO* gene with stroke. Of the two SNPs tested, covering genetic variability in this region, one was associated with stroke risk. This effect seemed to be largely driven by the ischemic stroke subset where the strength of association was improved, perhaps reflecting somewhat distinct pathological mechanisms for the hemorrhagic and ischemic subtypes. Validation in independent populations is now warranted.

The identification of a synergistic interaction between *IL6* and *MPO* contributing to stroke susceptibility highlights the importance of testing for epistasis and illustrates the complexity of the inflammatory processes in stroke. It indicates that susceptibility may be modulated not only by a variety of genetic factors but also by non-linear gene-gene interactions, as had been previously shown by others [Flex et al. 2004; Liu et al. 2009; Palmer et al. 2010]. Liu and colleagues (2009), in particular, investigated the existence of gene-gene interactions between five candidate genes and stroke and found that individuals with a combination of polymorphisms in three of these genes had an increased risk of thrombotic stroke. Two additional studies also report that the risk of stroke increases with the number of high risk genotypes in pro-inflammatory gene polymorphisms carried by an individual, suggesting that such polymorphisms act synergistically [Flex et al. 2004; Palmer et al. 2010]. Finally, our finding is in agreement with a previous *in vitro* functional study, showing that enzymatically inactive MPO induced IL-6 secretion in a dose and time-dependent manner by endothelial cells [Lefkowitz et al. 2000].

The genetic factors influencing outcome after a stroke event are far less studied than genetic risk factors. The importance of inflammation after stroke onset and the correlation between inflammatory marker levels and infarct volume or patient's outcome [Smith et al. 2004; Sotgiu et al. 2006], led us to investigate the role of several inflammatory genes in stroke outcome at three months. We found an *IL6* two-marker haplotype associated with patient's outcome at three months. These results are intriguing, since IL-6 is one of the cytokines induced after stroke, playing a fundamental role in the inflammatory injury that follows a stroke event, but equally known to have neuroprotective effects in later stages after stroke [Herrmann et al. 2003]. IL-6 levels have been correlated with stroke severity, 12 months mortality, clinical outcome and

brain infarct volume [Smith et al. 2004]. However, our results require validation in independent, larger population samples.

IL-6 is a pleiotropic cytokine, with both pro- and anti-inflammatory functions and a low level of expression in the brain under normal physiologic conditions [Luheshi and Rothwell 1996]. However, increased levels of IL-6 have been detected after a stroke event [Clark et al. 1999]. Increased expression of IL-6 has also been found in atherosclerotic plaques [Schieffer et al. 2000], suggesting that the identified *IL6* association with stroke susceptibility may be mediated by atherosclerosis progression. Likewise, abundant MPO-positive cells are present in sites of atherosclerotic plaque rupture, and this molecule may contribute to stroke through destabilization of the atherosclerotic plaques [Sugiyama et al. 2001]. It would be very interesting to correlate *IL6* and *MPO* genetic variants with the carotid intimal-media wall thickness, which is a marker for atherosclerosis [Mattace Raso et al. 1999]. However, this data is only available for a small percentage of our patients, precluding this analysis for now.

The complex interplay between genetic background, clinical and life-style factors and the environment may ultimately regulate the onset, acute phase and outcome of stroke. In the present study we present supporting evidence for a role of the *IL6* and *MPO* inflammatory genes in stroke susceptibility, and show that stroke risk is modulated by main gene effects together with clinical and life-style factors as well as by gene-gene interactions. Our findings are compatible and strengthen previous genetic and biological observations, highlighting the need of further functional studies, particularly in view of the possible utility of IL-6 as a diagnostic and/or prognostic biomarker for stroke.

Acknowledgements and funding

The authors are grateful to all study participants and their families. The authors wish to thank Dr. Marinho Falcão, Dr. Carlos Dias and the whole team at Instituto Nacional de Saúde Dr. Ricardo Jorge, and all the clinicians that recruited study subjects from the following hospitals: H. S. João, H. do Espírito Santo, Centro Hospitalar do Funchal, H. Conde S. Bento, H. S. José, H. S. Marcos, H. Garcia de Orta, H. de Faro, Centro Hospitalar de Coimbra, Centro Hospitalar de Vila Nova de Gaia, H. Infante D.Pedro, Serviços de Assistência Médico-Social do Sindicato dos Bancários do Sul e

Ilhas, H. Sto. António dos Capuchos, H. Sto. António, H. Distrital de Mirandela, H Sta. Maria, H. de Egas Moniz, H. Prof. Doutor Fernando Fonseca and H. S. Pedro de Vila Real. The authors also wish to thank the technical assistance provided by the Genotyping Unit at Instituto Gulbenkian de Ciência. This work was supported by the Marie Curie International Reintegration Grant 513760, the Marie Curie Intra-European Fellowship 024563, the FCT grant PTDC/SAU-GMG/64426/2006, and fellowships from FCT and the Portuguese Instituto do Emprego e Formação Profissional.

References

Barrett JC, Fry B, Maller J, Daly MJ. (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21:263-265

Bastard JP, Maachi M, Lagathu C, Kim MJ, Caron M, Vidal H, Capeau J, Feve B. (2006) Recent advances in the relationship between obesity, inflammation, and insulin resistance. Eur Cytokine Netw 17:4-12

Bis JC, Heckbert SR, Smith NL, Reiner AP, Rice K, Lumley T, Hindorff LA, Marciante KD, Enquobahrie DA, Monks SA, Psaty BM. (2008) Variation in inflammation-related genes and risk of incident nonfatal myocardial infarction or ischemic stroke. Atherosclerosis 198:166-173

Chung Y, Lee SY, Elston RC, Park T. (2007) Odds ratio based multifactor-dimensionality reduction method for detecting gene-gene interactions. Bioinformatics 23:71-76

Clark WM, Rinker LG, Lessov NS, Hazel K, Eckenstein F. (1999) Time course of IL-6 expression in experimental CNS ischemia. Neurol Res 21:287-292

Correale J, Villa A. (2004) The neuroprotective role of inflammation in nervous system injuries. J Neurol 251:1304-1316

Flex A, Gaetani E, Papaleo P, Straface G, Proia AS, Pecorini G, Tondi P, Pola P, Pola R. (2004) Proinflammatory genetic profiles in subjects with history of ischemic stroke. Stroke 35:2270-2275

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D. (2002) The structure of haplotype blocks in the human genome. Science 296:2225-2229

Hansson GK, Libby P. (2006) The immune response in atherosclerosis: a double-edged sword. Nat Rev Immunol 6:508-519

Herrmann O, Tarabin V, Suzuki S, Attigah N, Coserea I, Schneider A, Vogel J, Prinz S, Schwab S, Monyer H, Brombacher F, Schwaninger M. (2003) Regulation of body temperature and neuroprotection by endogenous interleukin-6 in cerebral ischemia. J Cereb Blood Flow Metab 23:406-415

Hosmer DW, Lemeshow S. (2000) Applied Logistic Regression. New York: John Wiley & Sons, Inc.

Hoy A, Leininger-Muller B, Poirier O, Siest G, Gautier M, Elbaz A, Amarenco P, Visvikis S. (2003) Myeloperoxidase polymorphisms in brain infarction. Association with infarct size and functional outcome. Atherosclerosis 167:223-230

Jakulin A, Bratko I. (2003) Analyzing attribute interactions. Lect. Notes Artif. Intell. 2838:229-240

Krug T, Manso H, Gouveia L, Sobral J, Xavier JM, Albergaria I, Gaspar G, Correia M, Viana-Baptista M, Simões RM, Pinto AN, Taipa R, Ferreira C, Fontes JR, Silva MR, Gabriel JP, Matos I, Lopes G, Ferro JM, Vicente AM, Oliveira SA. (2010) Kalirin: a novel genetic risk factor for ischemic stroke. Hum Genet 127:513-523

Lakhan SE, Kirchgessner A, Hofer M. (2009) Inflammatory mechanisms in ischemic stroke: therapeutic approaches. J Transl Med 7:97

Lanktree MB, Dichgans M, Hegele RA. (2010) Advances in genomic analysis of stroke: what have we learned and where are we headed? Stroke 41:825-832

Lefkowitz DL, Roberts E, Grattendick K, Schwab C, Stuart R, Lincoln J, Allen RC, Moguilevsky N, Bollen A, Lefkowitz SS. (2000) The endothelium and cytokine secretion: the role of peroxidases as immunoregulators. Cell Immunol 202:23-30

Liu J, Sun K, Bai Y, Zhang W, Wang X, Wang Y, Wang H, Chen J, Song X, Xin Y, Liu Z, Hui R. (2009) Association of three-gene interaction among MTHFR, ALOX5AP and NOTCH3 with thrombotic stroke: a multicenter case-control study. Hum Genet 125:649-656

Lucas SM, Rothwell NJ, Gibson RM. (2006) The role of inflammation in CNS injury and disease. Br J Pharmacol 147 Suppl 1:S232-240

Luheshi G, Rothwell N. (1996) Cytokines and fever. Int Arch Allergy Immunol 109:301-307

Manso H, Krug T, Sobral J, Albergaria I, Gaspar G, Ferro JM, Oliveira SA, Vicente AM. (2010) Variants of the Matrix Metalloproteinase-2 but not the Matrix Metalloproteinase-9 genes significantly influence functional outcome after stroke. BMC Med Genet 11:40

Mattace Raso F, Rosato M, Talerico A, Cotronei P, Mattace R. (1999) Intimal-medial thickness of the common carotid arteries and lower limbs atherosclerosis in the elderly. Minerva Cardioangiol 47:321-327

Moore JH. (2003) The ubiquitous nature of epistasis in determining susceptibility to common human diseases. Hum Hered 56:73-82

Moore JH, Gilbert JC, Tsai CT, Chiang FT, Holden T, Barney N, White BC. (2006) A flexible computational framework for detecting, characterizing, and interpreting statistical patterns of epistasis in genetic studies of human disease susceptibility. J Theor Biol 241:252-261

Moutsopoulos NM, Madianos PN. (2006) Low-grade inflammation in chronic infectious diseases: paradigm of periodontal infections. Ann N Y Acad Sci 1088:251-264

Palmer CN, Kimber CH, Doney AS, Proia AS, Morris AD, Gaetani E, Quarta M, Smith RC, Pola R. (2010) Combined effect of inflammatory gene polymorphisms and the risk of ischemic stroke in a prospective cohort of subjects with type 2 diabetes: a Go-DARTS study. Diabetes 59:2945-2948

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559-575

R: A language and Environment for Statistical Computing. (2004). Vienna

Rinaldo A, Bacanu SA, Devlin B, Sonpar V, Wasserman L, Roeder K. (2005) Characterization of multilocus linkage disequilibrium. Genet Epidemiol 28:193-206

Ritchie MD, Hahn LW, Roodi N, Bailey LR, Dupont WD, Parl FF, Moore JH. (2001) Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. Am J Hum Genet 69:138-147

Rodríguez-Yáñez M, Castillo J. (2008) Role of inflammatory markers in brain ischemia. Curr Opin Neurol 21:353-357

Schieffer B, Schieffer E, Hilfiker-Kleiner D, Hilfiker A, Kovanen PT, Kaartinen M, Nussberger J, Harringer W, Drexler H. (2000) Expression of angiotensin II and interleukin 6 in human coronary atherosclerotic plaques: potential implications for inflammation and plaque instability. Circulation 101:1372-1378

Schindhelm RK, van der Zwan LP, Teerlink T, Scheffer PG. (2009) Myeloperoxidase: a useful biomarker for cardiovascular disease risk stratification? Clin Chem 55:1462-1470

Smith CJ, Emsley HC, Gavin CM, Georgiou RF, Vail A, Barberan EM, del Zoppo GJ, Hallenbeck JM, Rothwell NJ, Hopkins SJ, Tyrrell PJ. (2004) Peak plasma interleukin-6 and other peripheral markers of inflammation in the first week of ischaemic stroke correlate with brain infarct volume, stroke severity and long-term outcome. BMC Neurol 4:2

Sotgiu S, Zanda B, Marchetti B, Fois ML, Arru G, Pes GM, Salaris FS, Arru A, Pirisi A, Rosati G. (2006) Inflammatory biomarkers in blood of patients with acute brain ischemia. Eur J Neurol 13:505-513

Sugiyama S, Okada Y, Sukhova GK, Virmani R, Heinecke JW, Libby P. (2001) Macrophage myeloperoxidase regulation by granulocyte macrophage colony-stimulating factor in human atherosclerosis and implications in acute coronary syndromes. Am J Pathol 158:879-891

Tso AR, Merino JG, Warach S. (2007) Interleukin-6 174G/C polymorphism and ischemic stroke: a systematic review. Stroke 38:3070-3075

Wang Q, Tang XN, Yenari MA. (2007) The inflammatory response in stroke. J Neuroimmunol 184:53-68

Yamada Y, Metoki N, Yoshida H, Satoh K, Ichihara S, Kato K, Kameyama T, Yokoi K, Matsuo H, Segawa T, Watanabe S, Nozawa Y. (2006) Genetic risk for ischemic and hemorrhagic stroke. Arterioscler Thromb Vasc Biol 26:1920-1925

Figures

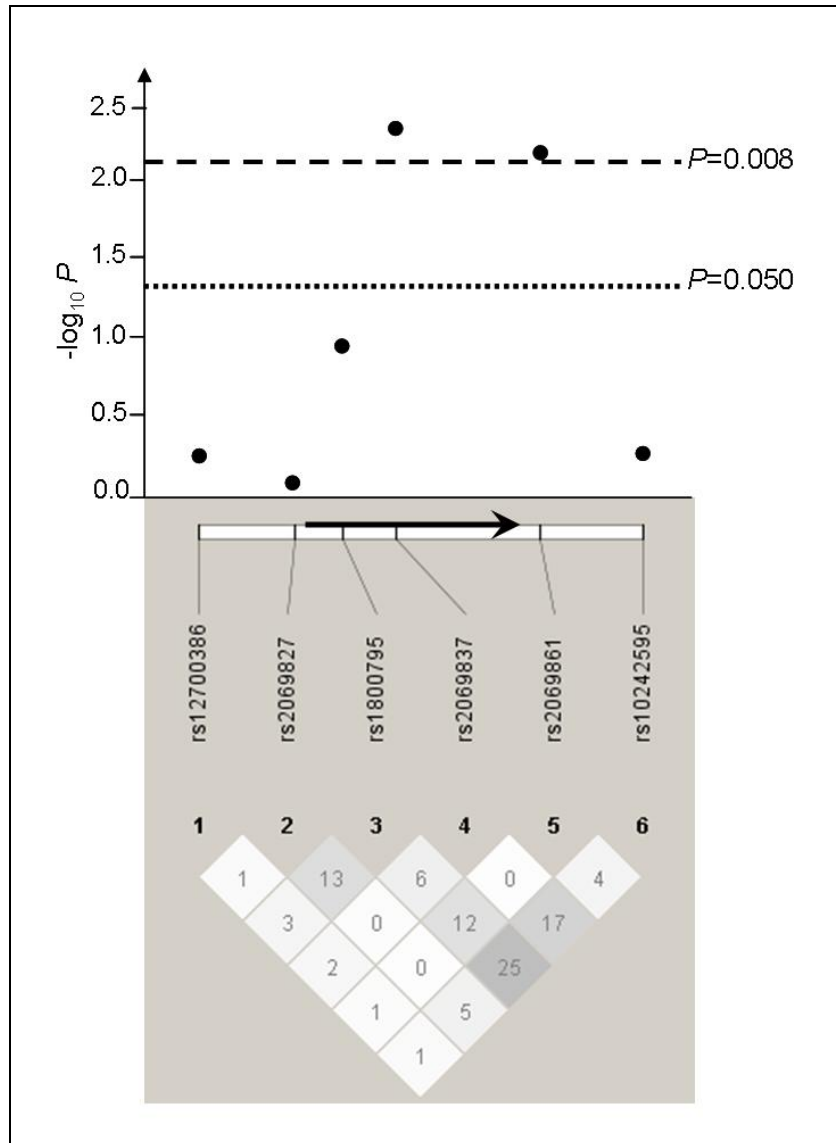

Figure 1 - *IL6* association results ($-\log_{10} P$) with stroke susceptibility and pairwise r^2 among genotyped SNPs in our population sample. The positions of the six SNPs relative to the *IL6* gene (represented by an arrow) are indicated. The magnitude of linkage disequilibrium (r^2) is represented by the white-black gradient shading and the values within each diamond. Association results above the line $-\log_{10}P=1.3$ are considered significant ($P<0.050$); those above $-\log_{10}P=2.1$ survive Bonferroni correction ($P<0.008$). Linkage disequilibrium blocks were generated using the Gabriel et al [2002] method.

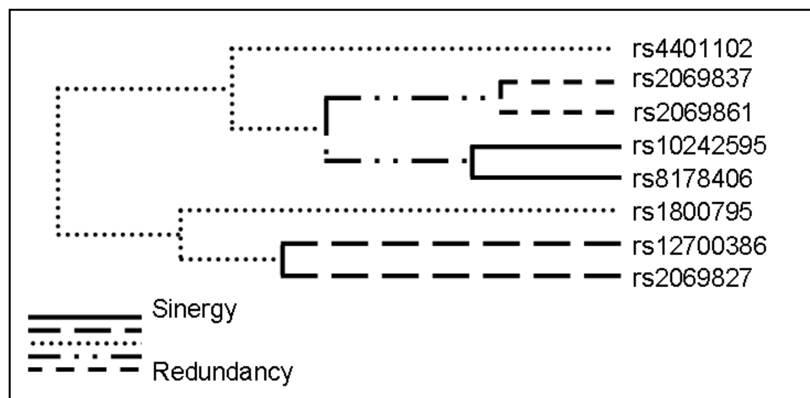

Figure 2 - Interaction dendrogram for the *IL6* and *MPO* polymorphisms in stroke susceptibility. The length of the dendrogram branch that connects two polymorphisms indicates the strength of interaction (the shorter the branch, the stronger is the interaction).

Tables

Table 1 – Demographic and clinical characteristics of the population sample.

Characteristic	Controls	Patients	P^{*}
Age, mean±SD (yrs)	62.9±6.8	52.2±9.1	$<10^{-4}$
Gender (male), n/N (%)	247/530 (46.6)	428/672 (63.7)	$<10^{-4}$
Stroke Type, n/N (%)			
Ischemic stroke	–	551/672 (82.0)	–
Hemorrhagic stroke	–	111/672 (16.5)	–
Unknow type of stroke	–	10/672 (1.5)	–
Stroke Risk Factors, n/N (%)			
Hypertension (>85-140 mmHg)	193/513 (37.6)	369/601 (61.4)	$<10^{-4}$
Diabetes	59/501 (11.8)	102/628 (16.2)	0.033
Hypercholesterolemia (cholesterol >200 mg/dL)	328/520 (63.1)	385/623 (61.8)	0.657
Smoking	147/512 (28.7)	308/660 (46.7)	$<10^{-4}$
Drinking	218/505 (43.2)	388/662 (58.6)	$<10^{-4}$

SD – standard deviation, yrs – years.

[*] Mann-Whitney test or Pearson's χ^2 test.

Table 2 – Demographic and clinical characteristics of stroke patients analyzed for outcome at three months.

Characteristic	Good Recovery (mRS≤1)	Poor Recovery (mRS>1)	*P* *
Age and Gender			
Age, mean±SD (yrs)	50.8±9	52.5±8.5	0.028
Gender (male), n/N (%)	174/276 (63.0)	174/270 (64.4)	0.734
Past History, n/N (%)			
Hypertension	159/241 (66.0)	143/240 (59.6)	0.147
Diabetes	36/259 (13.9)	47/246 (19.1)	0.115
Cardiac Disease	37/264 (14.0)	43/257 (16.7)	0.390
Stroke type, n/N (%)			$<10^{-4}$
Ischemic stroke	238/276 (86.2)	193/270 (71.5)	–
Hemorrhagic stroke	33/276 (12.0)	72/270 (26.7)	–
Unknow type of stroke	5/276 (1.8)	5/270 (1.9)	–
Stroke Features, n/N (%)			
Aphasia	53/258 (20.5)	98/250 (39.2)	$<10^{-4}$
Neglect	11/266 (4.1)	19/240 (7.9)	0.072
Dysphagia	15/270 (5.6)	25/251 (10.0)	0.059
Urinary Incontinence	5/272 (1.8)	15/251 (6.0)	0.014
Paresis	203/273 (74.4)	244/269 (90.7)	$<10^{-4}$
Altered consciousness	21/275 (7.6)	59/265 (22.3)	$<10^{-4}$
Medical complications	18/265 (6.8)	82/254 (32.3)	$<10^{-4}$
Neurologic complications	14/274 (5.1)	39/267 (14.6)	$2.03\text{x}10^{-4}$

SD – standard deviation, yrs – years.

*Mann-Whitney test or Pearson's χ^2 test.

Table 3 – Genotype and allele frequency distribution, and association with stroke susceptibility for the *IL6* and *MPO* SNPs.

Gene	SNP	Genotype	Genotype frequency		OR [95% CI]	*P*
			Controls, n (%)	Cases, n (%)		
IL6	**rs2069837**					
		A/A	365 (76.4)	461 (81.2)	0.66 [0.50-0.89]	**0.005**†
		G/A	105 (22.0)	102 (18.0)		
		G/G	8 (1.7)	5 (0.9)		
IL6	**rs2069861**					
		C/C	442 (91.9)	497 (86.9)	1.74 [1.15-2.63]	**0.007**†
		T/C	39 (8.1)	70 (12.2)		
		T/T	0 (0.0)	5 (0.9)		
MPO	**rs8178406**					
		T/T	151 (31.4)	221 (38.9)	0.78 [0.65-0.95]	**0.011**†
		T/C	254 (52.8)	262 (46.1)		
		C/C	76 (15.8)	85 (15.0)		

Only associated SNPs are shown. Odds Ratio (OR) >1 indicates increased probability of having a stroke for the carriers of the minor allele. CI – 95% Confidence Interval.

*OR [95% CI] and *P* for the log-additive genetic model after adjustment for significant covariates (gender, history of hypertension, diabetes, smoking status)

†Significant result after Bonferroni correction.

Table 4 – Gene x gene interaction models obtained using the multifactor-dimensionality reduction (MDR) method in stroke susceptibility.

Genes	Best model				CVC	TBA	*P**
	SNP1	SNP2	SNP3	SNP4			
IL1B _TNF	rs1143643(*IL1B*)	rs16944(*IL1B*)	rs2071590(*TNF*)	–	8/10	0.517	0.487
IL6 _TNF	rs10242595(*IL6*)	rs909253(*TNF*)	–	–	10/10	0.549	0.054
IL6 _IL1B	rs2069837(*IL6*)	rs10242595(*IL6*)	rs1143643(*IL1B*)	–	5/10	0.541	0.145
MPO _TNF	rs2071590(*TNF*)	rs8178406(*MPO*)	–	–	9/10	0.547	0.060
IL6 _MPO	rs10242595(*IL6*)	rs8178406(*MPO*)	–	–	9/10	0.556	**0.031**
MPO _IL1B	rs1143643(*IL1B*)	rs16944(*IL1B*)	rs8178406(*MPO*)	rs4401102(*MPO*)	9/10	0.538	0.160
IL1B _IL6 _TNF_MPO	rs10242595(*IL6*)	rs1143643(*IL1B*)	rs8178406(*MPO*)	rs4401102(*MPO*)	8/10	0.527	0.323

CVC– Cross Validation Consistency, TBA – Testing Balanced Accuracy.

*1000 permutations *P*

Table 5 – Odds Ratio (OR) of each genotype combination of *IL6* rs10242595 and *MPO* rs8178406 obtained using the OR- based MDR.

rs10242595	rs8178406	frequency (case:control)	OR [95% CI]
AA	CC	17:5	**2.80 [1.17-7.53]**
GA	CC	42:44	0.79 [0.59-1.18]
GG	CC	28:32	0.72 [0.51-1.18]
AA	TC	48:54	0.73 [0.56-1.06]
GA	TC	136:124	0.90 [0.76-1.12]
GG	TC	108:94	0.95 [0.78-1.21]
AA	TT	32:33	0.80 [0.57-1.28]
GA	TT	132:71	**1.53 [1.22-1.99]**
GG	TT	77:53	1.20 [0.92-1.66]

CI – Confidence Interval.

Supplementary tables

Supplementary table 1 – Association analysis results for *IL1B*, *IL6*, *MPO* and *TNF* markers and stroke susceptibility

gene	marker	position*	location*	MA	HWE *P*	missing genotypes (%)	MAF controls	MAF patients	*P* †	*P* ‡
IL1B	rs1143643	113588302	intronic/downstream	T	0.924	0.3	0.348	0.327	0.195	NS
	rs1143634	113590390	exonic/downstream	A	0.627	0.2	0.234	0.246	0.862	NS
	rs16944	113594867	upstream	A	0.771	5.4	0.357	0.354	0.591	NS
IL6	rs12700386	22763009	upstream	G	0.130	0.2	0.196	0.179	0.538	NS
	rs2069827	22765456	upstream	T	0.503	0.4	0.071	0.070	0.799	NS
	rs1800795	22766645	intronic	C	0.611	4.2	0.315	0.345	0.115	NS
	rs2069837	22768027	intronic	G	1.000	0.9	0.123	0.100	**0.005**	**0.032**
	rs2069861	22771654	downstream	T	0.615	0.3	0.044	0.068	**0.007**	**0.042**
	rs10242595	22774231	downstream	A	0.530	0.6	0.413	0.411	0.525	NS
MPO	rs8178406	56344443	downstream	C	0.061	0.9	0.422	0.377	**0.011**	**0.022**
	rs4401102	56360837	upstream	T	0.916	0.2	0.291	0.287	0.892	NS
TNF	rs2071590	31539768	upstream	A	0.062	0.3	0.331	0.302	0.070	NS
	rs909253	31540313	upstream	G	0.526	0.2	0.290	0.319	0.159	NS
	rs3093662	31544189	intronic/downstream	G	0.820	0.3	0.105	0.092	0.587	NS

HWE – Hardy-Weinberg equilibrium, MA – minor allele, MAF – minor allele frequency, NS – non-significant.

*According to Ensembl Release 56 – September 2009.

†*P* for the log-additive genetic model after adjustment for significant covariates.

‡*P* after Bonferroni correction.

Supplementary table 2 – Association analysis results for *IL1B*, *IL6* and *TNF* haplotypes and stroke susceptibility

Gene	Haplotypes		Haplotype frequency	Controls (%)	Cases (%)	χ^2	*P*
IL1B	rs1143643-rs1143634	CG	0.422	41.7	42.7	0.238	0.626
		TG	0.336	34.8	32.7	1.236	0.266
		CA	0.241	23.5	24.6	0.442	0.506
IL6	rs2069837-rs2069861-rs10242595	ACG	0.531	54.3	52.1	1.139	0.286
		ACA	0.303	29.0	31.3	1.457	0.227
		GCA	0.110	12.3	9.9	3.572	0.059
		ATG	0.057	4.4	6.7	6.011	**0.014**
TNF	rs2071590-rs909253-rs3093662	AAA	0.316	33.1	30.4	2.037	0.154
		GGA	0.306	29.0	31.8	2.215	0.137
		GAA	0.281	27.4	28.6	0.425	0.514
		GAG	0.098	10.5	9.2	1.131	0.288

Supplementary table 3 – Association analysis results for *IL1B*, *IL6*, *MPO* and *TNF* markers and stroke outcome.

gene	marker	position*	location*	MA	HWE *P*	missing genotypes (%)	MAF good recovery	MAF poor recovery	*P* †	*P* ‡
IL1B	rs1143643	113588302	intronic/downstream	T	0.147	0.4	0.317	0.345	0.671	NS
	rs1143634	113590390	exonic/downstream	A	0.908	0.0	0.248	0.243	0.887	NS
	rs16944	113594867	upstream	A	0.633	3.5	0.357	0.343	0.796	NS
IL6	rs12700386	22763009	upstream	G	0.886	0.4	0.179	0.185	0.684	NS
	rs2069827	22765456	upstream	T	0.719	0.5	0.060	0.071	0.585	NS
	rs1800795	22766645	intronic	C	0.377	5.7	0.311	0.370	**0.011**	NS
	rs2069837	22768027	intronic	G	0.334	0.9	0.086	0.110	0.908	NS
	rs2069861	22771654	downstream	T	0.076	0.2	0.062	0.070	0.179	NS
	rs10242595	22774231	downstream	A	0.287	1.1	0.405	0.423	0.236	NS
MPO	rs8178406	56344443	downstream	C	0.927	0.9	0.365	0.395	0.192	NS
	rs4401102	56360837	upstream	T	0.834	0.2	0.269	0.304	0.305	NS
TNF	rs2071590	31539768	upstream	A	0.269	0.5	0.325	0.294	0.241	NS
	rs909253	31540313	upstream	G	0.274	0.4	0.305	0.325	0.366	NS
	rs3093662	31544189	intronic/downstream	G	0.605	0.5	0.080	0.099	0.520	NS

HWE – Hardy-Weinberg equilibrium, MA – minor allele, MAF – minor allele frequency, NS – non-significant.

*According to Ensembl Release 56 – September 2009.

†*P* for the log-additive genetic model after adjustment for significant covariates.

‡*P* after Bonferroni correction.

Supplementary table 4 – Association analysis results for *IL1B*, *IL6* and *TNF* haplotypes and stroke outcome.

Gene	Haplotypes		Haplotype frequency	Good recovery (%)	Poor recovery (%)	χ^2	*P*
IL1B	rs1143643-rs1143634	CG	0.424	43.5	41.2	0.561	0.454
		TG	0.331	31.7	34.5	0.967	0.326
		CA	0.245	24.8	24.3	0.046	0.830
IL6	haplotype block 1: rs1800795-rs2069837	GA	0.561	60.1	52.1	7.025	**0.008***
		CA	0.340	31.3	36.9	3.771	0.052
		GG	0.098	8.6	11.0	1.763	0.184
IL6	haplotype block 2: rs2069861-rs10242595	CG	0.520	53.3	50.6	0.776	0.379
		CA	0.414	40.5	42.3	0.375	0.540
		TG	0.066	6.2	7.0	0.309	0.578
TNF	rs2071590-rs909253-rs3093662	GGA	0.315	30.5	32.5	0.462	0.497
		AAA	0.311	32.7	29.5	1.351	0.245
		GAA	0.284	28.7	28.2	0.039	0.844
		GAG	0.089	8.0	9.9	1.190	0.275

*Significant result after Bonferroni correction.

Supplementary table 5 – Gene x gene interaction models obtained using multifactor-dimensionality reduction method in stroke outcome.

Genes	Best model				CVC	TBA	*P**
	SNP1	SNP2	SNP3	SNP4			
IL1B _TNF	rs1143643(*IL1B*)	rs16944(*IL1B*)	rs909253(*TNF*)	–	1/10	0.5146	0.586
IL6 _TNF	rs12700386(*IL6*)	rs10242595(*IL6*)	rs2071590(*TNF*)	rs1800795(*IL6*)	6/10	0.5542	0.189
IL6 _IL1B	rs12700386(*IL6*)	rs10242595(*IL6*)	rs1800795(*IL6*)	–	5/10	0.5652	0.122
MPO _TNF	rs2071590(*TNF*)	rs909253(*TNF*)	rs8178406(*MPO*)	rs4401102(*MPO*)	8/10	0.4569	0.987
IL6 _MPO	rs12700386(*IL6*)	rs1800795(*IL6*)	rs10242595(*IL6*)	–	6/10	0.5672	0.085
MPO _IL1B	rs1143643(*IL1B*)	rs4401102(*MPO*)	–	–	7/10	0.5638	0.112
IL1B _IL6 _TNF _MPO	rs1143643(*IL1B*)	rs4401102(*MPO*)	–	–	7/10	0.5647	0.157

CVC – Cross Validation Consistency, TBA – Testing Balanced Accuracy.

*1000 permutations *P*

Chapter 4. Growth factor genes

Manso H, Krug T, Sobral J, Albergaria I, Gaspar G, Ferro JM, Oliveira SA, Vicente AM (2011) Evidence for genetic interactions among growth factor genes in stroke outcome (Submitted)

Abstract

Patient's outcome after stroke likely results from complex interactions between multiple molecules acting in different pathways. Through effects in angiogenesis, neurogenesis and neuroprotection, growth factors are thought to play a role in improving neurological function in stroke recovery. In this study, we hypothesized that variants of the brain-derived neurotrophic factor (*BDNF*), fibroblast growth factor 2 (*FGF2*) and vascular endothelial growth factor A (*VEGFA*) genes, as well as epistatic interactions among them, could influence functional outcome after stroke. A total of 546 stroke patients were analyzed. One *FGF2* SNP was nominally associated with stroke outcome ($_{uncorrected}P$=0.038). Analysis of epistasis yielded two significant gene-gene interaction models, after 1000-fold permutation testing. One model combined a *BDNF* SNP and three *FGF2* SNPs, with a statistically significant testing balanced accuracy (TBA) of 0.592 ($_{permuted}P$=0.026), cross-validation consistency [CVC] of 6/10, and a global odds ratio (OR) (95% confidence interval [CI]) of 4.15[2.86-6.04]. The second model included one *FGF2* SNP and two *VEGFA* SNPs which had a statistically significant TBA of 0.611 ($_{permuted}P$=0.002), CVC=3/10 and global OR[95%CI]=2.54[1.76-3.67]. The results provide evidence for gene interactions in stroke outcome and highlight the complexity of the recovery mechanisms after a stroke event.

Keywords: Cerebrovascular disease, Genetics, Interactions, Regeneration and recovery

Description

Stroke is a major health problem in developed countries, with a very high incidence and mortality and an elevated morbidity. Multiple biological pathways that can be neuroprotective or detrimental to the brain are activated after stroke, resulting in a series of biochemical, hemodynamic and neurophysiologic changes which ultimately determine stroke outcome [Mitsios et al. 2006]. The timely orchestration of complex interactions between growth factors such as brain-derived neurotrophic factor (BDNF), heparin-binding growth factor 2 (HBGF-2) and vascular endothelial growth factor A (VEGF-A), which mediate neurogenesis, angiogenesis and neuroprotection mechanisms that contribute to functional recovery after stroke [Chen et al. 2005; Krupinski et al. 1993; Nakatomi et al. 2002], is likely to influence neurological outcome (references in Supplementary Table 1).

Hypothesizing that variation within growth factor genes *BDNF*, *FGF2* (encoding the HBGF-2 protein) and *VEGFA* might influence patient's outcome after stroke, we conducted a genetic association study in 546 first-ever stroke patients evaluated at three months for functional recovery. Diagnosis and classification of patients according to the modified Rankin Scale (mRS) at three months were described in a previous study [Manso et al. 2010]. The study was approved by the appropriate Ethical Committees, and subjects gave informed consent. Patients were classified in poor (mRS>1) and good (mRS≤1) recovery groups, which showed similar profiles in terms of age, male:female ratio and stroke risk factors (Supplementary Table 2). As expected, clinical characteristics reflecting the severity of stroke (occurrence of aphasia, urinary incontinence, paresis, altered consciousness, medical and neurological complications during hospitalization) were significantly more frequent in patients with poor outcome at three months. Haplotype tagging single nucleotide polymorphisms (SNPs) in *BDNF* (N=11), *FGF2* (N=28) and *VEGFA* (N=8) genomic regions were genotyped using Sequenom MassARRAY technology (Sequenom, San Diego, USA). One SNP in the *FGF2* gene did not meet quality control criteria [described in Manso et al. 2010] due to low genotyping call rate (<90%) and was excluded. Using logistic regression, we compared genotype frequencies between the poor and good recovery patient groups, after adjusting for significant confounders reflecting stroke severity in a multivariate analysis model: type of stroke, history of hypertension, and occurrence of aphasia,

paresis, altered consciousness and medical complications during hospitalization. Results were corrected for multiple comparisons using the Bonferroni method. Further, we investigated the occurrence of non-additive gene-gene interactions between *BDNF*, *FGF2* and *VEGFA* in stroke outcome using the multifactor-dimensionality reduction (MDR) method [Ritchie et al. 2001]. The MDR is more powerful than traditional regression-based methods [Ritchie et al. 2001], with more than 80% power to detect interactions even in the absence of main effects [Ritchie et al. 2003]. Missing genotypes were imputed for each SNP using the PLINK software [Purcell et al. 2007] using genotypes from stroke patients and CEU HapMap individuals as reference. To correct for multiple testing, the MDR combines cross-validation and permutation procedures [Ritchie et al. 2001]. The statistical significance of the best models was calculated after 1000 permutations using the MDR Permutation Tool (v1.0, beta 2), and the OR and 95% CI for each genotype combination was determined using the OR-based MDR (OR-MDR) (v1.3-1) [Chung et al. 2007]. To determine the gain in information on outcome status by combining two variables together over that provided by the independent analysis of these variables, the MDR uses entropy measures [Jakulin and Bratko 2003]; a positive information gain (IG) indicates a synergistic effect, while a negative IG constitutes evidence for redundancy and a null IG indicates an independent effect. Entropy-based interaction dendrograms were used to interpret epistasis models [Moore et al. 2006].

Multivariate analysis did not provide any strong evidence for an independent effect of any of the tested genes in stroke outcome (Supplementary Table 3). A nominal association of *FGF2* rs12506776 with stroke outcome (P=0.038, Odds Ratio (OR) [95%Confidence Interval (CI)] = 0.60 [0.37-0.98]), did not withstand Bonferroni correction. Restricting the analysis to ischemic patients did not significantly improve these results (Supplementary Table 3). MDR analysis identified two gene interaction models, from all possible 2- to 4-marker models tested, as significant predictors of stroke outcome (Table 1). The interaction between *BDNF* SNP rs10835210 and three *FGF2* SNPs (rs167428, rs308379 and rs3804158) had a statistically significant testing balanced accuracy (TBA) of 0.592, thus correctly classifying 59.2% of the individuals tested (P=0.026 after 1000 fold permutations), a cross-validation consistency (CVC) of 6/10, indicating that the model was selected 6 times out of 10 cross validation subsets, and a global OR of 4.15 [95%CI=2.86-6.04]. The interaction dendrogram presented in figure 1A shows a non-linear (epistatic) synergistic interaction between *BDNF*

rs10835210 and *FGF2* rs3804158, while rs167428 and rs308379 have redundant and independent effects in the model, respectively. A second significant model provides evidence for an interaction between *FGF2* SNP rs167428 and two SNPs in *VEGFA* (rs3025000 and rs6900017), and was a better predictor of stroke outcome, with a higher TBA of 0.611 (P=0.002 after 1000 fold permutations), a CVC of 3/10 and a global OR of 2.54 [95%CI=1.76-3.67]. The interaction dendrogram (figure 1B) indicates that *FGF2* rs167428 and *VEGFA* rs6900017 interact in a synergistic manner, whereas rs3025000 is redundant in this model.

These results suggest that recovery after stroke can be modulated by epistatic interactions between the tested growth factor genes, a finding that now warrants validation in independent population samples. They further provide additional support for testing non-additive interactions in complex phenotypes, even in the absence of main gene effects, indicating that if a gene functions primarily through a complex network involving other genes, its effects may be missed by more traditional regression-based methods [Cordell 2009; Moore 2003]. The identified interactions are supported by previous *in vitro* and *in vivo* studies, showing that HBGF-2 regulates VEGF-A-induced angiogenesis [Mandriota and Pepper 1997], and that synergistic effects of HBGF-2, BDNF and neurotrophin-3 enhance neuronal survival and promoted axon regeneration [Logan et al. 2006]. Several studies using animal models of stroke have also shown that administration of specific growth factors after stroke leads to improvements in neurological function and induction of neuroprotective and repair mechanisms, including neurogenesis and angiogenesis [Chen et al. 2005; Ikeda et al. 2005; Schäbitz et al. 2007; Sun et al. 2003]. While direct biological interpretation of the present findings is not yet possible, testing for gene-gene interactions can be a valuable strategy towards the identification of biochemical pathways involved in stroke outcome, highlighting the multiplicity of neurotrophic factors and complex mechanisms underlying this trait.

Acknowledgements

The authors are grateful to all study participants and their families. The authors wish to thank Dr. Marinho Falcão, Dr. Carlos Dias and the whole team at Instituto Nacional de

Saúde Dr. Ricardo Jorge, all the clinicians that recruited study subjects and the technical assistance provided by the Genotyping Unit at Instituto Gulbenkian de Ciência. This work was supported by the Marie Curie International Reintegration Grant 513760, the Marie Curie Intra-European Fellowship 024563, the FCT grant PTDC/SAU-GMG/64426/2006, and fellowships from FCT and the Portuguese Instituto do Emprego e Formação Profissional.

References

Chen J, Zhang C, Jiang H, Li Y, Zhang L, Robin A, Katakowski M, Lu M, Chopp M. (2005) Atorvastatin induction of VEGF and BDNF promotes brain plasticity after stroke in mice. J Cereb Blood Flow Metab 25:281-290

Cordell HJ. (2009) Detecting gene-gene interactions that underlie human diseases. Nat Rev Genet 10:392-404

Ikeda N, Nonoguchi N, Zhao MZ, Watanabe T, Kajimoto Y, Furutama D, Kimura F, Dezawa M, Coffin RS, Otsuki Y, Kuroiwa T, Miyatake S. (2005) Bone marrow stromal cells that enhanced fibroblast growth factor-2 secretion by herpes simplex virus vector improve neurological outcome after transient focal cerebral ischemia in rats. Stroke 36:2725-2730

Jakulin A, Bratko I. (2003) Analyzing attribute interactions. Lect. Notes Artif. Intell. 2838:229-240

Krupinski J, Kaluza J, Kumar P, Wang M, Kumar S. (1993) Prognostic value of blood vessel density in ischaemic stroke. Lancet 342:742

Lloyd-Jones D, Adams R, Carnethon M, De Simone G, Ferguson TB, Flegal K, Ford E, Furie K, Go A, Greenlund K, Haase N, Hailpern S, Ho M, Howard V, Kissela B, Kittner S, Lackland D, Lisabeth L, Marelli A, McDermott M, Meigs J, Mozaffarian D, Nichol G, O'Donnell C, Roger V, Rosamond W, Sacco R, Sorlie P, Stafford R, Steinberger J, Thom T, Wasserthiel-Smoller S, Wong N, Wylie-Rosett J, Hong Y, Subcommittee AHASCaSS. (2009) Heart disease and stroke statistics--2009 update: a report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Circulation 119:e21-181

Logan A, Ahmed Z, Baird A, Gonzalez AM, Berry M. (2006) Neurotrophic factor synergy is required for neuronal survival and disinhibited axon regeneration after CNS injury. Brain 129:490-502

Mandriota SJ, Pepper MS. (1997) Vascular endothelial growth factor-induced in vitro angiogenesis and plasminogen activator expression are dependent on endogenous basic fibroblast growth factor. J Cell Sci 110 (Pt 18):2293-2302

Manso H, Krug T, Sobral J, Albergaria I, Gaspar G, Ferro JM, Oliveira SA, Vicente AM. (2010) Variants of the Matrix Metalloproteinase-2 but not the Matrix Metalloproteinase-9 genes significantly influence functional outcome after stroke. BMC Med Genet 11:40

Mitsios N, Gaffney J, Kumar P, Krupinski J, Kumar S, Slevin M. (2006) Pathophysiology of acute ischaemic stroke: an analysis of common signalling mechanisms and identification of new molecular targets. Pathobiology 73:159-175

Moore JH. (2003) The ubiquitous nature of epistasis in determining susceptibility to common human diseases. Hum Hered 56:73-82

Moore JH, Gilbert JC, Tsai CT, Chiang FT, Holden T, Barney N, White BC. (2006) A flexible computational framework for detecting, characterizing, and interpreting statistical patterns of epistasis in genetic studies of human disease susceptibility. J Theor Biol 241:252-261

Nakatomi H, Kuriu T, Okabe S, Yamamoto S, Hatano O, Kawahara N, Tamura A, Kirino T, Nakafuku M. (2002) Regeneration of hippocampal pyramidal neurons after ischemic brain injury by recruitment of endogenous neural progenitors. Cell 110:429-441

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559-575

Raab S, Plate KH. (2007) Different networks, common growth factors: shared growth factors and receptors of the vascular and the nervous system. Acta Neuropathol 113:607-626

Ritchie MD, Hahn LW, Moore JH. (2003) Power of multifactor dimensionality reduction for detecting gene-gene interactions in the presence of genotyping error, missing data, phenocopy, and genetic heterogeneity. Genet Epidemiol 24:150-157

Ritchie MD, Hahn LW, Roodi N, Bailey LR, Dupont WD, Parl FF, Moore JH. (2001) Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. Am J Hum Genet 69:138-147

Schäbitz WR, Steigleder T, Cooper-Kuhn CM, Schwab S, Sommer C, Schneider A, Kuhn HG. (2007) Intravenous brain-derived neurotrophic factor enhances poststroke sensorimotor recovery and stimulates neurogenesis. Stroke 38:2165-2172

Sun Y, Jin K, Xie L, Childs J, Mao XO, Logvinova A, Greenberg DA. (2003) VEGF-induced neuroprotection, neurogenesis, and angiogenesis after focal cerebral ischemia. J Clin Invest 111:1843-1851

Figures

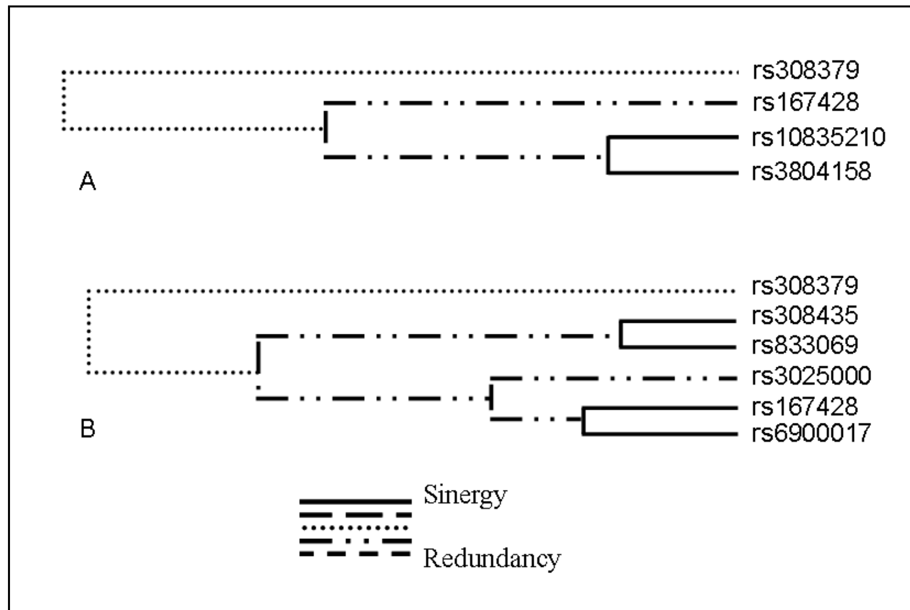

Figure 1 – A) Interaction dendrogram for the *BDNF* and *FGF2* polymorphisms in stroke outcome, showing a synergistic effect between the *BDNF* rs10835210 and *FGF2* rs3804158, an independent effect for *FGF2* rs308379 and redundancy for *FGF2* rs167428. B) Interaction dendrogram for the *FGF2* and *VEGFA* polymorphisms in stroke outcome, showing a synergistic interaction between *FGF2* rs167428 and *VEGFA* rs6900017, and redundancy for *VEGFA* rs3025000. The length of the dendrogram branch that connects two polymorphisms indicates the strength of interaction (the shorter the branch, the stronger is the interaction).

Tables

Table 1 – Gene x gene interaction models obtained using the multifactor-dimensionality reduction (MDR) method in stroke outcome.

Genes	Best model				global OR [95% CI]	CVC	TBA	P^*
	SNP1	SNP2	SNP3	SNP4				
BDNF_FGF2	rs10835210(*BDNF*)	rs167428(*FGF2*)	rs308379(*FGF2*)	rs3804158(*FGF2*)	4.15 [2.86-6.04]	6/10	0.592	**0.026**
BDNF_VEGFA	rs833069(*VEGFA*)	rs3025035(*VEGFA*)	rs6905288(*VEGFA*)		2.15 [1.50-3.07]	4/10	0.559	0.143
FGF2_VEGFA	rs167428(*FGF2*)	rs3025000(*VEGFA*)	rs6900017(*VEGFA*)		2.54 [1.76-3.67]	3/10	0.611	**0.002**
BDNF_FGF2_VEGFA	rs10835210(*BDNF*)	rs308441(*FGF2*)	rs308379(*FGF2*)	rs833069(*VEGFA*)	4.07 [2.76-5.99]	2/10	0.553	0.236

CVC – Cross Validation Consistency, TBA – Testing Balanced Accuracy.

*P after 1000 permutations

Supplementary tables

Supplementary table 1 – Growth factors influence on stroke outcome and recovery.

Growth factor gene	Function	Reference
BDNF	Human neural stem cells overexpressing BDNF promoted functional recovery and neuroprotection after intracerebral hemorrhage	Lee *et al* . (2010) J Neurosci Res. 88(15):3282-94
	Summary of the BDNF effects in the vascular and nervous systems	Raab and Plate (2007) Acta Neuropathol 113:607-626
	BDNF administration stimulated neurogenesis and enhanced stroke recovery after cerebral ischemia	Schäbitz *et al* . (2007) Stroke 38:2165-2172
FGF2	Summary of the HBGF-2 effects in the vascular and nervous systems	Raab and Plate (2007) Acta Neuropathol 113:607-626
	Treatment with HBGF-2 led to neurogenesis and improved motor behaviour in ischemic stroke	Leker *et al* . (2007) Stroke 38:153-161
	Raised HBGF-2 expression in the penumbra of ischemic patients; elevated serum levels of serum HBGF-2 in patients with ischemic stroke or intracerebral hemorrhage	Issa *et al* . (2005) Angiogenesis 8:53-62
	HBGF-2 administration improved neurological outcome and reduced infarct volume	Ikeda *et al* . (2005) Stroke 36:2725-2730
VEGFA	High serum levels of growth factors were associated with good outcome and reduced brain lesion in intracerebral hemorrhage	Sobrino *et al* . (2009) J Cereb Blood Flow Metab, 29: 1968–1974
	Intracerebral hemorrhage induced cerebral angiogenesis and upregulation of VEGF-A	Tang *et al* . (2007) Brain Research, 1175: 134–142
	Summary of the VEGF-A effects in the vascular and nervous systems	Raab and Plate (2007) Acta Neuropathol 113:607-626
	VEGF-A administration improved functional outcome after ischemic stroke	Wang *et al* . (2006) Brain Res 1115:186-193
	VEGF-A administration reduced infarct size and induced neuroprotection, neurogenesis and angiogenesis	Sun *et al* . (2003) J Clin Invest 111:1843-1851

Supplementary table 2 – Demographic and clinical characteristics of stroke patients analyzed for outcome at three months. Good and poor recovery was assessed using the modified Rankin Scale (mRS), which measures the global disability of a patient after stroke. This scale has been used in clinical practice to evaluate the effect of stroke on patient's activities and as a primary end point in randomized clinical trials [Kasner 2006].

Characteristic	Good Recovery (mRS≤1)	Poor Recovery (mRS>1)	*P* [*]
Age and Gender			
Age, mean±SD (yrs)	50.8±9	52.5±8.5	0.028
Gender (male), n/N (%)	174/276 (63.0)	174/270 (64.4)	0.734
Past History, n/N (%)			
Hypertension	159/241 (66.0)	143/240 (59.6)	0.147
Diabetes	36/259 (13.9)	47/246 (19.1)	0.115
Cardiac Disease	37/264 (14.0)	43/257 (16.7)	0.390
Stroke type, n/N (%)			$<10^{-4}$
Ischemic stroke	238/276 (86.2)	193/270 (71.5)	–
Hemorrhagic stroke	33/276 (12.0)	72/270 (26.7)	–
Unknow type of stroke	5/276 (1.8)	5/270 (1.9)	–
Stroke Features, n/N (%)			
Aphasia	53/258 (20.5)	98/250 (39.2)	$<10^{-4}$
Neglect	11/266 (4.1)	19/240 (7.9)	0.072
Dysphagia	15/270 (5.6)	25/251 (10.0)	0.059
Urinary Incontinence	5/272 (1.8)	15/251 (6.0)	0.014
Paresis	203/273 (74.4)	244/269 (90.7)	$<10^{-4}$
Altered consciousness	21/275 (7.6)	59/265 (22.3)	$<10^{-4}$
Medical complications	18/265 (6.8)	82/254 (32.3)	$<10^{-4}$
Neurologic complications	14/274 (5.1)	39/267 (14.6)	2.03×10^{-4}

SD – standard deviation, yrs – years.

[*]Mann-Whitney test or Pearson's χ^2 test.

Supplementary table 3 – Association analysis results for *BDNF*, *FGF2* and *VEGFA* SNPs and stroke outcome.

gene	marker	position*	location*	MA	Whole sample						Ischemic Subset					
					HWE *P*	missing genotypes (%)	MAF good recovery	MAF poor recovery	*P* †	*P* ‡	HWE *P*	missing genotypes (%)	MAF good recovery	MAF poor recovery	*P* †	*P* ‡
FGF2	rs308393	123746619	upstream	C	1.000	**10.8**	0.143	0.144	-	-	0.536	**10.4**	0.140	0.149	-	-
	rs308395	123746942	upstream	G	0.058	1.3	0.158	0.170	0.907	NS	0.291	1.6	0.154	0.182	0.568	NS
	rs308420	123767943	intronic	A	0.734	0.4	0.071	0.067	0.130	NS	1.000	0.5	0.068	0.070	-	-
	rs308418	123768263	intronic	G	0.046	0.2	0.164	0.172	0.665	NS	0.226	0.2	0.158	0.181	0.975	NS
	rs308417	123768337	intronic	C	0.242	8.6	0.066	0.052	0.106	NS	0.380	7.7	0.065	0.048	0.259	NS
	rs308435	123772018	intronic	A	0.104	0.4	0.165	0.184	0.364	NS	0.169	0.5	0.158	0.180	0.233	NS
	rs11938826	123772614	intronic	G	0.142	0.5	0.158	0.155	0.254	NS	0.582	0.7	0.158	0.154	0.437	NS
	rs167428	123773439	intronic	C	0.145	0.2	0.281	0.299	0.457	NS	0.194	0.2	0.284	0.284	0.635	NS
	rs308439	123773579	intronic	G	0.274	0.2	0.060	0.071	0.914	NS	1.000	0.2	0.059	0.052	0.645	NS
	rs308441	123774065	intronic	T	0.519	1.1	0.196	0.228	0.424	NS	0.545	1.4	0.198	0.202	0.742	NS
	rs17472986	123774832	intronic	A	0.495	0.2	0.096	0.112	0.168	NS	0.411	0.2	0.097	0.102	0.375	NS
	rs308442	123774913	intronic	A	0.427	0.2	0.304	0.326	0.619	NS	0.574	0.2	0.310	0.311	0.998	NS
	rs17407577	123779341	intronic	C	0.690	0.4	0.064	0.050	0.781	NS	0.654	0.5	0.068	0.049	0.623	NS
	rs1960669	123782568	intronic	A	0.456	0.4	0.102	0.086	0.389	NS	0.275	0.5	0.105	0.089	0.417	NS
	rs308379	123782896	intronic	A	0.326	0.4	0.307	0.335	0.724	NS	0.187	0.5	0.316	0.333	0.709	NS
	rs308382	123783194	intronic	C	0.561	0.7	0.184	0.178	0.533	NS	0.520	0.7	0.190	0.178	0.438	NS
	rs12644427	123783387	intronic	G	0.713	6.4	0.146	0.133	0.111	NS	0.259	7.0	0.137	0.115	0.150	NS
	rs308388	123787079	intronic	A	0.707	1.1	0.355	0.354	0.130	NS	0.451	1.4	0.353	0.332	0.113	NS
	rs7694627	123788495	intronic	C	0.413	0.7	0.152	0.160	0.744	NS	0.858	0.9	0.157	0.164	0.978	NS
	rs17474021	123798711	intronic	C	0.411	2.0	0.154	0.157	0.681	NS	1.000	2.6	0.159	0.160	0.893	NS
	rs17006255	123798993	intronic	C	0.578	5.9	0.142	0.132	0.052	NS	0.490	5.6	0.135	0.110	**0.047**	NS
	rs17408557	123800617	intronic	G	0.585	0.4	0.140	0.134	1.000	NS	1.000	0.5	0.146	0.135	0.916	NS
	rs12506776	123803370	intronic	T	0.707	4.2	0.143	0.122	**0.038**	NS	0.361	3.7	0.136	0.103	**0.035**	NS
	rs12503378	123810734	intronic	G	0.719	4.9	0.144	0.137	0.083	NS	0.503	4.6	0.137	0.112	0.063	NS
	rs3804158	123814900	3' UTR	G	0.663	1.8	0.462	0.445	0.200	NS	0.491	1.6	0.460	0.428	0.196	NS
	rs6854081	123816707	3' UTR	G	0.380	0.4	0.147	0.134	0.861	NS	0.846	0.5	0.149	0.136	0.781	NS
	rs7683093	123818085	3' UTR	G	1.000	7.3	0.136	0.135	0.891	NS	0.676	7.2	0.140	0.140	0.963	NS
	rs1476217	123818511	3' UTR	C	0.283	6.6	0.349	0.344	0.200	NS	0.219	5.8	0.345	0.312	0.134	NS

Supplementary table 3 (cont.) – Association analysis results for *BDNF*, *FGF2* and *VEGFA* SNPs and stroke outcome.

gene	marker	position*	location*	MA	Whole sample						Ischemic Subset					
					HWE *P*	missing genotypes (%)	MAF good recovery	MAF poor recovery	*P* †	*P* ‡	HWE *P*	missing genotypes (%)	MAF good recovery	MAF poor recovery	*P* †	*P* ‡
BDNF	rs4923460	27656789	downstream	T	0.897	0.9	0.211	0.209	0.832	NS	0.459	0.7	0.205	0.204	0.811	NS
	rs925946	27667202	downstream	T	0.833	2.7	0.282	0.298	0.653	NS	0.728	2.8	0.298	0.304	0.859	NS
	rs7927728	27667472	downstream	A	0.195	1.8	0.067	0.081	0.548	NS	0.282	1.6	0.071	0.081	0.509	NS
	rs1519479	27667531	downstream	T	0.931	1.1	0.467	0.463	0.569	NS	0.497	1.2	0.460	0.463	0.422	NS
	rs2203877	27670910	downstream	C	0.794	2.7	0.467	0.465	0.516	NS	0.922	2.6	0.462	0.470	0.369	NS
	rs6265	27679916	exonic	T	0.412	0.5	0.191	0.200	0.746	NS	0.115	0.7	0.188	0.194	0.593	NS
	rs1401635	27693991	intronic	C	0.755	1.1	0.293	0.292	0.902	NS	0.730	1.2	0.306	0.293	0.869	NS
	rs10835210	27695910	intronic	A	0.425	0.9	0.400	0.419	0.276	NS	0.269	1.2	0.390	0.424	0.131	NS
	rs11030119	27728102	upstream/intronic	A	0.756	0.5	0.290	0.290	0.932	NS	0.728	0.7	0.301	0.289	0.859	NS
	rs962369	27734420	upstream/intronic	C	1.000	1.1	0.286	0.285	0.902	NS	1.000	1.2	0.297	0.284	0.877	NS
	rs11030121	27736207	upstream/intronic	T	0.362	7.1	0.317	0.333	0.358	NS	0.369	6.0	0.329	0.330	0.640	NS
VEGFA	rs25648	43738977	upstream/5'UTR/exonic	T	0.426	0.5	0.159	0.164	0.519	NS	0.591	0.5	0.158	0.164	0.375	NS
	rs833069	43742579	intronic	C	0.244	4.4	0.337	0.349	0.570	NS	0.369	5.1	0.321	0.330	0.819	NS
	rs3024994§	43743507	intronic	T	0.244	0.4	0.040	0.037	0.828	NS	0.178	0.5	0.034	0.044	0.336	NS
	rs3025000	43746169	intronic	T	0.062	0.4	0.295	0.288	0.342	NS	0.189	0.5	0.289	0.273	0.481	NS
	rs3025035	43751359	intronic	T	0.665	0.7	0.105	0.120	0.148	NS	0.809	0.9	0.103	0.126	0.084	NS
	rs6899540	43758324	downstream	C	0.856	0.7	0.123	0.147	0.888	NS	1.000	0.7	0.118	0.147	0.899	NS
	rs6900017	43758485	downstream	T	0.560	0.4	0.120	0.131	0.198	NS	1.000	0.5	0.119	0.127	0.163	NS
	rs6905288	43758873	downstream	G	1.000	0.9	0.396	0.382	0.120	NS	1.000	0.9	0.390	0.385	0.273	NS

HWE – Hardy-Weinberg equilibrium, MA – minor allele, MAF – minor allele frequency, NS – non-significant, UTR – untranslated region.

*According to Ensembl Release 56 – September 2009.

†*P* for the log-additive genetic model after adjustment for significant covariates.

‡*P* after Bonferroni correction.

§The log-additive model could not be performed for this SNP; *P* for the codominant model is presented.

Chapter 5. Nitric oxide synthase genes

Manso H, Krug T, Sobral J, Albergaria I, Gaspar G, Ferro JM, Oliveira SA, Vicente AM (2011) Variants within the nitric oxide synthase 1 gene are associated with stroke susceptibility (Submitted)

Abstract

Background – Animal studies have allowed important insights into the role of the nitric oxide synthase (NOS) enzymes in atherosclerosis, hypertension and stroke, as well as in the post-stroke process of damage control and recovery. In this study we tested the hypothesis that the *NOS1* and *NOS3* genes, respectively encoding neuronal NOS (nNOS) and endothelial NOS (eNOS), influence susceptibility and recovery after a stroke event.

Methods and results – We conducted a case-control association study in 551 ischemic stroke patients and 530 controls to assess the role of *NOS1* and *NOS3* variants in stroke susceptibility. Further, the same genes were tested for association with stroke outcome in a subset of 431 patients, and the effect of gene-gene interactions in both phenotypes was also investigated. Four *NOS1* polymorphisms and four haplotypes were significantly associated with stroke susceptibility after adjusting for demographic, clinical and life-style risk factors, and correcting for multiple testing using the false discovery rate (FDR) method (0.036< SNP FDR q<0.048 and 0.018<haplotype FDR q<0.032). *NOS1* variants were not associated with stroke outcome. We did not find any evidence for association between *NOS3* and stroke susceptibility or outcome, or for significant *NOS1*-*NOS3* synergistic interactions in either phenotype.

Conclusions – Our results highlight *NOS1* as a susceptibility factor for stroke, but do not corroborate previous *NOS3* association findings with stroke risk. nNOS is known to play a major role in atherosclerosis development and in blood flow regulation, and its influence in stroke may therefore be mediated through these two main clinical risk factors.

Keywords: Atherosclerosis, Hypertension, Nitric Oxide Synthase, Recovery, Stroke, Susceptibility.

Introduction

Nitric oxide (NO) is an important inter- and intracellular messenger in several types of cells. It is synthesized via the oxidative L-arginine pathway by three distinct isoforms of nitric oxide synthase (NOS) enzymes, neuronal NOS (nNOS), inducible NOS (iNOS) and endothelial NOS (eNOS), which are dependent on cofactor binding and dimerization to become active [Alderton et al. 2001]. Each of these enzymes shows a predominant tissue expression pattern, however with some overlap [Liu and Huang 2008]. nNOS is the predominant synthase in neurons, whereas eNOS and iNOS are mainly expressed in endothelial cells and macrophages, respectively.

Known risk factors for stroke, including atherosclerosis and hypertension, and stroke itself have been linked to abnormalities in NO signaling [Liu and Huang 2008]. NO produced by endothelial cells causes vasodilation and hypotension, and has several anti-thrombotic and antiatherosclerotic properties, such as inhibition of leukocyte adhesion and migration, and reduction of smooth muscle cells proliferation [Toda et al. 2009a]. In the central nervous system (CNS) or in the peripheral nervous system (PNS), NO acts as a neurotransmitter and regulates multiple autonomic nervous system processes, including vasodilation and smooth muscle cells relaxation in the gastrointestinal, respiratory and genitourinary tracts [Huang 1999; Liu and Huang 2008; Toda et al. 2009a]. Genetically engineered animals overexpressing or with a disruption in NOS-encoding genes have provided particularly valuable insights into the pathophysiology of NOS enzymes in several disorders [reviewed in Liu and Huang 2008]. Apolipoprotein E/nitric oxide synthase 1 (*Apoe*/*Nos1*) and *Apoe*/nitric oxide synthase 3 (*Nos3*) double knockouts (KO) develop greater atherosclerotic vascular lesions than *Apoe* KO mice, an animal model for human diet-induced atherosclerosis [Kuhlencordt et al. 2001; Kuhlencordt et al. 2006], indicating that nNOS and eNOS have a vasculoprotective role in atherosclerosis. Studies with *Nos3* KO mice and nNOS inhibitors have also shown the involvement of both eNOS and nNOS enzymes in maintenance of blood pressure [Shesely et al. 1996; Talman and Nitschke Dragon 2007]. Further, *Nos1* KO mice have decreased infarct volumes and less neurological deficits after ischemic stroke than controls, while *Nos3* KO mice develop larger infarct sizes, suggesting that nNOS contributes to tissue damage and eNOS has a protective role [Huang et al. 1996; Huang et al. 1994].

These observations implicate nNOS and eNOS in pathological conditions that increase stroke risk, like atherosclerosis and hypertension, and in the injury or recovery mechanisms that occur after stroke. Genes encoding these enzymes are therefore strong candidates for a role in stroke risk and recovery. Some polymorphisms within the *NOS3* gene (encoding the eNOS enzyme), such as the Glu298Asp polymorphism (rs1799983) known to impair the enzymatic activity of eNOS, have been previously tested for association with stroke [reviewed in Bersano et al. 2008]. However, contradictory results have been obtained in different population samples. On the other hand, no association study assessing the role of *NOS1* gene (encoding the nNOS enzyme) in stroke has been published so far. We thus tested the association of these two genes with stroke susceptibility and with stroke outcome at three months.

According to the STRING database [Jensen et al. 2009], which contains information on physical and functional protein-protein interactions, nNOS and eNOS are predicted functional partners. Given the likely ubiquity of gene-gene interactions in common human diseases [Moore 2003], we also investigated the existence of *NOS1*-*NOS3* interactions in stroke risk and outcome.

Methods

Population Sample

The study sample included 551 first-ever ischemic stroke patients, recruited through Neurology and Internal Medicine Departments of several hospitals in Portugal. The assessment methods, inclusion criteria and history on previous stroke risk factors were previously described [Manso et al. 2011]. 530 healthy controls with no clinical history of stroke were also enrolled. Since stroke is a late-onset disease, we included older healthy individuals to reduce the probability of misclassification as controls. Demographic, clinical and life-style risk factors were compared between patients and controls and adjusted for, if significantly different (Table 1). A subset of 431 patients, for whom clinical data on functional recovery was available, was included in the outcome analysis. Occurrence of permanent altered consciousness, urinary incontinence, medical and neurological complications, neglect, aphasia, paresis, gaze

paresis and dysphagia during hospitalization were clinical parameters indicative of stroke severity. Stroke outcome was assessed, by direct interview, using the modified Rankin Scale (mRS) at three months after the stroke event. This scale was used to classify patients into two groups: patients with mRS≤1 were scored as "good recovery" and patients with mRS>1 were scored as "poor recovery".

SNP selection and genotyping

Haplotype tagging SNPs were selected across the *NOS1* and *NOS3* genomic regions (gene ± 5kb) with the Haploview software (v4.2) [Barrett et al. 2005] (HapMap Release 21/phase II July 2006) to capture the complete genetic variability in these *loci*. 31 *NOS1* SNPs and 4 *NOS3* SNPs were genotyped using Sequenom iPLEX assays with allele detection by mass spectroscopy, using Sequenom MassARRAY technology (Sequenom, San Diego, California) and following the manufacturer's protocol. Primer sequences were designed using Sequenom's MassARRAY Assay Design 3.0 software. Genotyping of 6 *NOS1* SNPs and 1 *NOS3* SNP were performed using TaqMan® Pre-Designed SNP Genotyping Assays in an ABI PRISM 7900HT Sequence Detector System (Applied Biosystems, Foster City, USA). All genotyping plates contained quality control samples: four no-template controls, eight HapMap individuals, duplicated samples within and across genotyping plates, and three large pedigrees to check for Mendelian inconsistencies. SNPs in Hardy-Weinberg equilibrium (P>0.05) and with a genotyping call rate above 90% were further analyzed. Genotype determinations were blinded to affection and/or stroke outcome status.

Statistical analysis

To identify potential confounders, univariate analyses were performed. Specifically, demographic, clinical and life-style risk factors were compared between patients and controls, using the Pearson's χ^2 test and the Mann-Whitney test for discrete and continuous variables, respectively. Variables with a P<0.25 in univariate analysis (Table 1) or of particular clinical relevance were included in a logistic regression model using forward selection and were maintained in the model if they were associated with stroke susceptibility at a P≤0.05 level [Hosmer and Lemeshow 2000]. Covariates in the final logistic regression model were not correlated (-0.5<interaction i<0.5). The effect

of each genetic variable on stroke susceptibility was determined by logistic regression, after adjustment for the significant covariates. Odds ratio (OR) and 95% confidence intervals (95% CI) were computed for the log-additive model. The analysis of stroke outcome followed a similar procedure. Demographic and clinical data reflecting the severity of stroke was compared between patients with poor (mRS>1) and good (mRS≤1) outcome at three months to identify potential confounders (supplementary table 3). The effect of each genetic variable on patient's outcome was determined by logistic regression, after adjustment for the significant covariates. OR and 95% CI were also computed for the log-additive model. Univariate and logistic regression analyses were performed using MASS and SNPassoc packages of the R software [R: A language and Environment for Statistical Computing 2004] (v2.10.1), respectively. Haplotype blocks in the two genes were determined using the Gabriel et al. [2002] or the solid spin of linkage disequilibrium (LD) methods of the Haploview software (v4.2) [Barrett et al. 2005] and haplotype-based association analyses were performed using the same software. To reduce type I error, we applied the false discovery rate (FDR) multiple testing correction in individual SNP and haplotype-based association analyses. SNPs (or haplotypes) with q values ≤5% were considered significant, which resulted in a FDR≤5% among the significant SNPs (or haplotypes). q values were calculated using the qvalue package of the R software [R: A language and Environment for Statistical Computing 2004].

Gene-gene interaction analysis

The multifactor-dimensionality reduction (MDR) method (v2.0, beta 7.2) [Ritchie et al. 2001], a nonparametric and genetic model-free approach, was used to test for genetic interactions in association with stroke susceptibility and outcome. Briefly, by pooling multilocus genotypes into high and low risk groups, the MDR reduces the dimensionality of the data from N dimensions to one dimension. The new multilocus genotype attribute is then tested for its ability to classify and predict disease status, or good/poor outcome at three months. False-positive results due to multiple testing are reduced through combination of a cross-validation strategy and permutation testing [Moore 2003; Ritchie et al. 2001]. Missing genotypes were imputed for each SNP using the PLINK software (http://pngu.mgh.harvard.edu/purcell/plink/) [Purcell et al. 2007] and individual genotypes obtained in the present study and from CEU HapMap subjects

were used as reference. The best interaction models were chosen from among the best 2, 3 and 4-marker models, based on the testing balanced accuracy (TBA), which measures how often individuals are correctly classified in relation to disease or outcome status, and the cross-validation consistency (CVC) of 10 cross-validation intervals, which measures the number of times the MDR found the same set of *loci* across the cross validation subsets. The statistical significance of the best models was calculated after 1000 permutations using the MDR Permutation Tool (v1.0, beta 2). To interpret interaction models and identify non-linear interactions, MDR uses measures of entropy [Jakulin and Bratko 2003], which is the extent of uncertainty associated with a variable, to assess information gain (IG). When the combination of two or more SNPs gives a positive information gain (IG), there is evidence for a synergistic interaction; when IG is negative, SNPs are correlated; and if IG=0, the SNPs have independent effects.

Results

The demographic and clinical characteristics of investigated patients and controls are reported in Table 1. Hypertension, diabetes, smoking and alcohol consumption – four well-known stroke risk factors – were significantly more frequent in patients than in controls, as shown by univariate analyses. The frequency of stroke during sample collection was higher in males than females, as expected. Multivariate analyses were performed subsequent to univariate analyses to identify potential non-genetic confounders. The final logistic regression model included the covariates significant in the multivariate analysis model – gender, hypertension, diabetes and smoking – as well as genetic markers.

All SNPs genotyped in the *NOS1* and *NOS3* genes met quality control criteria and were further evaluated. Figure 1 shows the association results and the LD pattern among all genotyped markers in our sample. Seven *NOS1* SNPs were associated with ischemic stroke under a log-additive model, after adjusting for significant covariates ($0.004<P<0.042$) (Table 2, Supplementary table 1). Four of these SNPs (rs2293050, rs2139733, rs7308402 and rs1483757) remained significantly associated with ischemic stroke after correcting for multiple testing (0.036<FDR q<0.048). We observed that rs2293050 and rs2139733, two of the four significant SNPs, are in almost complete LD

(r^2 ≈0.97) and may therefore signal the same variant. Eight haplotypes in the *NOS1* gene were associated with stroke (0.001<*P*<0.046) (Table 3, Supplementary table 2), four of which remained significant after multiple testing correction (0.018<FDR *q*<0.032). Evidence for association of three of these four haplotypes was increased when compared to that of individually associated markers, with one specific haplotype increasing stroke susceptibility (haplotype A-T-T-G-T for markers rs7977109-rs11068438-rs11611788-rs11068445-rs7298903) and two others protecting from stroke (haplotype C-T-G-A-C for markers rs1607817-rs2293050-rs7314935-rs2139733-rs7309163 and haplotype G-A for markers rs547954-rs7308402) (Tables 2 and 3).

We carried out a bioinformatics analysis to investigate possible functional consequences for gene transcription and/or protein activity of the significant SNPs and haplotypes in *NOS1*. Since they are located in introns and are not splice-site variants, these SNPs are not likely to be functional, but may be in LD with variants affecting the activity and/or subcellular distribution of nNOS. The nNOS protein has five functional domains: PDZ, NO synthase, flavodoxin, FAD binding and NAD binding. The PDZ domain, encoded by part of exon 2, is required to anchor nNOS to the cellular membrane. rs7308402 and rs1483757, as well as haplotype G[rs547954]-A[rs7308402], located in neighbor intron 2, were significantly associated with stroke, and may be in LD with functional variants in this domain, affecting the subcellular distribution and/or activity of the protein. The genomic region from rs1607817 to rs7309163, encompassing the protective haplotype C-T-G-A-C and the associated SNPs rs2293050 and rs2139733, includes exons 5-9 which encode part of the NO synthase domain. Exonic variants in this region may therefore alter nNOS activity.

Two out of five *NOS3* SNPs were associated with stroke after adjusting for the previously mentioned significant covariates (rs1800783: *P*=0.014, OR[95%CI]=0.79 [0.65-0.95]; rs2373929: *P*=0.027, OR[95%CI]=1.24 [1.02-1.50]) (Table 2, Supplementary table 1), but none of these associations withstood multiple testing correction. No haplotype was significantly associated with stroke susceptibility (Supplementary table 2).

Gene-gene interactions are likely to have an important impact on complex phenotypes, including susceptibility to common stroke. Thus, we investigated the existence of gene-gene interactions between *NOS1* and *NOS3* using the MDR method. All possible 2- to 4-marker models were tested. The most significant interaction model in stroke susceptibility corresponded to an interaction between three SNPs in *NOS1*

(rs1093330, rs2139733 and rs7977109) and one SNP in *NOS3* (rs2373929). This model had a statistically significant TBA of 0.577 (*P*=0.010, based on 1000-fold permutations), thus being able to correctly classify approximately 58% of the individuals tested, a CVC of 9/10, indicating that the model was selected 9 times out of 10 cross validation subsets, and a global OR of 3.13 [95%CI=2.41-4.06]. Table 4 shows that six genotype combinations of these four SNPs protected from stroke, while two others increased susceptibility to this disease. The genetic effects identified by MDR were interpreted using the entropy-based interaction dendrogram presented in Figure 2A, which highlights the gain of information on case/control status by knowing the genotypes at the four SNPs. The interaction dendrogram shows a non-linear synergistic interaction between two *NOS1* SNPs, rs1093330 and rs7977109, whereas *NOS1* rs2139733 and *NOS3* rs2373929 have redundant and independent effects in the model, respectively. These results suggest that the effects of the *NOS1* and *NOS3* genes on stroke susceptibility are independent.

The impact of *NOS1* and *NOS3* genetic variants in patient's outcome at three months was also analyzed. This was investigated in a subset of 431 stroke patients for whom clinical information during hospitalization and at three months was available (Supplementary table 3). It was observed that clinical characteristics reflecting the severity of stroke (occurrence of aphasia, urinary incontinence, paresis, altered consciousness, medical and neurological complications during hospitalization) were significantly more frequent in patients with poor outcome at three months, as expected. SNP association analyses did not provide any strong evidence for an independent effect of *NOS1* or *NOS3* in stroke outcome (Supplementary tables 4 and 5). Since the MDR method may identify gene-gene interactions even in the absence of independent effects, we investigated the existence of *NOS1*-*NOS3* interactions that might influence patient's recovery after stroke. The best interaction model that was identified corresponded to a 4-marker combination between three SNPs in *NOS1* (rs11068428, rs2139733 and rs7977109) and one SNP in *NOS3* (rs743507). This model correctly classified approximately 57.3% of the individuals (TBA of 0.573, *P*=0.051 based on 1000-fold permutations), and was selected 6 times out of 10 cross-validation subsets (CVC of 6/10); its global OR was 5.08 [3.29-7.83]. The entropy-based interaction dendrogram presented in Figure 2B shows an epistatic interaction between one *NOS1* SNP (rs2139733) and one *NOS3* SNP (rs743507). Nevertheless, since the TBA of the

interaction model was only nominal ($_{permuted}P$=0.051), it is unlikely that a *NOS1-NOS3* interaction strongly contributes to patient's outcome after stroke.

Discussion

The *NOS3* gene is predominantly expressed in the endothelium, a key player in vasodilation, and *Nos3* KO mice are hypertensive and have vascular abnormalities [Shesely et al. 1996]. *NOS3* was thus identified early as a candidate gene for stroke and has been repeatedly tested for association with this disease. *NOS1* gene expression was initially found in neurons and the aggressive behaviour and enlarged stomachs that characterize *Nos1* KO mice [Huang et al. 1993] were not suggestive of a role of this enzyme in stroke. However, it is nowadays clear that nNOS has an important vasculoprotective role in atherosclerosis and is involved in blood pressure control [Kuhlencordt et al. 2006; Talman and Nitschke Dragon 2007; Toda et al. 2009b]. It has been suggested that NO released from autonomic nitrergic nerves innervating the systemic vasculature, together with endothelial NO, plays a role in vasodilation [reviewed in Toda et al. 2009b]. It is therefore plausible that nNOS influences stroke susceptibility through very different but complementary mechanisms from eNOS. In this study, we provide novel genetic evidence supporting this hypothesis, identifying a significant association of the *NOS1* gene with stroke susceptibility. In the tested population sample, four SNPs and four haplotypes were significantly associated with stroke, after adjusting for demographic, clinical and life-style risk factors. These SNPs were located in or near gene regions that are critical for protein function, namely the PDZ domain, required to anchor nNOS to the cellular membrane, and the NO synthase domain.

In contrast, we found no evidence supporting a role of the *NOS3* gene in stroke risk in our population set. Previously, several *NOS3* variants have been tested for association with stroke, with approximately half of the studies showing positive results and the others failing to confirm these associations [reviewed in Bersano et al. 2008]. Some negative studies might have been underpowered to replicate the initial findings, as their sample sizes were relatively small, but a meta-analysis with 1086 cases and 1089 controls was also negative [Bersano et al. 2008; Casas et al. 2004]. Another study with

the largest sample size thus far, involving the analysis of 1901 patients and 1747 controls, obtained positive results [Berger et al. 2007]. It is possible that contradictory results reflect true differences between populations with distinct ethnic backgrounds, but heterogeneity between studies cannot be excluded, e.g. in the analyzed phenotype (ischemic stroke vs. small vessel stroke), or the age of patients (children vs. adults). In the present work, we covered, for the first time, the whole genetic variability in this region to prevent potential negative association findings due to allelic or LD pattern heterogeneity across populations, which are plausible reasons for inconsistencies between studies. Further, our results were adjusted for significant confounders, not consistently done in all studies, thus decreasing the chance of finding positive associations due to non-genetic risk factors. An effort to carry out association studies in other populations, analyzing the whole genetic variability of this region and not only previously associated SNPs, and taking into account non-genetic risk factors, will be necessary for an adequate meta-analysis, providing the necessary data to properly dissect the contribution of *NOS3* to stroke risk.

Because the NOS isoforms have partially overlapping mechanisms of regulation, physiological functions and tissue distributions, and nNOS and eNOS are predicted functional partners according to the STRING database, we investigated the existence of *NOS1-NOS3* interactions. Identifying interactions between *loci* may contribute to better understanding of the pathophysiological pathways involved in stroke risk and recovery, but this is not often investigated in association studies. Our results, however, did not support any epistatic interactions between the *NOS1* and *NOS3* genes in stroke susceptibility, suggesting on the contrary that the effects of the two genes are independent. We also identified a non-additive interaction between two SNPs in *NOS1*. These markers are far apart from each other (approximately 34kb) and were not individually associated with stroke. It is possible that they are in LD with two functional variants whose interaction may influence folding kinetics and stability of the protein, but further studies are needed to test this hypothesis.

We found no evidence supporting a contribution of *NOS1* and *NOS3* gene variants to patient's recovery. This was unexpected, since previous animal studies have shown that absence of either gene has an impact in neurological outcome and/or infarct volumes after stroke [Huang et al. 1996; Huang et al. 1994]. It is possible that a *NOS1* and *NOS3* gene variant with a modest effect in patient's outcome was undetected in this smaller subset. No significant gene interaction model was identified for this phenotype.

nNOS-derived NO is an important molecule for synaptic plasticity and neuronal signaling, and it also controls blood flow and muscle contractility in the smooth, cardiac and skeletal muscles [Zhou and Zhu 2009]. eNOS and nNOS show atheroprotective effects in a mouse model of atherosclerosis and are known to regulate vasodilation through different mechanisms. NO liberated from the endothelium (produced through eNOS), brain neurons and perivascular nitrergic nerves (produced through nNOS) act on vascular smooth muscle cells to produce relaxation, lowering blood pressure [Toda et al. 2009b]. Endothelial dysfunction and impairment of nitrergic nerve function lead to systemic hypertension [Toda et al. 2009b]. The biological functions of nNOS seem therefore critical for atherosclerosis progression and development of hypertension, with subsequent influence on stroke risk, and may be underlying the positive *NOS1* association results. Interestingly, Nakata et al [2007] showed that statins, which have been used to treat hypercholesterolemia, up-regulate nNOS in human endothelial cells, rat vascular smooth muscle cells and mouse aortas, suggesting yet a different functional mechanism whereby nNOS may be regulating stroke risk through its expression in cells other than neurons. This observation also demonstrates a novel vascular effect of statins mediated through nNOS, suggesting that this enzyme could also be involved in the reduction of stroke incidence that was observed for statins, in addition to the lowering of cholesterol levels [Baigent et al. 2005]. Further understanding of specific nNOS-mediated signaling pathways may highlight these as critical targets for stroke prevention [Zhou and Zhu 2009].

References

Alderton WK, Cooper CE, Knowles RG. (2001) Nitric oxide synthases: structure, function and inhibition. Biochem J 357:593-615

Baigent C, Keech A, Kearney PM, Blackwell L, Buck G, Pollicino C, Kirby A, Sourjina T, Peto R, Collins R, Simes R, Collaborators CTT. (2005) Efficacy and safety of cholesterol-lowering treatment: prospective meta-analysis of data from 90,056 participants in 14 randomised trials of statins. Lancet 366:1267-1278

Barrett JC, Fry B, Maller J, Daly MJ. (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21:263-265

Berger K, Stögbauer F, Stoll M, Wellmann J, Huge A, Cheng S, Kessler C, John U, Assmann G, Ringelstein EB, Funke H. (2007) The glu298asp polymorphism in the nitric oxide synthase 3

gene is associated with the risk of ischemic stroke in two large independent case-control studies. Hum Genet 121:169-178

Bersano A, Ballabio E, Bresolin N, Candelise L. (2008) Genetic polymorphisms for the study of multifactorial stroke. Hum Mutat 29:776-795

Casas JP, Hingorani AD, Bautista LE, Sharma P. (2004) Meta-analysis of genetic studies in ischemic stroke: thirty-two genes involving approximately 18,000 cases and 58,000 controls. Arch Neurol 61:1652-1661

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D. (2002) The structure of haplotype blocks in the human genome. Science 296:2225-2229

Hosmer DW, Lemeshow S. (2000) Applied Logistic Regression. New York: John Wiley & Sons, Inc.

Huang PL. (1999) Neuronal and endothelial nitric oxide synthase gene knockout mice. Braz J Med Biol Res 32:1353-1359

Huang PL, Dawson TM, Bredt DS, Snyder SH, Fishman MC. (1993) Targeted disruption of the neuronal nitric oxide synthase gene. Cell 75:1273-1286

Huang Z, Huang PL, Ma J, Meng W, Ayata C, Fishman MC, Moskowitz MA. (1996) Enlarged infarcts in endothelial nitric oxide synthase knockout mice are attenuated by nitro-L-arginine. J Cereb Blood Flow Metab 16:981-987

Huang Z, Huang PL, Panahian N, Dalkara T, Fishman MC, Moskowitz MA. (1994) Effects of cerebral ischemia in mice deficient in neuronal nitric oxide synthase. Science 265:1883-1885

Jakulin A, Bratko I. (2003) Analyzing attribute interactions. Lect. Notes Artif. Intell. 2838:229-240

Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, Doerks T, Julien P, Roth A, Simonovic M, Bork P, von Mering C. (2009) STRING 8--a global view on proteins and their functional interactions in 630 organisms. Nucleic Acids Res 37:D412-416

Kuhlencordt PJ, Gyurko R, Han F, Scherrer-Crosbie M, Aretz TH, Hajjar R, Picard MH, Huang PL. (2001) Accelerated atherosclerosis, aortic aneurysm formation, and ischemic heart disease in apolipoprotein E/endothelial nitric oxide synthase double-knockout mice. Circulation 104:448-454

Kuhlencordt PJ, Hotten S, Schodel J, Rutzel S, Hu K, Widder J, Marx A, Huang PL, Ertl G. (2006) Atheroprotective effects of neuronal nitric oxide synthase in apolipoprotein e knockout mice. Arterioscler Thromb Vasc Biol 26:1539-1544

Liu VW, Huang PL. (2008) Cardiovascular roles of nitric oxide: a review of insights from nitric oxide synthase gene disrupted mice. Cardiovasc Res 77:19-29

Manso H, Krug T, Sobral J, Albergaria I, Gaspar G, Ferro JM, Oliveira SA, Vicente AM. (2011) Variants in the inflammatory *IL6* and *MPO* genes modulate stroke susceptibility through main effects and gene-gene interactions. J Cereb Blood Flow Metab (*in press*)

Moore JH. (2003) The ubiquitous nature of epistasis in determining susceptibility to common human diseases. Hum Hered 56:73-82

Nakata S, Tsutsui M, Shimokawa H, Yamashita T, Tanimoto A, Tasaki H, Ozumi K, Sabanai K, Morishita T, Suda O, Hirano H, Sasaguri Y, Nakashima Y, Yanagihara N. (2007) Statin treatment upregulates vascular neuronal nitric oxide synthase through Akt/NF-kappaB pathway. Arterioscler Thromb Vasc Biol 27:92-98

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559-575

R: A language and Environment for Statistical Computing. (2004). Vienna

Ritchie MD, Hahn LW, Roodi N, Bailey LR, Dupont WD, Parl FF, Moore JH. (2001) Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. Am J Hum Genet 69:138-147

Shesely EG, Maeda N, Kim HS, Desai KM, Krege JH, Laubach VE, Sherman PA, Sessa WC, Smithies O. (1996) Elevated blood pressures in mice lacking endothelial nitric oxide synthase. Proc Natl Acad Sci U S A 93:13176-13181

Talman WT, Nitschke Dragon D. (2007) Neuronal nitric oxide mediates cerebral vasodilatation during acute hypertension. Brain Res 1139:126-132

Toda N, Ayajiki K, Okamura T. (2009a) Cerebral blood flow regulation by nitric oxide: recent advances. Pharmacol Rev 61:62-97

Toda N, Ayajiki K, Okamura T. (2009b) Control of systemic and pulmonary blood pressure by nitric oxide formed through neuronal nitric oxide synthase. J Hypertens 27:1929-1940

Zhou L, Zhu DY. (2009) Neuronal nitric oxide synthase: structure, subcellular localization, regulation, and clinical implications. Nitric Oxide 20:223-230

Figures

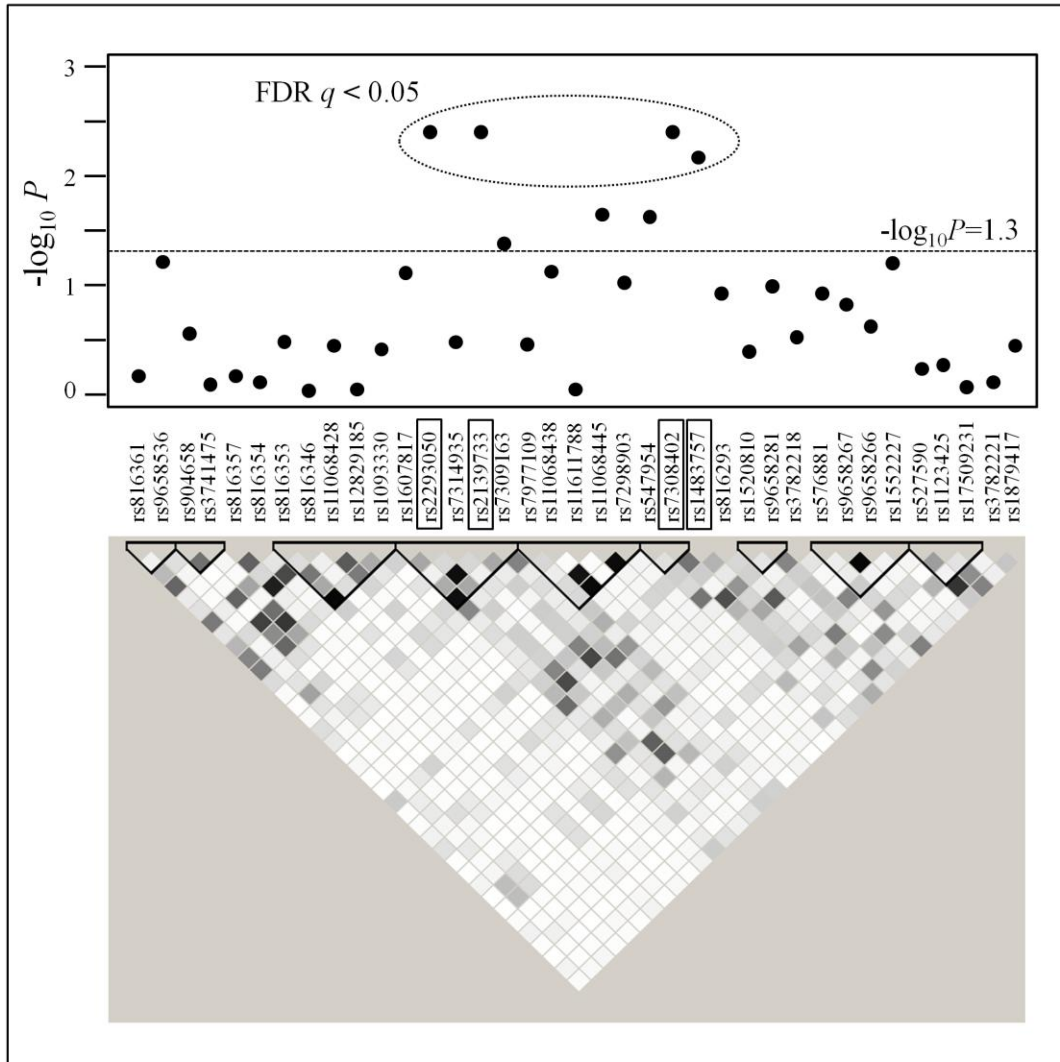

Figure 1 – *NOS1* association results ($-\log_{10} P$) with stroke risk and pairwise r^2 among genotyped SNPs in our population sample. The relative positions of the 37 *NOS1* SNPs are indicated. The white-black gradient shading represents the magnitude of linkage disequilibrium (r^2). Association results above the line $-\log_{10}P=1.3$ are considered significant ($P<0.050$); SNPs withstanding the multiple testing correction (FDR $q<0.05$) are indicated.

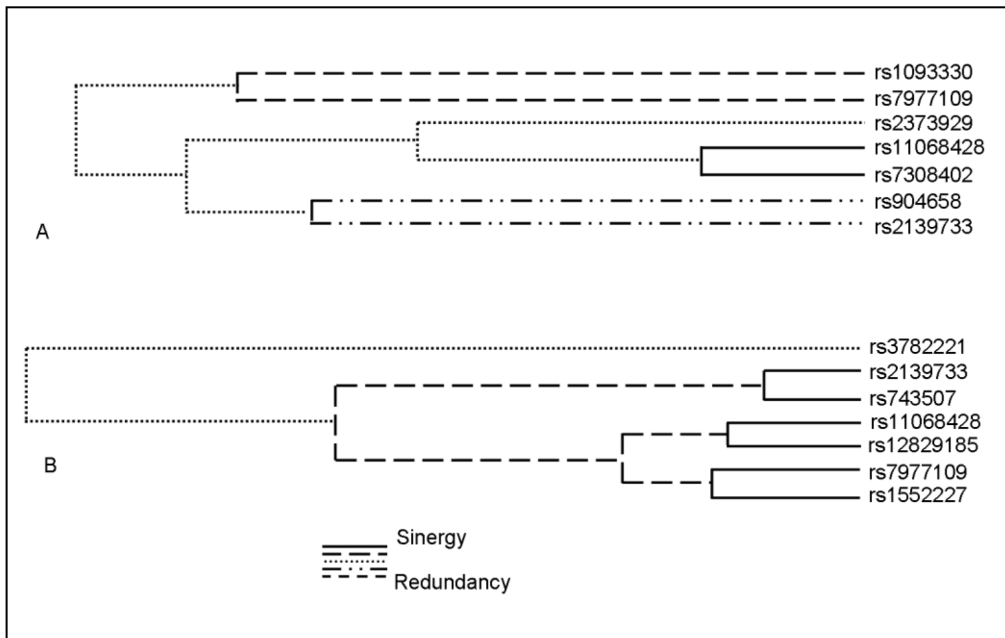

Figure 2 – Interaction dendrograms for the *NOS1* and *NOS3* polymorphisms in stroke susceptibility (A) and outcome (B). The length of the dendrogram branch that connects two polymorphisms indicates the strength of interaction (the shorter the branch, the stronger is the interaction).

Tables

Table 1 – Demographic and clinical characteristics of the population sample.

Characteristic	Controls	Patients	P^*
Age, mean±SD (yrs)	62.9±6.8	52.2±9.2	$<10^{-4}$
Gender (male), n/N (%)	247/530 (46.6)	352/551 (63.9)	$<10^{-4}$
Stroke Risk Factors, n/N (%)			
Hypertension (>85-140 mmHg)	193/513 (37.6)	278/490 (56.7)	$<10^{-4}$
Diabetes	59/501 (11.8)	89/522 (17.0)	0.017
Hypercholesterolemia (cholesterol >200 mg/dL)	328/520 (63.1)	320/511 (62.6)	0.880
Smoking	147/512 (28.7)	262/542 (48.3)	$<10^{-4}$
Drinking	218/505 (43.2)	321/544 (59.0)	$<10^{-4}$

SD – standard deviation, yrs – years.

*Mann-Whitney test or Pearson's χ^2 test.

Table 2 – Genotype frequency distribution and association with stroke susceptibility for the *NOS1* and *NOS3* SNPs.

Gene	SNP	Genotype	Genotype frequency		OR [95% CI]	P^*	FDR q †
			Controls, n (%)	Cases, n (%)			
NOS1	rs2293050						
		C/C	145 (30.1)	188 (39.8)	0.76 [0.63-0.92]	**0.004**	**0.036**
		T/C	244 (50.7)	206 (43.6)			
		T/T	92 (19.1)	78 (16.5)			
	rs2139733						
		T/T	138 (30.3)	182 (40.7)	0.76 [0.62-0.92]	**0.004**	**0.036**
		A/T	231 (50.8)	190 (42.5)			
		A/A	86 (18.9)	75 (16.8)			
	rs7309163						
		C/C	220 (47.3)	194 (42.6)	1.23 [1.01-1.51]	**0.042**	0.165
		C/T	203 (43.7)	196 (43.1)			
		T/T	42 (9.0)	65 (14.3)			
	rs11068445						
		G/G	319 (71.7)	349 (76.9)	0.72 [0.55-0.96]	**0.023**	0.110
		A/G	111 (24.9)	100 (22.0)			
		A/A	15 (3.4)	5 (1.1)			
	rs547954						
		G/G	293 (61.3)	264 (56.1)	1.29 [1.03-1.61]	**0.024**	0.110
		G/A	162 (33.9)	174 (36.9)			
		A/A	23 (4.8)	33 (7.0)			
	rs7308402						
		G/G	232 (48.2)	262 (55.5)	0.74 [0.60-0.91]	**0.004**	**0.037**
		G/A	199 (41.4)	177 (37.5)			
		A/A	50 (10.4)	33 (7.0)			
	rs1483757						
		A/A	143 (29.7)	179 (37.8)	0.77 [0.64-0.93]	**0.007**	**0.048**
		G/A	242 (50.2)	214 (45.2)			
		G/G	97 (20.1)	80 (16.9)			
NOS3	rs1800783						
		T/T	124 (25.9)	144 (31.1)	0.79 [0.65-0.95]	**0.014**	0.068
		A/T	248 (51.9)	231 (49.9)			
		A/A	106 (22.2)	88 (19.0)			
	rs2373929						
		G/G	193 (39.9)	160 (33.7)	1.24 [1.02-1.50]	**0.027**	0.068
		A/G	216 (44.6)	231 (48.6)			
		A/A	75 (15.5)	84 (17.7)			

Only associated SNPs are shown. Odds Ratio (OR) >1 indicates increased probability of having a stroke for the carriers of the minor allele. CI – Confidence Interval, FDR – false discovery rate.

*OR [95% CI] and *P* for the log-additive genetic model after adjustment for significant covariates (gender, history of hypertension, diabetes and smoking status)

† FDR *q* values

Table 3 – Haplotype frequency distribution of the *NOS1* gene and association with stroke susceptibility.

Gene	Haplotypes		Haplotype frequency	Controls (%)	Cases (%)	χ^2	***P***	FDR *q**
NOS1	rs816361-rs9658536	CA	0.092	10.4	8.0	3.967	**0.046**	0.119
	rs1607817-rs2293050-rs7314935-rs2139733-rs7309163	ACGTT	0.330	30.5	35.4	5.819	**0.016**	0.068
		CTGAC	0.305	33.5	27.6	8.635	**0.003**	**0.018**
	rs7977109-rs11068438-rs11611788-rs11068445-rs7298903	ATTGT	0.303	27.3	33.1	8.437	**0.004**	**0.032**
		AATAC	0.132	14.8	11.6	4.565	**0.033**	0.119
	rs547954-rs7308402	GA	0.283	31.6	25.1	11.408	**0.001**	**0.018**
		AG	0.235	21.1	25.8	6.558	**0.010**	**0.032**
	rs576881-rs9658267-rs9658266-rs1552227	AGGT	0.264	28.8	24.1	6.116	**0.013**	0.067

Only associated haplotypes are shown. FDR – false discovery rate.

* FDR *q* values

Table 4 – Frequency and odds ratio (OR) of the significant genotype combinations of rs1093330, rs2139733, rs7977109 and rs2373929 in stroke susceptibility. The OR-based MDR was used in the analysis.

rs1093330	rs2139733	rs7977109	rs2373929	frequency (case:control)	OR [95% CI]
TT	AA	GG	AG	1:9	0.11 [0.06-0.85]
CT	AA	AG	AG	2:10	0.20 [0.11-0.89]
CC	TT	GG	AG	5:14	0.35 [0.21-0.96]
CC	AT	AG	GG	6:16	0.37 [0.22-0.93]
CT	AT	GG	GG	6:16	0.37 [0.22-0.93]
CT	AT	AG	GG	13:30	0.42 [0.30-0.80]
CT	TT	AG	GG	21:10	2.05 [1.10-4.31]
CT	AT	AG	AG	43:20	2.10 [1.35-3.51]

CI – Confidence Interval.

Supplementary tables

Supplementary table 1 – Association analysis results for *NOS1* and *NOS3* markers and stroke susceptibility.

gene	marker	position[*]	location[*]	MA	HWE *P*	missing genotypes (%)	MAF controls	MAF patients	*P* [†]	FDR *q* [‡]
NOS1	rs816361	117655131	intronic	G	0.632	2.9	0.353	0.339	0.676	
	rs9658536	117657521	intronic	A	0.820	0.4	0.106	0.080	0.062	
	rs904658	117665544	intronic	C	0.513	3.6	0.384	0.358	0.275	
	rs3741475	117669914	3'UTR/exonic	A	0.355	2.7	0.253	0.248	0.827	
	rs816357	117682766	intronic	G	0.111	0.2	0.125	0.122	0.698	
	rs816354	117684716	intronic	G	1.000	1.4	0.106	0.106	0.766	
	rs816353	117684780	intronic	T	0.434	0.2	0.492	0.473	0.331	
	rs816346	117690874	intronic	A	0.378	0.4	0.112	0.109	0.924	
	rs11068428	117693817	intronic/exonic	A	0.563	0.1	0.344	0.321	0.360	
	rs12829185	117694020	intronic	T	0.634	0.2	0.240	0.242	0.918	
	rs1093330	117696534	intronic	C	0.250	4.3	0.497	0.479	0.388	
	rs1607817	117712186	intronic	A	1.000	0.6	0.310	0.358	0.078	
	rs2293050	117718822	intronic	T	0.661	0.6	0.448	0.372	**0.004**	**0.036**
	rs7314935	117718837	intronic	A	1.000	0.9	0.102	0.088	0.338	
	rs2139733	117726742	intronic	A	0.587	5.6	0.446	0.366	**0.004**	**0.036**
	rs7309163	117729274	intronic	T	0.533	4.1	0.309	0.366	**0.042**	0.165
	rs7977109	117730340	intronic	G	0.258	0.6	0.494	0.481	0.357	
	rs11068438	117737545	intronic	A	0.324	0.1	0.157	0.123	0.075	
	rs11611788	117738376	intronic	C	0.068	0.0	0.066	0.065	0.907	
	rs11068445	117744929	intronic	A	0.405	6.2	0.161	0.117	**0.023**	0.110
	rs7298903	117747210	intronic	C	0.307	0.3	0.151	0.118	0.095	
	rs547954	117754506	intronic	A	0.793	1.2	0.210	0.258	**0.024**	0.110
	rs7308402	117759447	intronic	A	0.614	0.6	0.316	0.251	**0.004**	**0.037**
	rs1483757	117761540	intronic	G	0.726	0.4	0.457	0.385	**0.007**	**0.048**
	rs816293	117762699	intronic	G	0.705	0.4	0.358	0.400	0.125	

Supplementary table 1 (cont.) – Association analysis results for *NOS1* and *NOS3* markers and stroke susceptibility.

gene	marker	position*	location*	MA	HWE *P*	missing genotypes (%)	MAF controls	MAF patients	*P* †	FDR *q* ‡
NOS1	rs1520810	117765189	intronic	A	0.517	0.5	0.213	0.234	0.405	
	rs9658281	117767578	intronic	T	0.582	0.3	0.272	0.238	0.102	
	rs3782218	117771511	intronic/upstream	T	0.756	0.4	0.168	0.152	0.299	
	rs576881	117772835	intronic/upstream	G	0.439	0.5	0.340	0.377	0.121	
	rs9658267	117777520	intronic	A	0.209	5.4	0.122	0.103	0.151	
	rs9658266	117777535	intronic	C	0.314	0.9	0.125	0.108	0.240	
	rs1552227	117779035	intronic	T	1.000	0.2	0.288	0.241	0.064	
	rs527590	117781918	intronic	T	1.000	0.6	0.251	0.257	0.563	
	rs1123425	117786105	intronic	G	0.862	0.0	0.464	0.449	0.559	
	rs17509231	117794323	intronic	T	0.669	0.2	0.116	0.114	0.890	
	rs3782221	117795881	intronic	A	0.914	0.5	0.279	0.280	0.791	
	rs1879417	117803515	upstream	C	0.330	0.0	0.432	0.455	0.372	
NOS3	rs12703107	150683629	upstream	T	0.547	0.4	0.236	0.252	0.086	
	rs1800783	150689397	intronic	A	0.432	1.9	0.478	0.443	**0.014**	0.068
	rs3918186	150702432	intronic	T	0.788	0.6	0.090	0.092	0.922	
	rs743507	150707488	intronic	C	0.617	0.2	0.320	0.297	0.066	
	rs2373929	150714812	downstream	A	0.515	0.0	0.372	0.421	**0.027**	0.068

FDR – false discovery rate, HWE – Hardy-Weinberg equilibrium, MA – minor allele, MAF – minor allele frequency, NS – non-significant, UTR – untranslated region.

*According to Ensembl Release 56 – September 2009.

† *P* for the log-additive genetic model after adjustment for significant covariates.

‡ FDR *q* values.

Supplementary table 2 – Association analysis results for *NOS1* and *NOS3* haplotypes and stroke susceptibility.

Gene	Haplotypes		Haplotype frequency	Controls (%)	Cases (%)	χ^2	*P*	FDR q *
NOS1	haplotype block 1: rs816361-rs9658536	CG	0.561	54.2	58.0	3.110	0.078	
		GG	0.346	35.2	34.0	0.331	0.565	
		CA	0.092	10.4	8.0	3.967	**0.046**	0.119
	haplotype block 2: rs904658-rs3741475	AG	0.627	61.5	63.9	1.358	0.244	
		CA	0.249	25.1	24.7	0.036	0.850	
		CG	0.123	13.3	11.2	2.146	0.143	
	haplotype block 3: rs816353-rs816346-rs11068428-rs12829185-rs1093330	GGGCT	0.512	50.5	52.0	0.499	0.480	
		TGATC	0.240	23.9	24.0	0.003	0.954	
		TAGCC	0.110	11.1	10.8	0.038	0.845	
		TGACC	0.090	10.2	7.9	3.179	0.075	
		TGGCC	0.042	4.0	4.5	0.303	0.582	
	haplotype block 4: rs1607817-rs2293050-rs7314935-rs2139733-rs7309163	ACGTT	0.330	30.5	35.4	5.819	**0.016**	0.068
		CTGAC	0.305	33.5	27.6	8.635	**0.003**	**0.018**
		CCGTC	0.246	23.4	25.8	1.647	0.199	
		CTAAC	0.094	10.2	8.6	1.669	0.196	
	haplotype block 5: rs7977109-rs11068438-rs11611788-rs11068445-rs7298903	GTTGT	0.490	49.8	48.2	0.541	0.462	
		ATTGT	0.303	27.3	33.1	8.437	**0.004**	**0.032**
		AATAC	0.132	14.8	11.6	4.565	**0.033**	0.119
		ATCGT	0.066	6.6	6.6	0.004	0.947	

Supplementary table 2 (cont.) – Association analysis results for *NOS1* and *NOS3* haplotypes and stroke susceptibility.

Gene	Haplotypes		Haplotype frequency	Controls (%)	Cases (%)	χ^2	*P*	FDR *q* *
NOS1	haplotype block 6: rs547954-rs7308402	GG	0.482	47.3	49.1	0.761	0.383	
		GA	0.283	31.6	25.1	11.408	**0.001**	**0.018**
		AG	0.235	21.1	25.8	6.558	**0.010**	**0.032**
	haplotype block 7: rs1520810-rs9658281	TC	0.521	51.4	52.8	0.423	0.516	
		TT	0.255	27.3	23.8	3.435	0.064	
		AC	0.224	21.4	23.5	1.339	0.247	
	haplotype block 8: rs576881-rs9658267-rs9658266-rs1552227	GGGC	0.358	33.9	37.7	3.441	0.064	
		AGGT	0.264	28.8	24.1	6.116	**0.013**	0.067
		AGGC	0.262	24.9	27.3	1.589	0.208	
		AACC	0.113	12.3	10.4	1.894	0.169	
	haplotype block 9: rs527590-rs1123425-rs17509231	CAC	0.424	41.8	43.0	0.295	0.587	
		TGC	0.249	24.7	25.1	0.044	0.835	
		CGC	0.207	21.6	19.8	1.038	0.308	
		CAT	0.115	11.5	11.4	0.001	0.974	
NOS3	haplotype block 1: rs12703107-rs1800783	GA	0.454	47.3	43.6	2.947	0.086	
		GT	0.302	29.1	31.2	1.082	0.298	
		TT	0.238	23.1	24.4	0.480	0.488	
	haplotype block 2: rs3918186-rs743507	AT	0.604	59.3	61.5	1.088	0.297	
		AC	0.305	31.7	29.3	1.466	0.226	
		TT	0.088	8.7	8.8	0.013	0.910	

* FDR *q* values. FDR – false discovery rate.

Supplementary table 3 - Demographic and clinical characteristics of ischemic stroke patients analyzed for outcome at three months

Characteristic	Good Recovery (mRS≤1)	Poor Recovery (mRS>1)	*P* *
Age and Gender			
Age, mean±SD (yrs)	50.8±9.0	52.1±8.7	0.137
Gender (male), n/N (%)	154/238 (64.7)	122/193 (63.2)	0.748
Past History, n/N (%)			
Hypertension	130/205 (63.4)	83/171 (48.5)	0.004
Diabetes	34/225 (15.1)	36/180 (20.0)	0.196
Cardiac Disease	35/226 (15.5)	35/184 (19.0)	0.344
Stroke Features, n/N (%)			
Aphasia	44/224 (19.6)	63/180 (35.0)	$5x10^{-4}$
Neglect	9/231 (3.9)	14/171 (8.2)	0.067
Dysphagia	13/235 (5.5)	19/181 (10.5)	0.060
Urinary Incontinence	2/235 (0.9)	12/182 (6.6)	0.001
Paresis	175/235 (74.5)	173/192 (90.1)	$<10^{-4}$
Altered consciousness	9/238 (3.8)	33/190 (17.4)	$<10^{-4}$
Medical complications	10/231 (4.3)	48/180 (26.7)	$<10^{-4}$
Neurologic complications	12/236 (5.1)	22/192 (11.5)	0.015

SD – standard deviation, yrs – years.

*Mann-Whitney test or Pearson's χ^2 test.

Supplementary table 4 – Association analysis results for *NOS1* and *NOS3* markers and stroke outcome.

gene	marker	position*	location*	MA	HWE *P*	missing genotypes (%)	MAF good recovery	MAF poor recovery	*P* †
NOS1	rs816361	117655131	intronic	G	1.000	2.6	0.320	0.372	0.297
	rs9658536	117657521	intronic	A	0.756	0.5	0.093	0.070	0.165
	rs904658	117665544	intronic	C	0.754	3.0	0.393	0.344	0.104
	rs3741475	117669914	3'UTR/exonic	A	0.898	1.6	0.247	0.257	0.975
	rs816357	117682766	intronic	G	0.829	0.2	0.127	0.127	0.928
	rs816354	117684716	intronic	G	0.800	1.2	0.102	0.111	0.748
	rs816353	117684780	intronic	T	1.000	0.2	0.487	0.477	0.285
	rs816346	117690874	intronic	A	1.000	0.5	0.108	0.117	0.999
	rs11068428	117693817	intronic/exonic	A	1.000	0.0	0.334	0.321	0.494
	rs12829185	117694020	intronic	T	0.602	0.2	0.233	0.260	0.398
	rs1093330	117696534	intronic	C	0.768	4.4	0.496	0.481	0.264
	rs1607817	117712186	intronic	A	0.917	0.2	0.368	0.344	0.113
	rs2293050	117718822	intronic	T	**0.049**	0.7	0.371	0.361	0.975
	rs7314935	117718837	intronic	A	1.000	1.4	0.100	0.068	0.143
	rs2139733	117726742	intronic	A	**0.032**	4.9	0.367	0.356	0.976
	rs7309163	117729274	intronic	T	0.916	2.8	0.376	0.349	0.090
	rs7977109	117730340	intronic	G	0.923	0.5	0.475	0.479	0.217
	rs11068438	117737545	intronic	A	0.203	0.2	0.127	0.135	0.823
	rs11611788	117738376	intronic	C	1.000	0.0	0.059	0.070	0.262
	rs11068445	117744929	intronic	A	0.371	3.5	0.123	0.130	0.699
	rs7298903	117747210	intronic	C	0.277	0.0	0.124	0.132	0.823
	rs547954	117754506	intronic	A	0.804	1.4	0.251	0.285	0.159
	rs7308402	117759447	intronic	A	1.000	0.7	0.243	0.233	0.955
	rs1483757	117761540	intronic	G	0.357	0.5	0.376	0.388	0.968
	rs816293	117762699	intronic	G	0.418	0.5	0.405	0.370	0.225

Supplementary table 4 (cont.) – Association analysis results for *NOS1* and *NOS3* markers and stroke outcome.

gene	marker	position*	location*	MA	HWE *P*	missing genotypes (%)	MAF good recovery	MAF poor recovery	*P* †
	rs1520810	117765189	intronic	A	0.700	0.2	0.253	0.249	0.693
	rs9658281	117767578	intronic	T	0.679	0.0	0.216	0.233	0.666
	rs3782218	117771511	intronic/upstream	T	0.319	0.5	0.122	0.164	0.250
	rs576881	117772835	intronic/upstream	G	0.146	0.7	0.382	0.356	0.949
	rs9658267	117777520	intronic	A	0.600	5.3	0.117	0.088	0.240
	rs9658266	117777535	intronic	C	0.804	1.2	0.125	0.084	0.133
	rs1552227	117779035	intronic	T	1.000	0.2	0.209	0.275	0.123
	rs527590	117781918	intronic	T	0.902	0.9	0.265	0.277	0.908
	rs1123425	117786105	intronic	G	0.628	0.0	0.468	0.451	0.195
	rs17509231	117794323	intronic	T	0.333	0.0	0.101	0.122	0.249
	rs3782221	117795881	intronic	A	0.818	0.7	0.297	0.301	0.581
	rs1879417	117803515	upstream	C	0.561	0.0	0.462	0.440	0.863
NOS3	rs12703107	150683629	upstream	T	0.801	0.7	0.276	0.236	0.246
	rs1800783	150689397	intronic	A	0.624	2.3	0.436	0.471	0.109
	rs3918186	150702432	intronic	T	1.000	0.7	0.085	0.099	0.618
	rs743507	150707488	intronic	C	0.361	0.0	0.305	0.301	0.518
	rs2373929	150714812	downstream	A	0.921	0.0	0.424	0.422	0.985

HWE – Hardy-Weinberg equilibrium, MA – minor allele, MAF – minor allele frequency, NS – non-significant., UTR – untranslated region.

*According to Ensembl Release 56 – September 2009.

†*P* for the log-additive genetic model after adjustment for significant covariates.

Supplementary table 5 – Association analysis results for *NOS1* and *NOS3* haplotypes and stroke outcome.

Gene	Haplotypes		Haplotype frequency	Good recovery (%)	Poor recovery (%)	χ^2	*P*	FDR *q* *
NOS1	haplotype block 1: rs904658-rs3741475	AG	0.628	60.8	65.2	1.767	0.184	
		CA	0.251	24.5	25.8	0.184	0.668	
		CG	0.120	14.5	9.0	6.028	**0.014**	0.107
	haplotype block 2: rs816353-rs816346-rs11068428-rs12829185-rs1093330	GGGCT	0.514	50.8	52.1	0.129	0.719	
		TGATC	0.246	23.3	26.1	0.890	0.346	
		TAGCC	0.111	10.7	11.7	0.192	0.661	
		TGACC	0.083	10.1	6.0	4.634	**0.031**	0.122
		TGGCC	0.044	4.6	4.1	0.115	0.735	
	haplotype block 3: rs1607817-rs2293050-rs7314935-rs2139733-rs7309163	ACGTT	0.355	36.3	34.4	0.312	0.576	
		CTGAC	0.273	26.2	28.7	0.660	0.417	
		CCGTC	0.267	25.3	28.4	0.999	0.318	
		CTAAC	0.083	9.6	6.6	2.586	0.108	
	haplotype block 4: rs11068438-rs11611788-rs11068445-rs7298903-rs547954	TTGTG	0.543	56.6	51.4	2.314	0.128	
		TTGTA	0.262	24.9	27.9	0.959	0.327	
		ATACG	0.126	12.2	13.2	0.201	0.654	
		TCGTG	0.064	5.9	7.0	0.442	0.506	
	haplotype block 5: rs7308402-rs1483757	GA	0.619	62.4	61.2	0.140	0.708	
		AG	0.238	24.3	23.2	0.138	0.711	
		GG	0.143	13.3	15.6	0.941	0.332	

Supplementary table 5 (cont.) – Association analysis results for *NOS1* and *NOS3* haplotypes and stroke outcome.

Gene	Haplotypes		Haplotype frequency	Good recovery (%)	Poor recovery (%)	χ^2	*P*	FDR *q* *
NOS1	haplotype block 6: rs1520810-rs9658281	TC	0.525	53.0	51.8	0.123	0.725	
		AC	0.251	25.3	24.9	0.026	0.873	
		TT	0.224	21.6	23.3	0.345	0.557	
	haplotype block 7: rs576881-rs9658267-rs9658266-rs1552227	GGGC	0.371	38.2	35.7	0.556	0.456	
		AGGC	0.284	28.7	27.9	0.076	0.782	
		AGGT	0.237	20.6	27.5	5.618	**0.018**	0.107
		AACC	0.103	11.7	8.5	2.371	0.124	
	haplotype block 8: rs527590-rs1123425-rs17509231	CAC	0.425	42.6	42.4	0.003	0.956	
		TGC	0.267	26.1	27.4	0.176	0.675	
		CGC	0.194	20.7	17.7	1.265	0.261	
		CAT	0.110	10.1	12.2	0.951	0.329	
NOS3	haplotype block 1: rs12703107-rs1800783	GA	0.441	42.3	46.4	1.494	0.222	
		GT	0.300	30.0	30.0	0.000	0.997	
		TT	0.248	26.3	23.0	1.291	0.256	
		TA	0.010	1.4	0.6	1.269	0.260	
	haplotype block 2: rs3918186-rs743507	AT	0.611	61.7	60.4	0.158	0.691	
		AC	0.298	29.8	29.7	0.000	0.982	
		TT	0.086	7.8	9.6	0.817	0.366	

FDR – false discovery rate. * FDR *q*values

Chapter 6. Genome-wide association study in stroke outcome

Manso H, Krug T, Sobral J, Albergaria I, Gaspar G, Ferro JM, Oliveira SA, Vicente AM (2011) A genome-wide association study using DNA pooling identifies *BBS9* and *GLIS3* as novel *loci* influencing patient's outcome after stroke.

Abstract

Stroke is a major cause of morbidity and mortality in developed countries, with large numbers of stroke survivors left with variable degrees of disability. Given the increased life expectancy of populations, finding adequate treatments that promote recovery of patients is a priority task, requiring the elucidation of the molecular pathways influencing brain recovery. Few studies, however, have assessed the role of genes in stroke outcome. The present report describes a pilot genome-wide association study (GWAS) of 262,264 single nucleotide polymorphisms (SNPs) to identify the genetic factors contributing to patient's outcome at three months. We compared SNP allele frequencies in a pool of non-disabled stroke patients (as assessed by the modified Rankin scale [mRS]=0), with a pool of severely disabled or deceased patients (mRS≥3), and selected the 100 most interesting SNPs for validation by individual genotyping. 36 SNPs were validated, showing significant differences between patients with extremely good and extremely poor outcome at three months at a significance level of 5% ($1.7x10^{-4}<P<0.049$). These SNPs were further analyzed in a larger sample including patients with the whole range of mRS scores, using a more clinically sensible threshold set at mRS≤1 for good recovery and mRS>1 for poor recovery. 15 SNPs were associated with stroke outcome ($4.3x10^{-4}<_{uncorrected}P<0.047$), six of which remained associated after adjusting for stroke severity parameters ($0.002<_{uncorrected}P<0.039$). Two of these SNPs, rs10273634 and rs10974334, are located within the Bardet-Biedl syndrome 9 (*BBS9*) and GLIS family zinc finger 3 (*GLIS3*) genes. In addition, we found a *GLIS3* haplotype significantly associated with stroke outcome ($_{uncorrected}P=0.004$, false discovery rate [FDR] $q=0.024$). Our results suggest that the *BBS9* and *GLIS3* genes may play a role in stroke outcome. This work represents an initial effort towards the elucidation of stroke outcome and recovery pathways using a genome-wide approach, and further supports the feasibility and accuracy of DNA pooling strategies for GWAS of complex traits.

Introduction

Besides being one of the leading causes of mortality in developed countries, secondary only to cancer and ischemic heart disease, stroke is also a main cause of disability, with a significant proportion of stroke survivors requiring institutional care and/or remaining permanently disabled [Asplund et al. 1998; Hankey et al. 2002]. Patients surviving a first-ever stroke also have an increased risk of death and/or of suffering a second event in the following years [Hankey et al. 2002; Hardie et al. 2004]. With increasing life expectancy, this public health problem tends to worsen, with significant social and economic consequences. Thus, it is essential not only to improve stroke preventive strategies, but also to find therapies that reduce stroke-induced disability and improve recovery.

A large body of data from animal studies show that genetic factors influence the severity of brain damage and the recovery process after stroke [Atochin et al. 2010; Chang et al. 2011; Hyakkoku et al. 2010; Jeffs et al. 1997]. Furthermore, family history of stroke was shown to be associated with stroke outcome, while this association was not observed for stroke severity or mortality at 90 days [Jood et al. 2005; Lisabeth et al. 2005]. Still, this field of stroke genetics remains almost unexplored. Until now, relatively few candidate genes have been tested for association with stroke outcome. The apolipoprotein E (*APOE*) gene is the most widely tested candidate, with some studies finding an association between this gene and stroke outcome, while others fail to confirm this association [Martínez-González and Sudlow 2006; McCarron et al. 2000; McCarron et al. 1998; Sarzynska-Dlugosz et al. 2007; Treger et al. 2003]. The heterogeneity between studies regarding the clinical instruments used to assess patient's outcome and the time of assessment may in part explain the discrepancies [Martínez-González and Sudlow 2006]. Other candidate genes analyzed are involved in platelet adhesion and aggregation (*PLAT* and *SERPINE1* genes), in thrombosis (*GP1BA* and *ITGB3* genes), in neurovascular injury and/or neurovascular remodeling (*MMP2* and *MMP9* genes), in metabolic pathways regulating oxidative stress in the brain (*MTHFR* and *GSTO1* genes) and in inflammatory pathways (*IL6*, *MPO*, *IL1B*, *TNF* and *PTGS2* genes) [Maguire et al. 2010; Manso et al. 2010; Manso et al. 2011; Peddareddygari et al. 2009].

Thus far, no genome-wide association study (GWAS) for stroke outcome has been reported. While GWAS strategies allow searching for variants that affect a common trait, without making any prior assumptions, the single nucleotide polymorphism (SNP) microarrays that are used are extremely expensive, rendering these studies beyond the reach of many laboratories. A suitable alternative to overcome this difficulty is to pool DNA from different individuals and to estimate the SNP allele frequencies from the DNA pools using microarrays, a strategy known as allelotyping. SNPs associated with the phenotype in an initial phase can then be confirmed by individual genotyping. Several studies have shown the feasibility, reliability and accuracy of a pooling experiment on high-density genotyping arrays to estimate allele frequencies, as compared to individual genotyping [Anantharaman and Chew 2009; Bossé et al. 2009; Docherty et al. 2007; Meaburn et al. 2006], and DNA pooling has successfully been used to identify genes associated with several traits [Brown et al. 2008; Butcher et al. 2008; Diergaarde et al. 2010; Meaburn et al. 2008; Shifman et al. 2008a; Shifman et al. 2008b]. For instance, Shifman et al [2008b] identified an intronic SNP in the reelin gene (*RELN*) that confers a sex-specific risk of schizophrenia and this association was replicated in four additional independent population samples. Brown and colleagues [2008] identified a new melanoma risk *locus* on chromosome 20 that was replicated in two other samples. The gama-glutamyltransferase 1 (*GGT1*) gene was also identified as a risk *locus* for pancreatic cancer using a pooled sample strategy [Diergaarde et al. 2010]. Results from these and other studies show that it is possible to perform a GWAS in a cost-effective way by combining the advantages of microarrays to genotype thousands of SNPs and DNA pooling.

In this study, we present a pilot GWAS with stroke outcome in pooled samples. This work was conducted in three stages: 1) pooling-based association analysis of two pools of patients classified in the extremes of a clinical outcome assessment instrument, the modified Rankin Scale (mRS); 2) validation of the pooling strategy by individual genotyping of the 100 most interesting SNPs; 3) association analysis with stroke outcome of validated SNPs in a larger sample of stroke patients using a more clinically sensible mRS cut-off for good and poor recovery.

Materials and Methods

Participants and outcome classification

414 first-ever stroke patients below 65 years of age were recruited through Neurology and Internal Medicine Departments of several hospitals in Portugal. Stroke was defined as a focal neurological deficit of sudden or rapid onset lasting more than 24 hours, and classified as ischemic based on brain imaging (computed tomography and/or magnetic resonance imaging). Stroke diagnosis was confirmed by a neurologist. Age, gender, information on previous stroke risk factors and detailed clinical data during hospitalization were collected for the majority of patients. Occurrence of aphasia, neglect, paresis, gaze paresis, dysphagia, permanent altered consciousness, urinary incontinence and medical and neurological complications during hospitalization were clinical parameters indicative of stroke severity. The demographic and clinical characteristics of this sample set are shown in Table 1.

Stroke outcome was assessed, by direct interview, using the mRS at three months after the stroke event. The mRS measures the global disability of a patient and is employed in clinical practice to evaluate the effect of stroke on the patient's daily activities and life style. Patients are classified in seven grades, with mRS=0 indicating no disability and mRS=6 indicating death. In the first step of the present study, to increase the distinction between outcomes and the homogeneity within the pooled samples and thus improve our chances for gene discovery, we constructed two pools of patients with more extreme mRS scores, mRS=0 (no disability symptoms) (N=87) and mRS≥3 (moderate to severe disability and death) (N=100). For the final joint association analysis, in a larger population set individually genotyped, we used a more clinically sensible cut-off set at mRS≤1 (no symptoms or some symptoms but able to perform all usual activities) for good recovery and mRS>1 (unable to perform all usual activities to bedridden and death) for poor recovery. With this classification, a total of 230 individuals were included in the good recovery group (64.3% males and 35.7% females), and 184 in the poor recovery group (63.0% males and 37.0% females).

Study design

The study design is delineated in Figure 1 and was carried out as follows:

a) DNA pool construction

All DNA samples had been previously genotyped with excellent results (>90% call rates) and were thus known to be of adequate quality for the present study. A DNA sample from each individual was quantified in duplicate using the NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE) and a third DNA quantification was performed if the coefficient of variation (CV) (the standard deviation divided by the mean) of the two independent measurements was >5%.

Each DNA sample was diluted to a concentration of 50ng/μL and equimolar amounts of DNA from each individual were pooled by manual pipetting. To assess variance in allele frequency estimates due to pipetting and technical errors, the two pools were created in triplicate (replicates of pool formation) and each replicate was assayed on three replicate chips (technical replicates), yielding a total of nine measurements per pool.

b) Allelotyping of pooled DNA

Pools were allelotyped using the 250K Affymetrix GeneChip® Mapping Assay – Nsp I according to the standard protocol for individual DNA samples (see the GeneChip® Mapping 500K Assay Manual for full protocol). Approximately 262,000 SNPs can be typed using this chip, with each SNP allele interrogated by 6 or 10 probes. Each microarray was washed and stained using the GeneChip® Fluidics Station 450, and was scanned using the GeneChip® Scanner 3000. For quality control checks, the manufacturer's individual reference DNA was also assayed on a separate microarray (accuracy of genotyping call = 99.7%). This sample was used as a positive control to test for assay performance.

c) Estimation of SNP-MaP allele frequency

The raw probe intensities were extracted from the CEL files using the SNPMaP package [Davis et al. 2009] of the R software [R: A language and Environment for Statistical Computing 2004] and quantile-normalized across arrays. An estimate of the allele frequency in pooled DNA was calculated for each SNP based on the relative intensities of the two alleles (relative allele signal [RAS]). The allele frequency estimates in each pool were the average frequencies across the nine measurements (3 replicates of pool formation x 3 technical replicates). The CV for each SNP was also calculated across the nine replicates. The CV is a relative measure of the variability of the data, with a CV<10% indicating low variability [Allaby 1999]. SNPs showing high variability across replicates (CV>10%), which greatly increase the likelihood of false positives [Macgregor 2007], were excluded.

d) Selection of SNPs for individual genotyping

Since there is no consensus on the best strategy for SNP selection, we chose the 100 most interesting markers based on four plausible strategies, and assessed the performance of each upon validation by individual genotyping. We initially selected the SNPs that had the largest allele frequency differences between the two pools of extremely good and poor outcome [Abraham et al. 2008] and the lowest Student's *t*-test p-values for the differences between allele frequency estimates [Baum et al. 2008; Diergaarde et al. 2010; Meaburn et al. 2008]. For these two approaches, we established somewhat arbitrary cut-offs by visually inspecting the slopes of the curves of allele frequency differences and Student's *t*-test p-values, and defining a point where this slope tended to stabilize. Additionally, we selected the SNPs that were clustered in 3 or more consecutive markers within 100kb from each other, or clustered in 3 or more consecutive markers within the same gene (according to RefSeq database), amongst the top 1,000 SNPs with larger allele frequency differences between pools and the top 1,000 SNPs with lower *t*-test p-values. If two or more SNPs were in linkage disequilibrium (LD) (r^2>0.5) with each other, only one of them was selected. A script was written in R to perform this analysis using some functions from the matrixStats and ClassComparison (OOMPA project) packages.

e) Individual genotyping and statistical analyses

Genotyping of 96 SNPs was performed using the Sequenom iPLEX assays with allele detection by mass spectroscopy, using Sequenom MassARRAY technology (Sequenom, San Diego, California) and following the manufacturer's protocol. Primer sequences were designed using Sequenom's MassARRAY Assay Design 3.0 software. The remaining 4 SNPs were genotyped using TaqMan® Pre-Designed or Custom SNP Genotyping Assays, in an ABI PRISM 7900HT Sequence Detector System (Applied Biosystems, Foster City, USA). Genotyping was performed blinded to patient's functional outcome at three months. Extensive quality control (QC) was carried out, using no-template controls, 8 HapMap individuals, duplicated samples within and across genotyping plates, and one large pedigree to check for Mendelian inconsistencies. SNPs showing genotypic inconsistencies in the HapMap and/or duplicated samples (n=3), with a genotyping call rate below 90% (n=8) and/or showing deviation from Hardy-Weinberg equilibrium (P<0.01) (n=1) were excluded. In total, 88 SNPs met QC criteria and were further analyzed.

Association analysis using logistic regression was performed for each SNP to validate the pooling results, in the initial sample set. Validated SNPs (P<0.05) were further genotyped in an additional sample (patients with mRS=1 or 2) and association analyses were carried out in the combined sample, comparing the good recovery patient group (mRS≤1) with the poor recovery patient group (mRS>1). The effect of each genetic variable on patient's outcome was determined by logistic regression, before and after adjustment for the significant covariates (history of hypertension, occurrence of aphasia, paresis, altered consciousness and medical complications during hospitalization). SNP-based association analyses were performed using the SNPassoc package of the R software [R: A language and Environment for Statistical Computing 2004]. Haplotype blocks in the GLIS family zinc finger 3 (*GLIS3*) gene were determined using the default method [Gabriel et al. 2002] of the Haploview software (v4.2) [Barrett et al. 2005] and haplotype-based association analyses were performed using the same software. To reduce the type I error, we applied the false discovery rate (FDR) multiple testing correction in individual SNP and haplotype-based association analyses. SNPs (or haplotypes) with q values≤5% were considered significant, which resulted in a FDR≤5% among the significant SNPs (or haplotypes). q values were

calculated using the qvalue package of the R software [R: A language and Environment for Statistical Computing 2004].

Results

Allelotyping and pooling validation by individual genotyping

After allelotyping on genotyping arrays, the average allele frequency estimates were calculated for each SNP and compared between the two pools of patients with very good (mRS=0) and very poor (mRS≥3) outcome at three months. Since DNA pooling introduces experimental errors in allele frequency estimations that are related to pool construction and array variation, affecting the efficiency of pooling, only SNPs with low variability (CV<10%) among replicates were included in this exploratory analysis. 46 SNPs with the largest allele frequency differences between the two pools of good and poor outcome, 34 SNPs with the lowest Student's *t*-test p-values, as well as 14 SNPs clustering within 100kb and 15 SNPs clustering within the same gene, were selected for validation. Of these, nine markers were selected by more than one strategy, and a total of 100 were individually genotyped.

The efficiency of the pooling-based association analysis was determined by the number of SNPs, among these selected 100, that were associated with stroke outcome after individual genotyping. Individual genotyping was carried out in the same subset of 187 stroke patients with extremely good or extremely poor outcome at three months (mRS=0 and mRS≥3, respectively) to validate the pooling results. 12 SNPs were excluded due to failure of quality control measures and 88 SNPs were further analyzed. Table 2 shows the association analysis results after individual genotyping of these markers. 36 SNPs (approximately 41%) showed significant differences between patients with extremely good and extremely poor outcome at three months, at a significance level of 5% ($1.7 \times 10^{-4} < P < 0.049$) (Table 2). Of these, 13 SNPs had p-values below 0.005.

We investigated whether allelotyping provides a reliable estimate of the absolute allele frequency differences between the two groups of patients with extremely good and extremely poor outcome. After comparing the allele frequency differences estimated by allelotyping and genotyping for the 36 validated SNPs, we found that the

effect size determined by allelotyping was underestimated when compared with the individual genotyping results (Wilcoxon rank sum test, W=100, $P=6.9\times10^{-10}$) (Figure 2). This means that the observed allele frequency differences between groups are significantly larger than those estimated by allelotyping, and individual genotyping is therefore needed to obtain a reliable estimate of those differences.

We next evaluated the SNP selection strategies by comparing the percentage of true-positive markers obtained for each. SNPs were considered as true-positive or validated if there was a significant difference between the two groups of patients at a significance level of 5%. Selecting SNPs according to allele frequency difference between groups and selecting consecutive SNPs showed better performances, as 56.8% and 57.1%, respectively, of the SNPs selected in this way were validated by individual genotyping. Choosing SNPs according to the *t*-test p-values showed a poor performance, with a SNP validation of 20.7%. Selecting SNPs within the same gene showed an intermediate performance, as 28.6% of SNPs were validated.

Association analysis in the combined sample

The 36 SNPs with association p-values below 0.05 were genotyped in an additional sample of 227 patients with mRS=1 or 2, and a joint association analysis was carried out in the combined sample. This test was performed to verify if the association with stroke outcome of the 36 validated SNPs still held when the cut-off for good/poor outcome was set between 1 and 2 (mRS≤1 vs. mRS>1), which is more clinically sensible. 15 out of 36 SNPs were associated with stroke outcome at three months in the combined sample ($4.3\times10^{-4}<P<0.047$) and three other markers showed a trend towards association (Table 3). Six SNPs remained associated after adjusting for significant covariates (history of hypertension, and occurrence of aphasia, paresis, altered consciousness and medical complications during hospitalization) ($0.002<P<0.039$). Two of these SNPs, rs10273634 and rs10974334, were intragenic and located in the Bardet-Biedl syndrome 9 (*BBS9* on 7p14) and GLIS family zinc finger 3 (*GLIS3* on 9p24.2) genes, respectively. Another associated SNP, rs290916, was located downstream from a novel processed transcript (RP11-428L9.1-001), for which there is little information. The other three SNPs were located in intergenic regions and far from other known *loci*: rs1243659 is 27kb from the olfactory receptor 6S1 (*OR6S1*) gene; rs7664979 is 61kb and 48kb from the hematopoietic prostaglandin D synthase

(*HPGDS*) and PDZ and LIM domain 5 (*PDLIM5*) genes, respectively; and rs9293983 is 232kb from the collagen type XII alpha 1 (*COL12A1*) gene. While the functional consequences of these intergenic polymorphisms are currently unknown, we can hypothesize that they (or other variants in LD with these markers) are located in regulatory regions (e.g. enhancers) and may influence the expression levels of distantly transcribed genes. None of these 6 associations withstood correction for multiple testing, although rs10273634 on *BBS9* had a nominal FDR *q* value (*q*=0.057).

Since rs10974334 and two other genotyped SNPs, rs7024250 and rs1000128, were located within the same gene (*GLIS3*), we performed a haplotype-based association analysis in the combined sample. One haplotype in this gene (T[rs7024250]-A[rs1000128]-G[rs10974334]) was significantly associated with stroke outcome after correcting for multiple testing (*P*=0.004, FDR *q*=0.024).

Discussion

In contrast with stroke susceptibility, few studies have assessed the role of candidate genes in stroke outcome and no GWAS has previously been reported. We carried out an association study of 262,264 SNPs with stroke outcome using, in the first stage, a DNA pooling strategy, as a cost effective way of conducting a GWAS. The efficiency of the DNA pooling strategy was determined by the number of SNPs validated by individual genotyping. In this case, the percentage of SNP validation was approximately 41%, which is considerably higher than the percentage obtained in some previous pooling studies: 16% [Butcher et al. 2008], 12% [Meaburn et al. 2008], 17% [Shifman et al. 2008a] and 31% [Shifman et al. 2008b], studies that, however, identified relevant genes subsequently replicated in independent samples. Such variation may be explained by different study designs. In this study we improved our methods to obtain a better performance. Our pools were created in triplicate and were assayed on replicate arrays, for a better control of variability in pooling construction and array technique, which strongly influence the efficiency of pooling [Macgregor 2007]. Two studies with lower percentages of SNP validation did not perform any of these replicates [Butcher et al. 2008; Meaburn et al. 2008] and the two other did not have replicates of pool construction [Shifman et al. 2008a; Shifman et al. 2008b], possibly explaining their

lower efficiency. Interestingly, Meaburn et al [2008] selected SNPs based on *t*-test p-values, which also had a poor performance in our study (12% vs. 20.7%, respectively).

For this pilot study, SNP exclusion was based on a CV>10% among replicates to reduce the chance of pursuing false positives in the following stage. This proved to be a very stringent criterion, excluding a large number of markers from further analysis. For instance, two excluded SNPs, rs10486524 and rs7857436, were located within the *BBS9* and *GLIS3* genes, respectively. Although they had a slightly higher CV (11% and 12%, respectively) than the two SNPs associated with stroke outcome, their predicted allele frequency differences between groups were similar, and should thus be individually genotyped for validation. A less stringent cut-off (e.g. CV>15%) for SNP exclusion may be advantageous, as any false positive result will be identified in the stage of individual genotyping, and the risk of missing associated markers will be decreased.

We also found that two SNP selection strategies were clearly better performers: selecting SNPs based on the allele frequency differences between groups and the clustering strategies improved the number of validated markers. The other two approaches, however, were still valuable and can complement the allele frequency differences and clustering strategies. For instance, the *GLIS3* SNP that showed the second highest evidence for association with stroke outcome in the combined sample belonged to a cluster of three consecutive markers within the same gene, but because they are far apart from each other (>100kb), they would have been missed if we did not inspect clusters within the same gene.

Six polymorphisms were associated with patient's outcome independently of stroke severity parameters. These associations did not withstand multiple testing correction, and replication of positive findings in independent datasets is required. The putative functions of the three intergenic SNPs should be investigated, as these markers may be influencing the expression levels of distantly located genes and consequently of patient's outcome after stroke. Further studies are also needed to investigate the role of the new processed transcript RP11-428L9.1-001 and its potential relation to stroke outcome.

The *BBS9* gene encodes different isoforms of the PTHB1 protein, which are expressed in a variety of tissues, including the brain [Adams et al. 1999]. Mutations in this gene were identified in patients with Bardet-Biedl syndrome (BBS) (MIM ID: 209900) [Nishimura et al. 2005], a pleiotropic genetic disease that is associated with increased susceptibility to hypertension, diabetes and congenital heart disease [Elbedour

et al. 1994; Green et al. 1989; Harnett et al. 1988]. Obesity is one of the major clinical manifestations of BBS. Interestingly, it was observed that mice maintained in dietary energy restriction had smaller infarct volumes and less neurological impairment after stroke, which suggests that excessive energy intake or obesity can negatively influence stroke outcome [Arumugam et al. 2010].

Different isoforms of the zinc finger protein GLIS3 are encoded by alternative splicing of the *GLIS3* gene. GLIS3 is a transcription factor that contains five C_2H_2-type Krüppel like zinc finger motifs, and can act as a transcriptional activator and repressor [Kim et al. 2003]. This protein is expressed in a variety of tissues, including the brain [Kim et al. 2003]. Mutations within *GLIS3* cause a neonatal diabetes syndrome associated with several other conditions, like congenital hypothyroidism and hepatic fibrosis [Senée et al. 2006]. In addition, polymorphisms within *GLIS3* have been associated with type 1 diabetes [Barrett et al. 2009], and with glycemic traits and type 2 diabetes [Dupuis et al. 2010]; and it was observed that diabetes is associated with severe disability after stroke [Kaarisalo et al. 2005].

This work was an initial effort towards the elucidation of the biochemical pathways involved in stroke outcome and recovery, with the advantage of a genome-wide approach without setting an *a priori* hypothesis. Our results highlight two unexpected genes, as frequently happens in GWAS, and further studies are required to validate this hypothesis and to understand their connection to stroke-induced disability and/or stroke recovery processes. Namely, an association analysis needs to be conducted with haplotype tagging SNPs fully covering the genetic variability within these genes, and the results need to be replicated in independent population samples, which are currently being recruited by several research groups. As with most complex diseases and traits, large samples will be required for adequate power to detect low to moderate size effects or rare variants and, as such, pooling of resources and population samples is advised. The study results also further supported the feasibility of DNA pooling as a cost-effective strategy for GWAS, as previously shown by other studies of complex traits and, as a pilot study, it highlighted potential design improvements, which will be pursued in the future.

References

Abraham R, Moskvina V, Sims R, Hollingworth P, Morgan A, Georgieva L, Dowzell K, Cichon S, Hillmer AM, O'Donovan MC, Williams J, Owen MJ, Kirov G. (2008) A genome-wide association study for late-onset Alzheimer's disease using DNA pooling. BMC Med Genomics 1:44

Adams AE, Rosenblatt M, Suva LJ. (1999) Identification of a novel parathyroid hormone-responsive gene in human osteoblastic cells. Bone 24:305-313

Allaby M. (1999) Dictionary of Zoology. USA: Oxford University Press

Anantharaman R, Chew FT. (2009) Validation of pooled genotyping on the Affymetrix 500 k and SNP6.0 genotyping platforms using the polynomial-based probe-specific correction. BMC Genet 10:82

Arumugam TV, Phillips TM, Cheng A, Morrell CH, Mattson MP, Wan R. (2010) Age and energy intake interact to modify cell stress pathways and stroke outcome. Ann Neurol 67:41-52

Asplund K, Stegmayr B, Peltonen M. (1998) From the twentieth to the twenty-first century: a public health perspective on stroke. In: *Cerebrovascular Disease Pathophysiology, Diagnosis, and Management.* (Ginsberg MD, Bogousslavsky J, eds), Oxford, UK: Blackwell Science, pp 901-918

Atochin DN, Yuzawa I, Li Q, Rauwerdink KM, Malhotra R, Chang J, Brouckaert P, Ayata C, Moskowitz MA, Bloch KD, Huang PL, Buys ES. (2010) Soluble guanylate cyclase alpha1beta1 limits stroke size and attenuates neurological injury. Stroke 41:1815-1819

Barrett JC, Clayton DG, Concannon P, Akolkar B, Cooper JD, Erlich HA, Julier C, Morahan G, Nerup J, Nierras C, Plagnol V, Pociot F, Schuilenburg H, Smyth DJ, Stevens H, Todd JA, Walker NM, Rich SS, Consortium TDG. (2009) Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. Nat Genet 41:703-707

Barrett JC, Fry B, Maller J, Daly MJ. (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21:263-265

Baum AE, Akula N, Cabanero M, Cardona I, Corona W, Klemens B, Schulze TG, Cichon S, Rietschel M, Nöthen MM, Georgi A, Schumacher J, Schwarz M, Abou Jamra R, Höfels S, Propping P, Satagopan J, Detera-Wadleigh SD, Hardy J, McMahon FJ. (2008) A genome-wide association study implicates diacylglycerol kinase eta (DGKH) and several other genes in the etiology of bipolar disorder. Mol Psychiatry 13:197-207

Bossé Y, Bacot F, Montpetit A, Rung J, Qu HQ, Engert JC, Polychronakos C, Hudson TJ, Froguel P, Sladek R, Desrosiers M. (2009) Identification of susceptibility genes for complex diseases using pooling-based genome-wide association scans. Hum Genet 125:305-318

Brown KM, Macgregor S, Montgomery GW, Craig DW, Zhao ZZ, Iyadurai K, Henders AK, Homer N, Campbell MJ, Stark M, Thomas S, Schmid H, Holland EA, Gillanders EM, Duffy DL, Maskiell JA, Jetann J, Ferguson M, Stephan DA, Cust AE, Whiteman D, Green A, Olsson H, Puig S, Ghiorzo P, Hansson J, Demenais F, Goldstein AM, Gruis NA, Elder DE, Bishop JN, Kefford RF, Giles GG, Armstrong BK, Aitken JF, Hopper JL, Martin NG, Trent JM, Mann GJ, Hayward

NK. (2008) Common sequence variants on 20q11.22 confer melanoma susceptibility. Nat Genet 40:838-840

Butcher LM, Davis OS, Craig IW, Plomin R. (2008) Genome-wide quantitative trait locus association scan of general cognitive ability using pooled DNA and 500K single nucleotide polymorphism microarrays. Genes Brain Behav 7:435-446

Chang CF, Chen SF, Lee TS, Lee HF, Shyue SK. (2011) Caveolin-1 deletion reduces early brain injury after experimental intracerebral hemorrhage. Am J Pathol 178:1749-1761

Davis OS, Plomin R, Schalkwyk LC. (2009) The SNPMaP package for R: a framework for genome-wide association using DNA pooling on microarrays. Bioinformatics 25:281-283

Diergaarde B, Brand R, Lamb J, Cheong SY, Stello K, Barmada MM, Feingold E, Whitcomb DC. (2010) Pooling-based genome-wide association study implicates gamma-glutamyltransferase 1 (GGT1) gene in pancreatic carcinogenesis. Pancreatology 10:194-200

Docherty SJ, Butcher LM, Schalkwyk LC, Plomin R. (2007) Applicability of DNA pools on 500 K SNP microarrays for cost-effective initial screens in genomewide association studies. BMC Genomics 8:214

Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, Wheeler E, Glazer NL, Bouatia-Naji N, Gloyn AL, Lindgren CM, Mägi R, Morris AP, Randall J, Johnson T, Elliott P, Rybin D, Thorleifsson G, Steinthorsdottir V, Henneman P, Grallert H, Dehghan A, Hottenga JJ, Franklin CS, Navarro P, Song K, Goel A, Perry JR, Egan JM, Lajunen T, Grarup N, Sparsø T, Doney A, Voight BF, Stringham HM, Li M, Kanoni S, Shrader P, Cavalcanti-Proença C, Kumari M, Qi L, Timpson NJ, Gieger C, Zabena C, Rocheleau G, Ingelsson E, An P, O'Connell J, Luan J, Elliott A, McCarroll SA, Payne F, Roccasecca RM, Pattou F, Sethupathy P, Ardlie K, Ariyurek Y, Balkau B, Barter P, Beilby JP, Ben-Shlomo Y, Benediktsson R, Bennett AJ, Bergmann S, Bochud M, Boerwinkle E, Bonnefond A, Bonnycastle LL, Borch-Johnsen K, Böttcher Y, Brunner E, Bumpstead SJ, Charpentier G, Chen YD, Chines P, Clarke R, Coin LJ, Cooper MN, Cornelis M, Crawford G, Crisponi L, Day IN, de Geus EJ, Delplanque J, Dina C, Erdos MR, Fedson AC, Fischer-Rosinsky A, Forouhi NG, Fox CS, Frants R, Franzosi MG, Galan P, Goodarzi MO, Graessler J, Groves CJ, Grundy S, Gwilliam R, Gyllensten U, Hadjadj S, et al. (2010) New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet 42:105-116

Elbedour K, Zucker N, Zalzstein E, Barki Y, Carmi R. (1994) Cardiac abnormalities in the Bardet-Biedl syndrome: echocardiographic studies of 22 patients. Am J Med Genet 52:164-169

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D. (2002) The structure of haplotype blocks in the human genome. Science 296:2225-2229

Green JS, Parfrey PS, Harnett JD, Farid NR, Cramer BC, Johnson G, Heath O, McManamon PJ, O'Leary E, Pryse-Phillips W. (1989) The cardinal manifestations of Bardet-Biedl syndrome, a form of Laurence-Moon-Biedl syndrome. N Engl J Med 321:1002-1009

Hankey GJ, Jamrozik K, Broadhurst RJ, Forbes S, Anderson CS. (2002) Long-term disability after first-ever stroke and related prognostic factors in the Perth Community Stroke Study, 1989-1990. Stroke 33:1034-1040

Hardie K, Hankey GJ, Jamrozik K, Broadhurst RJ, Anderson C. (2004) Ten-year risk of first recurrent stroke and disability after first-ever stroke in the Perth Community Stroke Study. Stroke 35:731-735

Harnett JD, Green JS, Cramer BC, Johnson G, Chafe L, McManamon P, Farid NR, Pryse-Phillips W, Parfrey PS. (1988) The spectrum of renal disease in Laurence-Moon-Biedl syndrome. N Engl J Med 319:615-618

Hyakkoku K, Hamanaka J, Tsuruma K, Shimazawa M, Tanaka H, Uematsu S, Akira S, Inagaki N, Nagai H, Hara H. (2010) Toll-like receptor 4 (TLR4), but not TLR3 or TLR9, knock-out mice have neuroprotective effects against focal cerebral ischemia. Neuroscience 171:258-267

Jeffs B, Clark JS, Anderson NH, Gratton J, Brosnan MJ, Gauguier D, Reid JL, Macrae IM, Dominiczak AF. (1997) Sensitivity to cerebral ischaemic insult in a rat model of stroke is determined by a single genetic locus. Nat Genet 16:364-367

Jood K, Ladenvall C, Rosengren A, Blomstrand C, Jern C. (2005) Family History in Ischemic Stroke Before 70 Years of Age. The Sahlgrenska Academy Study on Ischemic Stroke. Stroke 36:1383-1387

Kaarisalo MM, Räihä I, Sivenius J, Immonen-Räihä P, Lehtonen A, Sarti C, Mähönen M, Torppa J, Tuomilehto J, Salomaa V. (2005) Diabetes worsens the outcome of acute ischemic stroke. Diabetes Res Clin Pract 69:293-298

Kim YS, Nakanishi G, Lewandoski M, Jetten AM. (2003) GLIS3, a novel member of the GLIS subfamily of Krüppel-like zinc finger proteins with repressor and activation functions. Nucleic Acids Res 31:5513-5525

Lisabeth LD, Smith MA, Brown DL, Uchino K, Morgenstern LB. (2005) Family history and stroke outcome in a bi-ethnic, population-based stroke surveillance study. BMC Neurol 5:20

Macgregor S. (2007) Most pooling variation in array-based DNA pooling is attributable to array error rather than pool construction error. Eur J Hum Genet 15:501-504

Maguire J, Thakkinstian A, Levi C, Lincz L, Bisset L, Sturm J, Scott R, Whyte S, Attia J. (2010) Impact of COX-2 rs5275 and rs20417 and GPIIIa rs5918 Polymorphisms on 90-Day Ischemic Stroke Functional Outcome: A Novel Finding. J Stroke Cerebrovasc Dis

Manso H, Krug T, Sobral J, Albergaria I, Gaspar G, Ferro JM, Oliveira SA, Vicente AM. (2010) Variants of the Matrix Metalloproteinase-2 but not the Matrix Metalloproteinase-9 genes significantly influence functional outcome after stroke. BMC Med Genet 11:40

Manso H, Krug T, Sobral J, Albergaria I, Gaspar G, Ferro JM, Oliveira SA, Vicente AM. (2011) Variants in the inflammatory *IL6* and *MPO* genes modulate stroke susceptibility through main effects and gene-gene interactions. J Cereb Blood Flow Metab (*in press*)

Martínez-González NA, Sudlow CL. (2006) Effects of apolipoprotein E genotype on outcome after ischaemic stroke, intracerebral haemorrhage and subarachnoid haemorrhage. J Neurol Neurosurg Psychiatry 77:1329-1335

McCarron MO, Muir KW, Nicoll JA, Stewart J, Currie Y, Brown K, Bone I. (2000) Prospective study of apolipoprotein E genotype and functional outcome following ischemic stroke. Arch Neurol 57:1480-1484

McCarron MO, Muir KW, Weir CJ, Dyker AG, Bone I, Nicoll JA, Lees KR. (1998) The apolipoprotein E epsilon4 allele and outcome in cerebrovascular disease. Stroke 29:1882-1887

Meaburn E, Butcher LM, Schalkwyk LC, Plomin R. (2006) Genotyping pooled DNA using 100K SNP microarrays: a step towards genomewide association scans. Nucleic Acids Res 34:e27

Meaburn EL, Harlaar N, Craig IW, Schalkwyk LC, Plomin R. (2008) Quantitative trait locus association scan of early reading disability and ability using pooled DNA and 100K SNP microarrays in a sample of 5760 children. Mol Psychiatry 13:729-740

Nishimura DY, Swiderski RE, Searby CC, Berg EM, Ferguson AL, Hennekam R, Merin S, Weleber RG, Biesecker LG, Stone EM, Sheffield VC. (2005) Comparative genomics and gene expression analysis identifies BBS9, a new Bardet-Biedl syndrome gene. Am J Hum Genet 77:1021-1033

Peddareddygari LR, Dutra AV, Levenstien MA, Sen S, Grewal RP. (2009) An analysis of methylenetetrahydrofolate reductase and glutathione S-transferase omega-1 genes as modifiers of the cerebral response to ischemia. BMC Neurol 9:37

R: A language and Environment for Statistical Computing. (2004). Vienna

Sarzynska-Dlugosz I, Gromadzka G, Baranska-Gieruszczak M, Ciesielska A, Czlonkowska A. (2007) APOE does not predict poor outcome 1 year after ischemic stroke. Neurol Res 29:64-69

Senée V, Chelala C, Duchatelet S, Feng D, Blanc H, Cossec JC, Charon C, Nicolino M, Boileau P, Cavener DR, Bougnères P, Taha D, Julier C. (2006) Mutations in GLIS3 are responsible for a rare syndrome with neonatal diabetes mellitus and congenital hypothyroidism. Nat Genet 38:682-687

Shifman S, Bhomra A, Smiley S, Wray NR, James MR, Martin NG, Hettema JM, An SS, Neale MC, van den Oord EJ, Kendler KS, Chen X, Boomsma DI, Middeldorp CM, Hottenga JJ, Slagboom PE, Flint J. (2008a) A whole genome association study of neuroticism using DNA pooling. Mol Psychiatry 13:302-312

Shifman S, Johannesson M, Bronstein M, Chen SX, Collier DA, Craddock NJ, Kendler KS, Li T, O'Donovan M, O'Neill FA, Owen MJ, Walsh D, Weinberger DR, Sun C, Flint J, Darvasi A. (2008b) Genome-wide association identifies a common variant in the reelin gene that increases the risk of schizophrenia only in women. PLoS Genet 4:e28

Treger I, Froom P, Ring H, Friedman G. (2003) Association between apolipoprotein E4 and rehabilitation outcome in hospitalized ischemic stroke patients. Arch Phys Med Rehabil 84:973-976

Figures

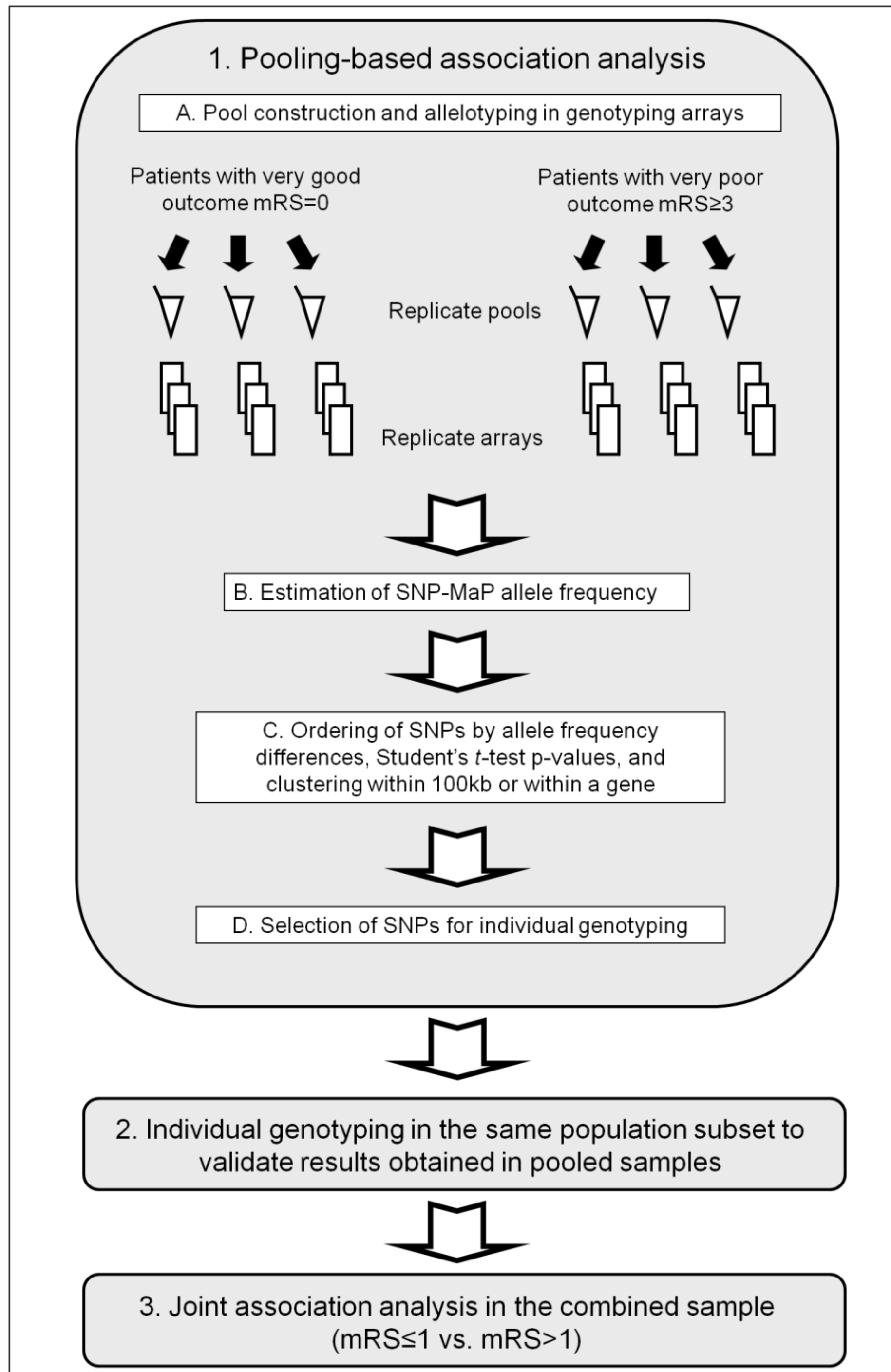

Figure 1 – Study design of the present study. The mRS at three months was used to create a "very good recovery" and "very poor recovery" group of patients. Specifically, the "very good" and "very poor recovery" groups included patients with no symptoms (mRS=0) (N=87) and patients with moderate to severe disability or deceased (mRS≥3) (N=100), respectively. DNA samples from these individuals were pooled together in the corresponding groups followed by allelotyping on genotyping arrays. SNPs were prioritized for individual genotyping to validate our pooling results. Validated SNPs, i.e. showing differences between the two groups ($P<0.05$), were genotyped in an additional sample of patients with

mRS=1 and 2 (N=227), followed by joint association analysis with stroke outcome in the combined sample. For the joint association analysis, the cut-off for good/poor recovery was set between 1 and 2, because it is more clinically sensible: patients with mRS≤1 were assigned to the "good recovery" group and patients with mRS>1 were assigned to the "poor recovery" group.

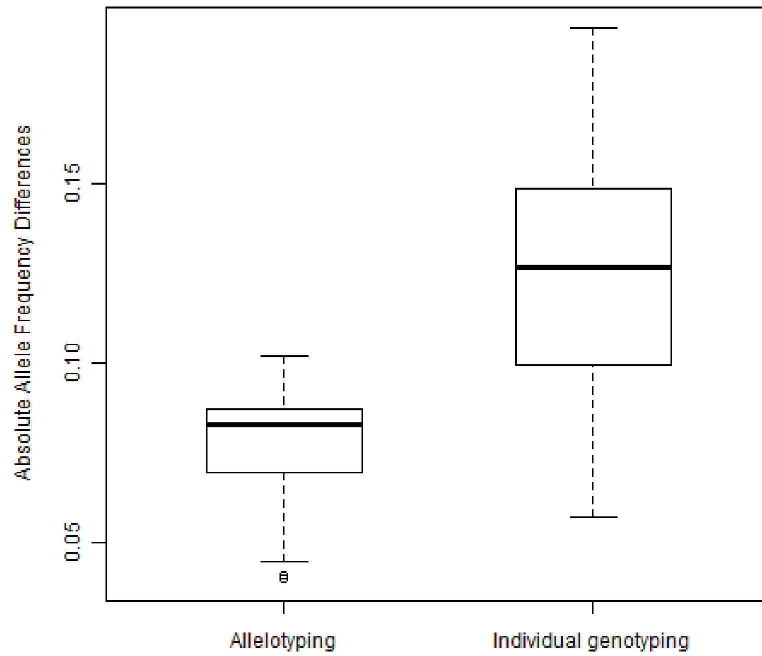

Figure 2 – Boxplots showing the distribution of the allele frequency differences between patients with very good (mRS=0) and very poor (mRS≥3) outcome at three months that were estimated by allelotyping and by individual genotyping. This figure is derived from the results of the 36 validated SNPs.

Tables

Table 1 – Demographic and clinical characteristics of ischemic stroke patients analyzed for outcome at three months

Characteristic	Good Recovery (mRS≤1)	Poor Recovery (mRS>1)	P^*
Age and Gender			
Age, mean±SD (yrs)	50.7±9.0	52.1±8.9	0.125
Gender (male), n/N (%)	148/230 (64.3)	116/184 (63.0)	0.784
Past History, n/N (%)			
Hypertension	127/198 (64.1)	77/163 (47.2)	0.001
Diabetes	34/217 (15.7)	35/173 (20.2)	0.241
Cardiac Disease	33/218 (15.1)	33/175 (18.9)	0.327
Stroke Features, n/N (%)			
Aphasia	43/217 (19.8)	56/173 (32.4)	0.005
Neglect	9/223 (4.0)	14/165 (8.5)	0.067
Dysphagia	13/227 (5.7)	19/175 (10.9)	0.060
Urinary Incontinence	2/227 (0.9)	12/175 (6.9)	0.001
Paresis	168/227 (74.0)	165/184 (89.7)	$<10^{-4}$
Altered consciousness	8/230 (3.5)	31/181 (17.1)	$<10^{-4}$
Medical complications	10/223 (4.5)	43/171 (25.1)	$<10^{-4}$
Neurologic complications	12/228 (5.3)	19/183 (10.4)	0.051

SD – standard deviation, yrs – years.

*Mann-Whitney test or Pearson's χ^2 test.

Table 2 – Results of individual genotyping for the SNPs selected in the pooling stage.

SNP ID	Chr	position	gene	Rank based on allelotyping		selection strategies*	SNP type	allele	validation sample		
				difference	t-test				freq Good Recov (mRS=0)	freq Poor Recov (mRS≥3)	*P*†
rs12751992	1	63246918	*LOC199897*	12	36	1	A/G	G	41.3	54.7	**0.011**
rs705547	1	64503365		22	624	1	A/G	G	45.9	62.8	**0.002**
rs17300340	1	178588341	*ACBD6*	21	78	1	C/T	C	80.2	89.4	**0.026**
rs1453766	2	5309265		>1000	35	2	G/T	G	56.4	65.8	0.069
rs6761743	2	84331467		>1000	18	2	C/T	C	72.4	75.5	0.493
rs4673324	2	205926709	*PARD3B*	27	148	1	A/G	A	75.3	68.1	0.148
rs2600753	2	237214589		>1000	30	2	C/T	C	69.2	74.5	0.270
rs7425755	2	240352740	*LOC150935*	>1000	4	2	C/T	C	66.9	71.1	0.399
rs6793158	3	55210273		>1000	25	2	A/G	G	56.4	59.0	0.626
rs811322	3	139812552	*FAIM*	13	99	1	A/G	A	63.1	46.8	**0.003**
rs7639507	3	156171973		>1000	10	2	A/G	G	69.2	75.8	0.156
rs4469109	4	61049121		>1000	9	2	A/G	G	92.4	89.4	0.284‡
rs7664979	4	95544420		>1000	22	2	C/T	C	58.4	71.3	**0.012**
rs12508742	4	134733668		124	>1000	3	G/T	T	66.9	55.9	**0.041**
rs2587163	4	186457029	*SNX25*	23	327	1	C/G	G	65.1	48.9	**0.001**
rs10060876	5	2611690		34	>1000	1	A/G	A	91.9	85.3	0.051
rs1150462	5	5964422		9	>1000	1	C/T	T	68.8	61.2	0.137
rs1328254	5	49596616		>1000	31	2	C/T	T	59.3	57.4	0.694
rs17627020	5	54203597		42	>1000	1	C/G	G	99.4	93.7	**0.010**
rs4958803	5	154774673		>1000	19	2	C/G	G	51.7	54.8	0.558

rs10050717	5	164099184		32	371	1	A/T	A	62.9	51.1	**0.035**
rs10071045	5	164779960		4	183	1	A/T	T	45.9	55.9	0.074
rs415223	6	4513282		>1000	26	2	C/T	T	79.1	73.7	0.240
rs423137	6	42768633		>1000	11	2	C/T	C	76.7	72.1	0.351
rs10484654	6	53990696	*C6orf142*	>1000	13	2	A/C	C	64.0	57.6	0.238
rs9293983	6	75618311		8	736	1	A/G	G	54.1	68.6	**0.004**
rs7769736	6	143879125		>1000	1	2	C/G	C	57.5	56.8	0.892
rs7811384	7	3568874	*SDK1*	26	937	1	A/G	A	66.3	53.7	**0.014**
rs17792244	7	9200867		16	>1000	1	A/G	G	45.3	54.3	0.108
rs10273634	7	33318154	*BBS9*	46	947	1	C/T	C	85.5	76.8	**0.027**
rs2598044	7	37856841	*TXNDC3*	5	>1000	1	A/G	G	59.9	77.7	**3.7x10^{-4}**
rs4588727	7	63859252		>1000	6	2	G/T	T	80.8	69.5	**0.013**
rs2928672	8	23466270	*SLC25A37*	>1000	14	2	A/G	G	66.3	58.4	0.100
rs7831651	8	51859645	*SNTG1*	>1000	27	2	G/T	G	67.4	70.2	0.563
rs11786647	8	57927246		>1000	28	2	C/G	G	73.6	67.4	0.201
rs2941422	8	76682389		31	227	1	C/T	T	66.5	53.7	**0.015**
rs1463186	8	93669646		152	949	3	C/T	C	50.0	57.4	0.123
rs1449243	8	93676222		448	>1000	3	G/T	T	72.7	69.5	0.462
rs11562770	8	119086110	*EXT1*	40	>1000	1	A/G	A	70.9	51.6	**1.7x10^{-4}**
rs17646599	8	137757137		37	>1000	1	A/G	A	86.6	80.5	0.131
rs7024250	9	3823480	*GLIS3*	29	>1000	1,4	G/T	T	95.3	89.5	**0.026**[‡]
rs1000128	9	4017185	*GLIS3*	141	>1000	4	A/G	A	73.8	64.7	**0.049**
rs10974334	9	4092339	*GLIS3*	288	276	4	G/T	T	65.7	81.1	**5.1x10^{-4}**
rs1576657	9	12429164		479	33	2	C/G	C	57.7	51.6	0.208
rs947403	10	1006601		39	580	1	C/T	T	52.4	41.5	**0.037**
rs7476580	10	1366089	*ADARB2*	30	>1000	1	C/T	C	61.2	70.0	0.088

rs290916	10	9000349		19	>1000	1	C/T	C	65.9	81.6	**4.0x10^{-4}**
rs1904006	10	53479033	*PRKG1*	18	453	1	C/T	T	83.1	74.2	**0.039**
rs7097525	10	123052530		253	16	2	C/T	C	55.8	60.6	0.358
rs12364058	11	18191988		>1000	24	2	C/T	T	56.5	50.5	0.266
rs17824184	11	76617563	*GDPD4*	6	603	1	A/G	A	80.2	68.4	**0.013**
rs10895266	11	101533105	*YAP1*	20	761	1	A/G	A	62.9	54.2	0.114
rs2513605	11	105823864		>1000	5	2	C/T	C	46.5	53.2	0.194
rs10502193	11	114169861		14	>1000	1	A/G	A	90.7	83.7	0.099
rs6590261	11	126806806		292	>1000	3	A/G	G	43.0	58.4	**0.003**
rs7947171	11	126838598		563	7	2,3	A/T	T	57.0	70.7	**0.003**
rs17187	11	126897371		474	>1000	3	C/G	G	88.4	75.8	**0.001**
rs11223888	11	133908599		35	>1000	1	C/T	C	44.2	51.6	0.156
rs12830815	12	52483533		1	885	1	A/G	G	52.9	68.1	**0.005**
rs10506321	12	52510591		171	>1000	3	A/G	G	52.3	66.3	**0.009**
rs12822967	12	64948897		11	232	1	A/G	A	63.4	55.3	0.115
rs10506806	12	77979113		25	>1000	1	C/T	C	75.6	68.0	0.109
rs7131780	12	101722143		7	83	1	C/G	G	79.1	65.4	**0.004**
rs10778623	12	107244706	*CMKLR1*	301	>1000	3	C/T	C	57.0	63.7	0.180
rs9532718	13	40542722	*WBP4*	>1000	2	2	A/C	A	57.6	47.3	**0.044**
rs7985623	13	59220111	*DIAPH3*	44	549	1	C/G	G	66.9	54.2	**0.019**
rs1243659	14	20151347		>1000	34	2	C/T	T	77.9	86.7	**0.026**
rs1295826	14	69186662	*KIAA0247*	>1000	8	2	G/T	G	83.1	92.0	**0.012**
rs7145567	14	77428041	*ADCK1*	38	240	1	A/G	A	65.7	53.8	**0.031**
rs2035801	15	96162993		>1000	23	2	A/G	A	62.9	61.1	0.719
rs1470984	15	99515982		>1000	20	2	C/T	C	57.6	50.5	0.154
rs4619425	16	7091509	*A2BP1*	99	>1000	4	C/G	G	82.6	75.8	0.112

rs11077149	16	7204783	*A2BP1*	241	>1000	4	C/T	C	52.9	56.5	0.535
rs8050137	16	7391457	*A2BP1*	387	>1000	4	C/T	T	85.5	81.6	0.329
rs3027232	17	7962790	*ALOXE3*	>1000	32	2	A/G	G	71.5	81.1	**0.037**
rs2970016	17	47531938	*CA10 CA10*	3	29	1,2	C/G	G	45.3	53.7	0.101
rs12962942	18	10275603		28	59	1	C/T	C	59.8	67.4	0.158
rs16942806	18	22667875		2	>1000	1	A/G	A	89.0	84.2	0.182
rs7258452	19	62139806		45	>1000	1	C/T	T	73.3	65.3	0.115
rs6108882	20	11187261		>1000	3	2	A/G	G	52.3	60.5	0.135
rs3761896	20	12937901	*SPTLC3*	>1000	217	3, 4	A/G	A	86.0	80.0	0.108
rs2073302	20	13000671	*SPTLC3*	416	665	3,4	C/T	C	56.4	69.5	**0.009**
rs1041327	20	14798101	*MACROD2*	223	397	4	A/G	G	84.3	86.8	0.512
rs6131684	20	15572156	*MACROD2*	43	177	1, 4	A/G	G	44.8	53.2	0.095
rs175793	20	15808386	*MACROD2*	695	>1000	4	C/T	C	87.2	84.2	0.418
rs467155	21	30227845	*GRIK1*	17	>1000	1	C/T	T	91.4	86.8	0.069
rs220161	21	42422362	*UMODL1*	725	>1000	3, 4	C/G	C	87.8	80.5	0.064
rs220162	21	42422585	*UMODL1*	144	740	3, 4	G/T	G	62.8	70.7	0.119

Chr – Chromosome, freq Good Recov – allele frequency (%) in patients with good outcome, freq Poor Recov – allele frequency (%) in patients with poor outcome
*Selection strategies: 1 – allele frequency difference, 2 – Student's *t*-test p-values, 3 – clustering within 100kb, 4 – clustering within a gene.
†*P* for the log-additive model.
‡*P* for the codominant model.

Table 3 – Association results in the combined sample for the 36 validated SNPs.

SNP ID	Chr	position	gene	freq Good Recov (mRS≤1)	freq Poor Recov (mRS>1)	$_{unadjusted}P^{*}$	$_{adjusted}P^{\dagger}$	OR [95% CI]†	FDR $q^{\ddagger}$
rs11562770	8	119086110	*EXT1*	60.0	54.6	0.112			
rs2598044	7	37856841	*TXNDC3*	69.1	74.9	0.077			
rs290916	10	9000349		68.5	79.2	**$4.3x10^{-4}$**	**0.016**	0.59 [0.39-0.92]	0.150
rs10974334	9	4092339	*GLIS3*	69.6	79.7	**$7.4x10^{-4}$**	**0.038**	0.64 [0.42-0.98]	0.155
rs2587163	4	186457029	*SNX25*	58.1	48.9	**0.008**	0.337		
rs17187	11	126897371		81.2	76.0	0.066			
rs705547	1	64503365		51.3	55.5	0.241			
rs7947171	11	126838598		65.1	68.2	0.349			
rs811322	3	139812552	*FAIM*	56.9	49.7	**0.047**	0.545		
rs6590261	11	126806806		51.1	58.2	**0.040**	0.703		
rs9293983	6	75618311		58.4	65.2	**0.047**	**0.039**	0.68 [0.46-0.98]	0.155
rs7131780	12	101722143		76.1	69.1	**0.028**	0.070		
rs12830815	12	52483533		57.9	63.0	0.149			
rs10506321	12	52510591		56.8	61.0	0.232			
rs2073302	20	13000671	*SPTLC3*	58.3	62.8	0.193			
rs17627020	5	54203597		98.7	95.1	**0.005**	0.090		
rs12751992	1	63246918	*LOC199897*	50.7	52.2	0.661			
rs7664979	4	95544420		59.8	67.4	**0.027**	**0.025**	0.65 [0.45-0.95]	0.155
rs1295826	14	69186662	*KIAA0247*	86.2	91.3	**0.026**	0.078		

SNP ID	Chr	position	gene	freq Good Recov (mRS≤1)	freq Poor Recov (mRS>1)	$_{\text{unadjusted}}P^*$	$_{\text{adjusted}}P^†$	OR [95% CI]†	FDR q‡
rs4588727	7	63859252		74.7	70.9	0.235			
rs17824184	11	76617563	*GDPD4*	78.3	72.1	0.050			
rs7811384	7	3568874	*SDK1*	58.3	56.3	0.558			
rs2941422	8	76682389		60.7	57.1	0.292			
rs7985623	13	59220111	*DIAPH3*	64.3	59.8	0.197			
rs1243659	14	20151347		80.0	87.1	**0.005**	**0.008**	0.50 [0.29-0.84]	0.109
rs17300340	1	178588341	*ACBD6*	81.3	86.7	**0.037**	0.169		
rs7024250	9	3823480	*GLIS3*	91.0	89.6	**0.020**	0.485		
rs10273634	7	33318154	*BBS9*	86.5	78.0	**7.7×10^{-4}**	**0.002**	2.21 [1.32-3.7]	0.057
rs7145567	14	77428041	*ADCK1*	58.0	55.9	0.563			
rs10050717	5	164099184		58.9	54.1	0.180			
rs947403	10	1006601		51.3	46.7	0.177			
rs3027232	17	7962790	*ALOXE3*	73.1	80.9	**0.009**	0.095		
rs1904006	10	53479033	*PRKG1*	78.6	76.9	0.555			
rs12508742	4	134733668		59.4	59.7	0.952			
rs9532718	13	40542722	*WBP4*	52.0	50.0	0.575			
rs1000128	9	4017185	*GLIS3*	67.3	63.3	0.221			

Odds Ratio (OR) >1 indicates increased probability of poor outcome at three months for the carriers of the minor allele.
Chr – Chromosome, CI – confidence interval, FDR – false discovery rate, freq Good Recov – allele frequency (%) in patients with good outcome, freq Poor Recov – allele frequency (%) in patients with poor outcome
$^*P$ for the log-additive model

†OR [95% CI] and *P* for the log-additive genetic model after adjustment for significant covariates (history of hypertension, and occurrence of aphasia, paresis, altered consciousness and medical complications during hospitalization)
‡FDR *q* values

Table 4 – Association analysis results for the *GLIS3* haplotypes and stroke outcome.

Gene	Haplotypes		Haplotype frequency	freq Good Recov (mRS≤1)	freq Poor Recov (mRS>1)	χ^2	*P*	FDR *q* *
GLIS3	rs7024250-rs1000128-rs10974334	TAT	0.434	0.418	0.455	1.187	0.276	
		TGT	0.221	0.199	0.248	2.775	0.096	
		TAG	0.168	0.201	0.126	8.430	**0.004**	**0.024**
		TGG	0.081	0.092	0.068	1.578	0.209	
		GAT	0.053	0.054	0.052	0.013	0.908	
		GGT	0.033	0.027	0.040	1.188	0.276	

FDR – false discovery rate, freq Good Recov – allele frequency (%) in patients with good outcome, freq Poor Recov – allele frequency (%) in patients with poor outcome
*FDR *q* values

Chapter 7. General discussion and final considerations

7.1 General discussion

Unraveling the genetic basis of stroke is a challenging task for several reasons. In most cases, stroke is a complex trait with polygenic etiology, i.e. reflecting the influence of many *loci* that modulate different pathophysiological processes. Furthermore, stroke is characterized by i) phenotypic heterogeneity – with different subtypes possibly reflecting different etiologies, ii) allelic or genetic heterogeneity – in which an identical phenotype might be caused by distinct mutations in the same or different genes, respectively, iii) phenocopy – some individuals without an inherited risk allele will have the disease due to random or environmental causes, iv) variable penetrance – some individuals carrying an inherited risk allele may not have a stroke (owing to epistatic interactions for example), and v) confounders – coexistence of other risk factors, like hypertension, in affected individuals can make it harder to assess the effects of a gene [Hassan and Markus 2000]. These features may partially explain the lack of replication or contradictory results of many published linkage and association studies on stroke. For instance, the association with stroke may be missed due to allelic or linkage disequilibrium (LD) pattern heterogeneity between populations if only the associated markers in the original study are tested in the replication study. When two populations have different ethnicities, distinct causal variants may exist in the same gene in these populations (allelic heterogeneity), which can be in LD with different genetic markers. It is also possible that a specific causal variant is associated with different genetic markers in the two populations owing to differences in LD patterns. Therefore, replication studies that are carried out in populations with different ethnicities and genetic backgrounds should analyze the complete genetic variation in a region/gene to reduce the likelihood of no association due to allelic or LD pattern heterogeneity. Other possible explanations for replication failure include differences in study designs or the existence of false positive results in the original study that were correctly non-replicated in the following studies.

To identify the genetic risk factors for stroke, a candidate gene approach was used in this thesis. Several studies have shown the important role of inflammation and oxidative stress in stroke susceptibility: known risk factors for stroke, such as

hypertension and atherosclerosis, are associated with an elevated systemic inflammatory profile, and the nitric oxide metabolism plays a role in atherosclerosis and blood pressure regulation [Bastard et al. 2006; Carvalho et al. 2006; Hansson and Libby 2006; Kuhlencordt et al. 2001; Kuhlencordt et al. 2006; Moutsopoulos and Madianos 2006; Shesely et al. 1996]. Four inflammatory genes (*IL1B*, *IL6*, *MPO* and *TNF*) and two genes involved in the oxidative stress processes (*NOS1* and *NOS3*) were thus selected and tested for association with stroke. These analyses were carried out in a sample of 672 stroke patients and 530 healthy controls for whom detailed clinical and life-style information on stroke risk factors were available. This made possible to identify and adjust our results for potential confounders, and to detect associations that are independent of these non-genetic risk factors.

NOS3 is a widely tested candidate gene, as *Nos3* knockout (KO) animals are hypertensive and develop greater atherosclerotic lesions. Earlier studies with *Nos1* KO animals, on the other hand, did not anticipate a role of the nNOS protein in stroke, possibly explaining the lack of interest on the *NOS1* gene. Further studies with nNOS inhibitors and *Apoe*/*Nos1* double knockout (KO) mice, however, have suggested that *NOS1* is also a strong candidate for a role in stroke risk [Kuhlencordt et al. 2006; reviewed in Toda et al. 2009a; reviewed in Toda et al. 2009b]. The work described herein support this latter hypothesis, while not confirming the effect of *NOS3* variants in this phenotype. We also showed that variants in the inflammatory genes *IL6* and *MPO* are positively associated with stroke and identified an epistatic interaction effect between them. The *IL6* findings support previous association results that have been obtained in different population samples, reinforcing *IL6* as a susceptibility gene for stroke. A methodological improvement in comparison to earlier studies is that, instead of analyzing a single SNP, we analyzed haplotype tagging SNPs covering the full genetic variation in that genomic region. This was done to increase the probability of replication of the *IL6* association, preventing potential negative results due to allelic or LD pattern heterogeneity between populations, and to clarify the association of this gene with stroke. Indeed, different *IL6* markers were associated with stroke in our sample when compared with previous studies, proving the effectiveness of our strategy. Overall, these results indicate that susceptibility to stroke may be modulated by main gene and gene-gene interaction effects, in addition to the well-known contribution of clinical and life-style risk factors. Furthermore, they are compatible and strengthen

previous evidence from other research areas for a role of inflammation and oxidative stress in stroke susceptibility.

Genome-wide association studies (GWAS) use high-throughput genotyping technologies to perform an unbiased search for genetic variants that underlie complex traits. In GWAS for stroke, several new candidate *loci* have been identified (http://genome.gov/gwastudies), but none was observed in two independent studies at a genome-wide significance level [Lanktree et al. 2010]. Although GWAS have identified previously unsuspected *loci* in stroke and other diseases, they also have limitations. Large sample sizes are required to detect variants that have small overall effects and testing for gene-gene interactions is computationally difficult due to the large dimension of the data. Furthermore, it is possible that the effect size of any given SNP is not large enough to reach the required stringent significance level of these studies (typically $<10^{-7}$). Finally, GWAS often involve the analysis of individuals from different populations, which can reduce power when genetic or allelic heterogeneity, as well as differences in LD patterns, occur between populations. Therefore, although the candidate gene approach has been gradually replaced by the genome-wide approach, it still remains a useful strategy to identify genetic risk factors for stroke, as was shown in this thesis.

Three out of six candidate genes tested (*IL6*, *MPO*, *NOS1*) showed positive association results. It would be interesting to perform the same association analyses in subgroups of patients with specific subtypes of ischemic stroke to investigate if the evidence for association increases, but this information was not available for the majority of stroke patients. Very stringent significance levels were used to minimize the problem of false positives due to multiple comparisons and to increase confidence about positive association findings. However, genetic markers are often not independent, which is one of the assumptions underlying Bonferroni correction, but associated due to LD. Therefore, it is possible that some association signals not withstanding Bonferroni correction were also true positive findings, since this correction is overly conservative. An alternative is to control for the expected proportion of false positive findings using false discovery rate (FDR) methods [Rice et al. 2008]. These are less computationally intensive and less conservative than permutations and Bonferroni methods, respectively. In the *NOS1*/*NOS3* study, although no association withstood Bonferroni correction, when the four SNPs with the smallest *P*-values are considered significant, less than 5% of these are expected to be false positives (FDR<5%) [Storey and Tibshirani 2003].

Replication of positive results in independent datasets may be the most adequate strategy to assess true positive findings, if strict criteria for study design and population ethnicity can be met, which is often not possible. As already mentioned, replication studies frequently show contradictory results or a lack of replication due to population differences or heterogeneity in study designs [Colhoun et al. 2003; Liu et al. 2008; Palmer and Cardon 2005].

No marker within the *TNF* and *IL1B* genes was associated with stroke susceptibility. Assuming an additive genetic model and disease allele frequency of 10-40%, our sample was well powered (90-99%) to detect a genotype relative risk of 1.5 with a type I error of 5%. Since this sample had enough power to detect common variants of low to moderate effect sizes and the genetic variability in the *TNF* and *IL1B* genomic regions was covered, it is unlikely that these genes have common variants that influence stroke risk, at least in the Portuguese population.

Although additive variance may account for more than 50% of total genetic variance [Hill et al. 2008], epistasis or non-additive interactions are documented in model organisms [Kroymann and Mitchell-Olds 2005; Shao et al. 2008] and several lines of evidence suggest that they are ubiquitous in common human diseases [Moore 2003]. Indeed, gene-gene interactions have been identified in association studies of autism [Coutinho et al. 2007], type 2 diabetes [Neuman et al. 2010] and stroke [Liu et al. 2009; Shen et al. 2007], among many others. Epistasis is, however, not often investigated, which may explain why it has not been widely observed in the etiology of many complex human traits [Greene et al. 2009]. The investigation of gene-gene interactions constitutes a novelty aspect of this work, since the effects of a gene may be missed if it is considered in isolation while functioning through complex networks involving other genes [Cordell 2009]. It is expected that detecting interactions between *loci* will lead to better understanding of the biological and biochemical pathways that underlie a disease [Cordell 2009]. In addition, epistasis may explain the lack of replication or inconsistency of results (in which an initially found protective allele is replicated as a risk allele) of some replication studies [Greene et al. 2009; Moore 2003]. The MDR method was used to investigate the existence of gene-gene interactions since it is more powerful than traditional regression-based methods [Ritchie et al. 2001] and it has increased power to detect interactions even in the absence of main effects [Ritchie et al. 2003]. Moreover, this method addresses the problem of multiple testing through combination of the cross-validation strategy and permutation testing [Moore 2003;

Ritchie et al. 2001]. The MDR identified an *IL6–MPO* interaction influencing stroke susceptibility. Interestingly, the identified interaction is in agreement with a previous *in vitro* study, showing that enzymatically inactive MPO induced IL-6 secretion in a dose and time-dependent manner by endothelial cells [Lefkowitz et al. 2000]. Further studies are now needed to understand the contribution of this interaction specifically in stroke [Moore 2003].

The “common disease, common variant” hypothesis argues that most of the genetic susceptibility to common diseases is due to genetic variants that are relatively frequent in the general population [Schork et al. 2009]. At the beginning of this work, this hypothesis was largely accepted in human genetics research. Thus, haplotype tagging SNPs with minor allele frequency ≥0.05 in the HapMap CEU population were selected in each candidate gene and tested for association with stroke susceptibility. As observed in previous association studies of complex traits [Bodmer and Bonilla 2008], the effect sizes that were detected in the *IL6*, *MPO* and *NOS1* genes are relatively small (OR<2). Furthermore, since the majority of the disease-associated polymorphisms are located in introns and are not splice-site or exonic variants, it is unlikely that they have themselves a functional role. Most probably, they are in LD with the functional variants. Thus, although the indirect association approach offered the opportunity to map disease genes as a consequence of association between tested markers and the true functional variants, understanding the contribution of these genes to disease etiology is only possible if the causal variants are identified [Bodmer and Bonilla 2008; Orr and Chanock 2008].

In the future, the association results of this work should be replicated in samples from the same ethnic population as the original study [McCarthy et al. 2008; Zondervan and Cardon 2007]. Since initial association studies frequently overestimate the true effect size, which is designated as the “winner’s curse” [Ioannidis 2008], replication studies must have adequate power to detect effect sizes that are usually smaller than those of the initial study [Colhoun et al. 2003]. If an association is not replicated, the potential sources of errors and bias responsible for the original association signals could be evaluated [McCarthy et al. 2008]. In addition, it would be desirable to replicate genetic associations in a different ethnic population to assess their generalization [McCarthy et al. 2008].

Given that the identified disease-associated variants in the *IL6*, *MPO* and *NOS1* genes are unlikely to have a functional role, the following step will be to sequence the

genomic regions defined by strong association signals to identify putatively causal variants. Relevant regions for protein function or expression, including promoters, exons, intron-exon boundaries and other splicing regions, and the 5' and 3' untranslated regions (UTRs), could be prioritized in a first stage. After putatively causal variants have been identified, computational approaches, and *in vitro* and *in vivo* studies are required to confirm if these variants are truly causal, to quantify their contribution to stroke and to elucidate their functional roles.

The studies on the genetic factors involved in stroke outcome are scarce, with the majority of them assessing the role of the *APOE* gene in this phenotype. The fact that patients are often left disabled after stroke, which has a huge social and economic impact, lead to the urgency of finding genes that may influence patient's outcome and recovery. Filling this gap was one of the main objectives of this work and the novel results obtained represent an important step towards this objective.

Several lines of research have shown the major contribution of inflammation, oxidative stress, neurogenesis, angiogenesis, neurovascular injury and neurovascular remodeling for stroke-associated brain damage and/or stroke recovery. Therefore, candidate genes involved in inflammatory processes (*IL1B*, *IL6*, *MPO* and *TNF*) and oxidative stress (*NOS1* and *NOS3*), as well as growth factor genes (*BDNF*, *FGF2* and *VEGFA*) and matrix metalloproteinase genes (*MMP2* and *MMP9*), were selected and tested for association with stroke outcome. Variants in the *MMP2* gene and epistatic interactions between the growth factor genes *BDNF*, *FGF2* and *VEGFA* were associated with patient's outcome at three months, constituting novel evidence for a role of these genetic factors in this complex phenotype. Recombinant tissue plasminogen activator (tPA), currently the only approved treatment for ischemic stroke, can activate MMPs, which in turn have been implicated in blood-brain barrier breakdown and neurovascular injury right after stroke [Adibhatla and Hatcher 2008; Zhao et al. 2006]. It is thus possible that *MMP2* gene variants influence patient's outcome after tPA treatment. On the other hand, and given that MMPs contribute to neurovascular remodeling during the later repair phase after stroke [Adibhatla and Hatcher 2008], the usefulness of MMP inhibitors as potentially therapies for stroke may be limited in time and dependent on *MMP2* genotype. Epistatic interaction effects between the *BDNF*, *FGF2* and *VEGFA* genes were identified in stroke outcome even in the absence of main gene effects, which

provides additional support for gene-gene interaction analysis in complex phenotypes. The existence of these interactions in stroke outcome had not been demonstrated before.

A pilot genome-wide association study was also carried out, allowing an unbiased search for genetic factors influencing stroke outcome, which led to novel findings: *BBS9* and *GLIS3* variants were associated with patient's outcome after stroke. Mutations in *BBS9* have been identified in patients with Bardet-Biedl syndrome, which is characterized by obesity, among other features [Nishimura et al. 2005]; excessive energy intake or obesity can negatively influence stroke outcome [Arumugam et al. 2010]. *GLIS3* variants have been associated with type 1 and type 2 diabetes [Barrett et al. 2009; Dupuis et al. 2010], a disease that is associated with severe disability after stroke [Kaarisalo et al. 2005].

The majority of the associated polymorphisms in *MMP2*, as well as in *BBS9* and *GLIS3*, are intronic, with no anticipated influence on splicing. These markers are probably in LD with the functional variants. The detected effect sizes are modest, like those obtained in other association studies of complex traits and in our analysis of stroke susceptibility. The advantage of the SNPSpD approach, which estimates the effective number of independent SNPs in the tested population by taking into account LD patterns between genotyped SNPs [Nyholt 2004], was demonstrated in the *MMP2*/*MMP9* association study. Clearly, the Bonferroni correction is extremely conservative in this case, where only 10 out of 21 *MMP2* SNPs were independent. Adjusting the significance level by the number of independent SNPs is a more reasonable solution to deal with the multiple testing issue. The sample size is not very large, but was adequately powered (82%) to detect a genotype relative risk of 1.5 with a type I error of 5% assuming an additive genetic model and allele frequency of 30%.

The GWAS that was performed using DNA from pooled samples represents one of the main novelties of this work, as until now no GWAS had been carried out in stroke outcome. The DNA pooling strategy has some limitations (e.g. it does not allow the analysis of haplotypes), but this pilot work confirms its feasibility and further supports the notion that this strategy can be used to perform GWAS in a cost-effective way. Since DNA pooling introduces extra experimental errors that can affect power, and to minimize the probability of prioritizing artifactual results, SNPs with high variability among replicates were eliminated and four different strategies were used to select SNPs for individual genotyping. Removing SNPs with high variability among replicates increased confidence for individual genotyping, but also restricted the number of SNPs

that could be assessed. This may be a caveat of this study, because some of the excluded SNPs may have an impact in patient's outcome at three months and, in a second stage, it may be advisable to return to the allelotyping data and be less stringent on this criterion. However, as an exploratory approach, we decided to focus on the most consistently highlighted SNPs. Two novel *loci* were identified, which may provide new insights into the pathophysiological mechanisms that underlie stroke recovery.

The extensive clinical data during hospitalization that was available for our patients permitted the identification of clinical variables that are significant predictors of patient's disability after three months. These variables reflect the severity of the event and, to a certain extent, patient's status at baseline, and were included in the logistic regression analysis to control for the effect of the severity of stroke in patient's outcome. It would be interesting to use the National Institute of Health Stroke Scale (NIHSS), a widely accepted severity scale, or the size and location of the cellular territory affected by stroke, but these were not available for the studied patients. Additionally, statistical approaches that use the original ordered data, like ordinal logistic regression, constitute an alternative to mRS dichotomization [Bath et al. 2007]. Because two classes of the response variable have few individuals, namely the mRS=5 (N=13) and mRS=6 (N=12), we did not employ ordinal logistic regression using the original ordered data, but decided to divide the response variable into two groups (mRS≤1 and mRS>1) and to use binary logistic regression in the analysis. The results here presented represent a significant progress in the field of stroke genetics, both in terms of the number of genes studied and the novelty of results obtained, and may provide important clues on the complex mechanisms that occur during patient's recovery.

7.2 Concluding remarks and future perspectives

Until now, no robust or consistent associations have been obtained in GWAS of stroke [Lanktree et al. 2010]. In the future, multiple well powered studies should try to replicate the GWAS findings in independent samples to assess the role of these genetic factors in stroke risk. A multistage GWAS that is being carried out, involving the collaboration of many groups (including ours) and the analysis of thousands of patients and controls, may help to clarify stroke etiology. Meta-analyses can also be performed, not only to increase the probability of detecting small effect sizes, which may solve some of the discrepancies in genetic association studies, but also to provide more precise estimates of effect sizes [Domingues-Montanari et al. 2010; Munafò and Flint 2004]. Meta-analyses do not, however, substitute adequately powered studies [Munafò and Flint 2004]. We and our collaborators carried out a meta-analysis of five SNPs and two haplotypes in the *ALOX5AP* gene to investigate the contribution of these genetic variants for stroke, since previous association results had been contradictory [Domingues-Montanari et al. 2010]. We found an association of one SNP, rs10507391, with ischemic stroke, suggesting that *ALOX5AP* is indeed a risk *locus* for stroke. SNPs that are not directly genotyped can be accurately estimated through imputation, and this can be incorporated into meta-analyses of GWAS that used different genotyping platforms and into individual studies, increasing the power of these studies [Li et al. 2009; Servin and Stephens 2007].

Restricting the analysis to younger patients is desirable, as the relative influence of genetics on stroke susceptibility decreases with age [Jerrard-Dunne et al. 2003]. In addition, since stroke is a heterogeneous disease in terms of clinical presentation, the complexity of the phenotype must be taken into account. One of the strategies that have been used is the analysis of intermediate phenotypes, like carotid intima-media thickness or leukoaraiosis, which are highly heritable [Carmelli et al. 1998; Turner et al. 2004]. The other alternative is to perform association analyses in less clinically heterogeneous subgroups of patients, because it was found that heritability of stroke depends on stroke subtypes [Jerrard-Dunne et al. 2003; Polychronopoulos et al. 2002] and some associations are only observed in specific subtypes [Gschwendtner et al. 2009; Gudbjartsson et al. 2009]. The analysis of samples from populations with

different ethnic backgrounds [Durbin et al. 2010] and from population isolates [Service et al. 2006] offers some advantages, including the identification of a wide variety of risk *loci* for stroke and etiological pathways [McCarthy et al. 2008]. Future studies should also assess the role of structural and rare variants in stroke, and investigate the existence of gene-gene and gene-environment interactions [Cirulli and Goldstein 2010; Cordell 2009; Thomas 2010].

Genetic studies on stroke outcome are still in their infancy, but this situation has to change in the near future, as these studies may identify novel therapeutic targets and lead to new pharmacological treatments to reduce stroke damage and/or to enhance recovery processes. The work here presented represents a significant change of this situation doubling the number of candidate genes tested for a role in stroke outcome. In addition, the first GWAS was carried out in this complex phenotype resulting in the identification of two previous unsuspected *loci* (*BBS9* and *GLIS3*). The non-hypothesis driven nature of GWAS is the main advantage of these studies and can lead to unexpected results, as demonstrated in this work. Similarly to what was already discussed for stroke susceptibility, future genetic studies should be better powered to detect small effects and should use imputation techniques to estimate SNPs that are not directly genotyped, thus increasing the power of these studies. Furthermore, the role of structural and rare variants, and the existence of gene-gene and gene-environment interactions in stroke outcome need to be properly assessed, and the study of population isolates or with different ethnic backgrounds will contribute to identify a wide variety of *loci* with significant impact in stroke outcome. It would be interesting to analyze patients with similarities in terms of stroke subtype, location and severity of the insult but with very different outcomes, since the influence of specific genes may become more evident. Future studies should also assess the impact of genetic factors on stroke outcome at the short and long term, because some molecules have opposite effects at different stages. For instance, it was observed that MMPs have a deleterious role right after stroke but have beneficial effects in later stages, possibly contributing to functional recovery [Rosell and Lo 2008].

Candidate gene and genome-wide association analyses, like those performed in this work, are only first steps in understanding the pathways contributing to stroke and underlying patient’s recovery afterwards. Integrating information from these and

functional genomic studies, including gene expression and proteomics studies, may provide important insights into the key pathways that are involved in these complex traits. Still, the ultimate goal resides in translating genetic findings into clinical advances, improving stroke prevention, diagnosis and treatment.

During the past five years, hundreds of *loci* have been associated with several common traits in GWAS [Hindorff et al. 2009], but most of them have reduced effects and explain only a small proportion of the heritability of those traits [Maher 2008; Manolio et al. 2009]. This "missing heritability" suggests that additional genetic variants, either unknown common variants with very small effects or structural and rare variants that are poorly or not detected by current methods, contribute to inherited risk [Maher 2008; Manolio et al. 2009; Pearson and Manolio 2008]. Epigenetics, reduced power to assess the impact of gene-gene and gene-environment interactions, or inaccurate estimates of heritability are other plausible explanations for the observed "missing heritability" [Maher 2008; Manolio et al. 2009; Pearson and Manolio 2008].

One of the major limitations of current GWAS is that only a portion of the total genetic variations is assessed [Ku et al. 2010]. Copy number variations (CNVs), generally defined as deletions and duplications that are larger than 1kb, are highly common in the human genome and interesting results have been obtained for several diseases [Ku et al. 2010]. For instance, it was observed that rare CNVs, disrupting genes involved in developmental pathways, are more frequent in patients with schizophrenia when compared to subjects without the disease [Walsh et al. 2008]. In addition to gene disruption, CNVs can contribute to disease by gene dosage alteration or by uncovering deleterious alleles, among other possible mechanisms [Estivill and Armengol 2007]. These structural variants can be detected using array-based methods, such as the SNP arrays and oligonucleotides comparative genomic hybridization arrays. Nevertheless, despite recent improvements in newer genotyping arrays, these methods have poor sensitivity to detect variants that are smaller than 5-10kb. Comparatively, sequencing-based methods have higher sensitivity to detect smaller CNVs and can also detect other structural variants, like short indels (with sizes ranging from 100bp to 1kb) and copy neutral variations (inversions and translocations), which cannot be investigated using SNP genotyping arrays [Ku et al. 2010]. In the future, studies need to assess a larger component of total genetic variations, including non-SNP genetic variants.

The "common disease, rare variant" hypothesis is an alternative to the "common disease, common variant" hypothesis, arguing that genetic susceptibility to common diseases is mainly influenced by multiple rare variants, possibly with relatively large penetrance [Schork et al. 2009]. Both common and rare variants may contribute to common diseases, but the relative impact of each type of variants is still unknown [Cirulli and Goldstein 2010]. Sequencing specific genes has demonstrated that a combination of rare variants can be associated with common diseases and disease-related phenotypes, including type I diabetes, blood pressure, obesity and low low-density lipoprotein (LDL) cholesterol [Ahituv et al. 2007; Cohen et al. 2005; Ji et al. 2008; Nejentsev et al. 2009]. To characterize the contribution of rare variants to the phenotype, researchers can use next generation sequencing technologies. The advantages of these technologies include the production of large quantities of sequence data cheaply and the ability to detect minor alleles accurately [Metzker 2010; Tucker et al. 2009]. Studies involving rare variants may focus on specific genes, on genomic regions identified in GWAS or linkage studies, on exons or the entire genomes [Bansal et al. 2010; Cirulli and Goldstein 2010]. While the costs associated with whole-exome and whole-genome sequencing remain high, two strategies to discover rare variants are to sequence individuals that are at the extreme ends of a quantitative trait (e.g. age at onset) or belonging to families with multiple affected individuals [Cirulli and Goldstein 2010; Manolio et al. 2009]. An alternative to DNA-sequencing is to use custom genotyping arrays to genotype variants identified in previous sequencing studies, such as the 1000 Genomes Project (http://www.1000genomes.org) [Durbin et al. 2010]. Pilot data from this project, whose aim is to characterize over 95% of human DNA variations with MAF≥1% in multiple populations but may also identify lower frequency variants, has been used to design the next generation genotyping arrays [Durbin et al. 2010]. For example, the Illumina's 2.5M genotyping array, which was launched last year, already contains common and rare variants from the 1000 Genomes Project. Data from this project will not only drive the development of more efficient genotyping arrays, but also accelerate fine mapping of genomic regions identified in GWAS and improve the power of imputation methods, which will allow better localization of disease-associated variants [Durbin et al. 2010; Ku et al. 2010].

In spite of several limitations, GWAS have provided new insights into the etiology of several diseases. Results from those studies suggest that autophagy is an important disease mechanism in Crohn's disease, as several polymorphisms associated

with disease susceptibility are located within or near genes related to this process, which was not anticipated, and *loci* identified in type 2 diabetes encode proteins involved in insulin secretion, rather than insulin signaling, which had been the focus of research [Hingorani et al. 2010]. Furthermore, GWAS have shown that some genetic regions influence the risk of different diseases. For example, different SNPs in the same region on chromosome 12 influence the risk of celiac disease, type 1 diabetes and myocardial infarction and the same SNP on chromosome 8 affects type 2 diabetes and prostate cancer risks [Hingorani et al. 2010]. Kalirin (*KALRN*), previously implicated in susceptibility to cardiovascular disorders, has also been associated with stroke, indicating that it may represent a common risk factor for vascular diseases [Krug et al. 2010]. These unexpected findings indicate that some common diseases, initially considered as very different, may have a partially overlapping etiology.

The primary objective of GWAS is the identification of biological pathways involved in a given trait, but their ultimate objective lies in the translation of findings into clinical advances [McCarthy et al. 2008; Pearson and Manolio 2008]. This ultimate goal will be one of the biggest challenges in the near future. New therapeutic targets may be identified within causal pathways, which could lead to new pharmaceutical and preventative approaches [McCarthy et al. 2008]. In addition, the knowledge of individual patterns of disease predisposition will lead to personalized medicine [McCarthy et al. 2008]. However, the majority of disease-associated variants identified have modest effect sizes. This implies that, for most individuals, using GWAS findings in screening for disease risk will provide limited information beyond that available from conventional risk factors [Hingorani et al. 2010; McCarthy et al. 2008]. Therefore, while the predictive genetic tests that have arisen commercially are hoped to revolutionize diagnostic approaches and population risk calculations, they may have little clinical value and can have serious adverse consequences for patients and their families [Janssens and Khoury 2006; McCarthy et al. 2008; Pearson and Manolio 2008]. Low frequency variants with intermediate penetrance effects, which are not detected by current GWAS approaches but could be identified by high-throughput sequencing technologies, are likely to be more clinically valuable [McCarthy et al. 2008]. Moreover, population-based cohort studies are required to analyze the effects of newly identified risk *loci*, providing information on their absolute risk and insights into the interactive effects with environmental factors [Hingorani et al. 2010]. Given the extraordinary progresses in recent years, new and exciting findings are expected in the

near future, as additional examples of the enormous complexity of human diseases and traits.

References

Abraham R, Moskvina V, Sims R, Hollingworth P, Morgan A, Georgieva L, Dowzell K, Cichon S, Hillmer AM, O'Donovan MC, Williams J, Owen MJ, Kirov G. (2008) A genome-wide association study for late-onset Alzheimer's disease using DNA pooling. BMC Med Genomics 1:44

Adams AE, Rosenblatt M, Suva LJ. (1999) Identification of a novel parathyroid hormone-responsive gene in human osteoblastic cells. Bone 24:305-313

Adams HP, Jr., Bendixen BH, Kappelle LJ, Biller J, Love BB, Gordon DL, Marsh EE, 3rd. (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

Adibhatla RM, Hatcher JF. (2008) Tissue plasminogen activator (tPA) and matrix metalloproteinases in the pathogenesis of stroke: therapeutic strategies. CNS Neurol Disord Drug Targets 7:243-253

Agerholm-Larsen B, Nordestgaard BG, Tybjaerg-Hansen A. (2000) ACE gene polymorphism in cardiovascular disease: meta-analyses of small and large studies in whites. Arterioscler Thromb Vasc Biol 20:484-492

Ahituv N, Kavaslar N, Schackwitz W, Ustaszewska A, Martin J, Hebert S, Doelle H, Ersoy B, Kryukov G, Schmidt S, Yosef N, Ruppin E, Sharan R, Vaisse C, Sunyaev S, Dent R, Cohen J, McPherson R, Pennacchio LA. (2007) Medical sequencing at the extremes of human body mass. Am J Hum Genet 80:779-791

Alderton WK, Cooper CE, Knowles RG. (2001) Nitric oxide synthases: structure, function and inhibition. Biochem J 357:593-615

Allaby M. (1999) Dictionary of Zoology. USA: Oxford University Press

Amarenco P, Bogousslavsky J, Caplan LR, Donnan GA, Hennerici MG. (2009) Classification of stroke subtypes. Cerebrovasc Dis 27:493-501

Anantharaman R, Chew FT. (2009) Validation of pooled genotyping on the Affymetrix 500 k and SNP6.0 genotyping platforms using the polynomial-based probe-specific correction. BMC Genet 10:82

Anderson CD, Biffi A, Rost NS, Cortellini L, Furie KL, Rosand J. (2010) Chromosome 9p21 in ischemic stroke: population structure and meta-analysis. Stroke 41:1123-1131

Andlin-Sobocki P, Jönsson B, Wittchen HU, Olesen J. (2005) Cost of disorders of the brain in Europe. Eur J Neurol 12 Suppl 1:1-27

Araki T, Milbrandt J. (2000) Ninjurin2, a novel homophilic adhesion molecule, is expressed in mature sensory and enteric neurons and promotes neurite outgrowth. J Neurosci 20:187-195

Arumugam TV, Phillips TM, Cheng A, Morrell CH, Mattson MP, Wan R. (2010) Age and energy intake interact to modify cell stress pathways and stroke outcome. Ann Neurol 67:41-52

Asplund K, Stegmayr B, Peltonen M. (1998) From the twentieth to the twenty-first century: a public health perspective on stroke. In: Cerebrovascular Disease Pathophysiology, Diagnosis, and

Management. (Ginsberg MD, Bogousslavsky J, eds), Oxford, UK: Blackwell Science, pp 901-918

Atochin DN, Yuzawa I, Li Q, Rauwerdink KM, Malhotra R, Chang J, Brouckaert P, Ayata C, Moskowitz MA, Bloch KD, Huang PL, Buys ES. (2010) Soluble guanylate cyclase alpha1beta1 limits stroke size and attenuates neurological injury. Stroke 41:1815-1819

Baigent C, Keech A, Kearney PM, Blackwell L, Buck G, Pollicino C, Kirby A, Sourjina T, Peto R, Collins R, Simes R, Collaborators CTT. (2005) Efficacy and safety of cholesterol-lowering treatment: prospective meta-analysis of data from 90,056 participants in 14 randomised trials of statins. Lancet 366:1267-1278

Baird AE. (2006) Blood biologic markers of stroke: improved management, reduced cost? Curr Atheroscler Rep 8:267-275

Bak S, Gaist D, Sindrup SH, Skytthe A, Christensen K. (2002) Genetic liability in stroke: a long-term follow-up study of Danish twins. Stroke 33:769-774

Bansal V, Libiger O, Torkamani A, Schork NJ. (2010) Statistical analysis strategies for association studies involving rare variants. Nat Rev Genet 11:773-785

Barrett JC, Clayton DG, Concannon P, Akolkar B, Cooper JD, Erlich HA, Julier C, Morahan G, Nerup J, Nierras C, Plagnol V, Pociot F, Schuilenburg H, Smyth DJ, Stevens H, Todd JA, Walker NM, Rich SS, Consortium TDG. (2009) Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. Nat Genet 41:703-707

Barrett JC, Fry B, Maller J, Daly MJ. (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21:263-265

Bastard JP, Maachi M, Lagathu C, Kim MJ, Caron M, Vidal H, Capeau J, Feve B. (2006) Recent advances in the relationship between obesity, inflammation, and insulin resistance. Eur Cytokine Netw 17:4-12

Bath PM, Gray LJ, Collier T, Pocock S, Carpenter J, Collaboration OAoST. (2007) Can we improve the statistical analysis of stroke trials? Statistical reanalysis of functional outcomes in stroke trials. Stroke 38:1911-1915

Baum AE, Akula N, Cabanero M, Cardona I, Corona W, Klemens B, Schulze TG, Cichon S, Rietschel M, Nöthen MM, Georgi A, Schumacher J, Schwarz M, Abou Jamra R, Höfels S, Propping P, Satagopan J, Detera-Wadleigh SD, Hardy J, McMahon FJ. (2008) A genome-wide association study implicates diacylglycerol kinase eta (DGKH) and several other genes in the etiology of bipolar disorder. Mol Psychiatry 13:197-207

Belmont JW, Leal SM. (2005) Complex phenotypes and complex genetics: an introduction to genetic studies of complex traits. Curr Atheroscler Rep 7:180-187

Berger K, Stögbauer F, Stoll M, Wellmann J, Huge A, Cheng S, Kessler C, John U, Assmann G, Ringelstein EB, Funke H. (2007) The glu298asp polymorphism in the nitric oxide synthase 3 gene is associated with the risk of ischemic stroke in two large independent case-control studies. Hum Genet 121:169-178

Berry FB, Miura Y, Mihara K, Kaspar P, Sakata N, Hashimoto-Tamaoki T, Tamaoki T. (2001) Positive and negative regulation of myogenic differentiation of C2C12 cells by isoforms of the multiple homeodomain zinc finger transcription factor ATBF1. J Biol Chem 276:25057-25065

Bersano A, Ballabio E, Bresolin N, Candelise L. (2008) Genetic polymorphisms for the study of multifactorial stroke. Hum Mutat 29:776-795

Bertina RM, Koeleman BP, Koster T, Rosendaal FR, Dirven RJ, de Ronde H, van der Velden PA, Reitsma PH. (1994) Mutation in blood coagulation factor V associated with resistance to activated protein C. Nature 369:64-67

Bevan S, Dichgans M, Gschwendtner A, Kuhlenbäumer G, Ringelstein EB, Markus HS. (2008) Variation in the PDE4D gene and ischemic stroke risk: a systematic review and meta-analysis on 5200 cases and 6600 controls. Stroke 39:1966-1971

Bhatnagar P, Scarborough P, Smeeton NC, Allender S. (2010) The incidence of all stroke and stroke subtype in the United Kingdom, 1985 to 2008: a systematic review. BMC Public Health 10:539

Bis JC, Heckbert SR, Smith NL, Reiner AP, Rice K, Lumley T, Hindorff LA, Marciante KD, Enquobahrie DA, Monks SA, Psaty BM. (2008) Variation in inflammation-related genes and risk of incident nonfatal myocardial infarction or ischemic stroke. Atherosclerosis 198:166-173

Bodmer W, Bonilla C. (2008) Common and rare variants in multifactorial susceptibility to common diseases. Nat Genet 40:695-701

Bossé Y, Bacot F, Montpetit A, Rung J, Qu HQ, Engert JC, Polychronakos C, Hudson TJ, Froguel P, Sladek R, Desrosiers M. (2009) Identification of susceptibility genes for complex diseases using pooling-based genome-wide association scans. Hum Genet 125:305-318

Broderick JP. (2005) The STICH trial: what does it tell us and where do we go from here? Stroke 36:1619-1620

Brown KM, Macgregor S, Montgomery GW, Craig DW, Zhao ZZ, Iyadurai K, Henders AK, Homer N, Campbell MJ, Stark M, Thomas S, Schmid H, Holland EA, Gillanders EM, Duffy DL, Maskiell JA, Jetann J, Ferguson M, Stephan DA, Cust AE, Whiteman D, Green A, Olsson H, Puig S, Ghiorzo P, Hansson J, Demenais F, Goldstein AM, Gruis NA, Elder DE, Bishop JN, Kefford RF, Giles GG, Armstrong BK, Aitken JF, Hopper JL, Martin NG, Trent JM, Mann GJ, Hayward NK. (2008) Common sequence variants on 20q11.22 confer melanoma susceptibility. Nat Genet 40:838-840

Butcher LM, Davis OS, Craig IW, Plomin R. (2008) Genome-wide quantitative trait locus association scan of general cognitive ability using pooled DNA and 500K single nucleotide polymorphism microarrays. Genes Brain Behav 7:435-446

Candelario-Jalil E, Yang Y, Rosenberg GA. (2009) Diverse roles of matrix metalloproteinases and tissue inhibitors of metalloproteinases in neuroinflammation and cerebral ischemia. Neuroscience 158:983-994

Carmelli D, DeCarli C, Swan GE, Jack LM, Reed T, Wolf PA, Miller BL. (1998) Evidence for genetic variance in white matter hyperintensity volume in normal elderly male twins. Stroke 29:1177-1181

Carvalho THF, Lopes OU, Tolentino-Silva FR. (2006) Baroreflex responses in neuronal nitric oxide synthase knockout mice (nNOS). Autonomic Neuroscience: Basic and Clinical 126-127:163-168

Casas JP, Hingorani AD, Bautista LE, Sharma P. (2004) Meta-analysis of genetic studies in ischemic stroke: thirty-two genes involving approximately 18,000 cases and 58,000 controls. Arch Neurol 61:1652-1661

Chang CF, Chen SF, Lee TS, Lee HF, Shyue SK. (2011) Caveolin-1 deletion reduces early brain injury after experimental intracerebral hemorrhage. Am J Pathol 178:1749-1761

Chen J, Zhang C, Jiang H, Li Y, Zhang L, Robin A, Katakowski M, Lu M, Chopp M. (2005) Atorvastatin induction of VEGF and BDNF promotes brain plasticity after stroke in mice. J Cereb Blood Flow Metab 25:281-290

Chung Y, Lee SY, Elston RC, Park T. (2007) Odds ratio based multifactor-dimensionality reduction method for detecting gene-gene interactions. Bioinformatics 23:71-76

Cirulli ET, Goldstein DB. (2010) Uncovering the roles of rare variants in common disease through whole-genome sequencing. Nat Rev Genet 11:415-425

Clark WM, Rinker LG, Lessov NS, Hazel K, Eckenstein F. (1999) Time course of IL-6 expression in experimental CNS ischemia. Neurol Res 21:287-292

Cohen J, Pertsemlidis A, Kotowski IK, Graham R, Garcia CK, Hobbs HH. (2005) Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. Nat Genet 37:161-165

Colhoun HM, McKeigue PM, Davey Smith G. (2003) Problems of reporting genetic associations with complex outcomes. Lancet 361:865-872

Consortium IH. (2003) The International HapMap Project. Nature 426:789-796

Cordell HJ. (2009) Detecting gene-gene interactions that underlie human diseases. Nat Rev Genet 10:392-404

Cordell HJ, Clayton DG. (2005) Genetic association studies. Lancet 366:1121-1131

Correale J, Villa A. (2004) The neuroprotective role of inflammation in nervous system injuries. J Neurol 251:1304-1316

Correia M, Silva MR, Matos I, Magalhães R, Lopes JC, Ferro JM, Silva MC. (2004) Prospective community-based study of stroke in Northern Portugal: incidence and case fatality in rural and urban populations. Stroke 35:2048-2053

Coutinho AM, Sousa I, Martins M, Correia C, Morgadinho T, Bento C, Marques C, Ataíde A, Miguel TS, Moore JH, Oliveira G, Vicente AM. (2007) Evidence for epistasis between SLC6A4 and ITGB3 in autism etiology and in the determination of platelet serotonin levels. Hum Genet 121:243-256

Cui Y, Li G, Li S, Wu R. (2010) Designs for linkage analysis and association studies of complex diseases. Methods Mol Biol 620:219-242

Dahlback B. (1995) New molecular insights into the genetics of thrombophilia. Resistance to activated protein C caused by Arg506 to Gln mutation in factor V as a pathogenic risk factor for venous thrombosis. Thromb Haemost 74:139-148

Davenport R, Dennis M. (2000) Neurological emergencies: acute stroke. J Neurol Neurosurg Psychiatry 68:277-288

Davis OS, Plomin R, Schalkwyk LC. (2009) The SNPMaP package for R: a framework for genome-wide association using DNA pooling on microarrays. Bioinformatics 25:281-283

Dawn Teare M, Barrett JH. (2005) Genetic linkage studies. Lancet 366:1036-1044

Deb P, Sharma S, Hassan KM. (2010) Pathophysiologic mechanisms of acute ischemic stroke: An overview with emphasis on therapeutic significance beyond thrombolysis. Pathophysiology 17:197-218

Di Carlo A, Lamassa M, Baldereschi M, Pracucci G, Basile AM, Wolfe CD, Giroud M, Rudd A, Ghetti A, Inzitari D. (2003) Sex differences in the clinical presentation, resource use, and 3-month outcome of acute stroke in Europe: data from a multicenter multinational hospital-based registry. Stroke 34:1114-1119

Dichgans M. (2007) Genetics of ischaemic stroke. Lancet Neurol 6:149-161

Diergaarde B, Brand R, Lamb J, Cheong SY, Stello K, Barmada MM, Feingold E, Whitcomb DC. (2010) Pooling-based genome-wide association study implicates gamma-glutamyltransferase 1 (GGT1) gene in pancreatic carcinogenesis. Pancreatology 10:194-200

Dixon RA, Diehl RE, Opas E, Rands E, Vickers PJ, Evans JF, Gillard JW, Miller DK. (1990) Requirement of a 5-lipoxygenase-activating protein for leukotriene synthesis. Nature 343:282-284

Docherty SJ, Butcher LM, Schalkwyk LC, Plomin R. (2007) Applicability of DNA pools on 500 K SNP microarrays for cost-effective initial screens in genomewide association studies. BMC Genomics 8:214

Domingues-Montanari S, Fernández-Cadenas I, del Rio-Espinola A, Corbeto N, Krug T, Manso H, Gouveia L, Sobral J, Mendioroz M, Fernández-Morales J, Alvarez-Sabin J, Ribó M, Rubiera M, Obach V, Martí-Fàbregas J, Freijo M, Serena J, Ferro JM, Vicente AM, Oliveira SA, Montaner J. (2010) Association of a genetic variant in the ALOX5AP with higher risk of ischemic stroke: a case-control, meta-analysis and functional study. Cerebrovasc Dis 29:528-537

Domingues-Montanari S, Mendioroz M, Rio-Espinola Ad, Fernández-Cadenas I, Montaner J. (2008) Genetics of stroke: a review of recent advances. Expert Rev. Mol. Diagn. 8:495-513

Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, Wheeler E, Glazer NL, Bouatia-Naji N, Gloyn AL, Lindgren CM, Mägi R, Morris AP, Randall J, Johnson T, Elliott P, Rybin D, Thorleifsson G, Steinthorsdottir V, Henneman P, Grallert H, Dehghan A, Hottenga JJ, Franklin CS, Navarro P, Song K, Goel A, Perry JR, Egan JM, Lajunen T, Grarup N, Sparsø T, Doney A, Voight BF, Stringham HM, Li M, Kanoni S, Shrader P, Cavalcanti-Proença C, Kumari M, Qi L, Timpson NJ, Gieger C, Zabena C, Rocheleau G, Ingelsson E, An P, O'Connell J, Luan J, Elliott A, McCarroll SA, Payne F, Roccasecca RM, Pattou F, Sethupathy P, Ardlie K, Ariyurek Y, Balkau B, Barter P, Beilby JP, Ben-Shlomo Y, Benediktsson R, Bennett AJ, Bergmann S, Bochud M, Boerwinkle E, Bonnefond A, Bonnycastle LL, Borch-Johnsen K, Böttcher Y, Brunner E, Bumpstead SJ, Charpentier G, Chen YD, Chines P, Clarke R, Coin LJ, Cooper MN, Cornelis M, Crawford G, Crisponi L, Day IN, de Geus EJ, Delplanque J, Dina C, Erdos MR, Fedson AC, Fischer-Rosinsky A, Forouhi NG, Fox CS, Frants R, Franzosi MG, Galan P, Goodarzi MO, Graessler J, Groves CJ, Grundy S, Gwilliam R, Gyllensten U, Hadjadj

S, et al. (2010) New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet 42:105-116

Durbin RM, Abecasis GR, Altshuler DL, Auton A, Brooks LD, Gibbs RA, Hurles ME, McVean GA, Consortium GP. (2010) A map of human genome variation from population-scale sequencing. Nature 467:1061-1073

Elbedour K, Zucker N, Zalzstein E, Barki Y, Carmi R. (1994) Cardiac abnormalities in the Bardet-Biedl syndrome: echocardiographic studies of 22 patients. Am J Med Genet 52:164-169

Emily M, Mailund T, Hein J, Schauser L, Schierup MH. (2009) Using biological networks to search for interacting loci in genome-wide association studies. Eur J Hum Genet 17:1231-1240

Estivill X, Armengol L. (2007) Copy number variants and common disorders: filling the gaps and exploring complexity in genome-wide association studies. PLoS Genet 3:1787-1799

Fardo DW, Becker KD, Bertram L, Tanzi RE, Lange C. (2009) Recovering unused information in genome-wide association studies: the benefit of analyzing SNPs out of Hardy-Weinberg equilibrium. Eur J Hum Genet 17:1676-1682

Faucourt M, Houliston E, Besnardeau L, Kimelman D, Lepage T. (2001) The pitx2 homeobox protein is required early for endoderm formation and nodal signaling. Dev Biol 229:287-306

Feigin VL, Lawes CM, Bennett DA, Barker-Collo SL, Parag V. (2009) Worldwide stroke incidence and early case fatality reported in 56 population-based studies: a systematic review. Lancet Neurol 8:355-369

Flex A, Gaetani E, Papaleo P, Straface G, Proia AS, Pecorini G, Tondi P, Pola P, Pola R. (2004) Proinflammatory genetic profiles in subjects with history of ischemic stroke. Stroke 35:2270-2275

Flossmann E, Schulz UG, Rothwell PM. (2004) Systematic review of methods and results of studies of the genetic epidemiology of ischemic stroke. Stroke 35:212-227

Franco RF, Trip MD, ten Cate H, van den Ende A, Prins MH, Kastelein JJ, Reitsma PH. (1999) The 20210 G-->A mutation in the 3'-untranslated region of the prothrombin gene and the risk for arterial thrombotic disease. Br J Haematol 104:50-54

Frosst P, Blom HJ, Milos R, Goyette P, Sheppard CA, Matthews RG, Boers GJ, den Heijer M, Kluijtmans LA, van den Heuvel LP. (1995) A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase. Nat Genet 10:111-113

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D. (2002) The structure of haplotype blocks in the human genome. Science 296:2225-2229

Girolamo F, Virgintino D, Errede M, Capobianco C, Bernardini N, Bertossi M, Roncali L. (2004) Involvement of metalloprotease-2 in the development of human brain microvessels. Histochem Cell Biol 122:261-270

Green JS, Parfrey PS, Harnett JD, Farid NR, Cramer BC, Johnson G, Heath O, McManamon PJ, O'Leary E, Pryse-Phillips W. (1989) The cardinal manifestations of Bardet-Biedl syndrome, a form of Laurence-Moon-Biedl syndrome. N Engl J Med 321:1002-1009

Greene CS, Penrod NM, Williams SM, Moore JH. (2009) Failure to replicate a genetic association may provide important clues about genetic architecture. PLoS One 4:e5639

Gretarsdottir S, Sveinbjornsdottir S, Jonsson HH, Jakobsson F, Einarsdottir E, Agnarsson U, Shkolny D, Einarsson G, Gudjonsdottir HM, Valdimarsson EM, Einarsson OB, Thorgeirsson G, Hadzic R, Jonsdottir S, Reynisdottir ST, Bjarnadottir SM, Gudmundsdottir T, Gudlaugsdottir GJ, Gill R, Lindpaintner K, Sainz J, Hannesson HH, Sigurdsson GT, Frigge ML, Kong A, Gudnason V, Stefansson K, Gulcher JR. (2002) Localization of a susceptibility gene for common forms of stroke to 5q12. Am J Hum Genet 70:593-603

Gretarsdottir S, Thorleifsson G, Manolescu A, Styrkarsdottir U, Helgadottir A, Gschwendtner A, Kostulas K, Kuhlenbäumer G, Bevan S, Jonsdottir T, Bjarnason H, Saemundsdottir J, Palsson S, Arnar DO, Holm H, Thorgeirsson G, Valdimarsson EM, Sveinbjörnsdottir S, Gieger C, Berger K, Wichmann HE, Hillert J, Markus H, Gulcher JR, Ringelstein EB, Kong A, Dichgans M, Gudbjartsson DF, Thorsteinsdottir U, Stefansson K. (2008) Risk variants for atrial fibrillation on chromosome 4q25 associate with ischemic stroke. Ann Neurol 64:402-409

Gretarsdottir S, Thorleifsson G, Reynisdottir ST, Manolescu A, Jonsdottir S, Jonsdottir T, Gudmundsdottir T, Bjarnadottir SM, Einarsson OB, Gudjonsdottir HM, Hawkins M, Gudmundsson G, Gudmundsdottir H, Andrason H, Gudmundsdottir AS, Sigurdardottir M, Chou TT, Nahmias J, Goss S, Sveinbjornsdottir S, Valdimarsson EM, Jakobsson F, Agnarsson U, Gudnason V, Thorgeirsson G, Fingerle J, Gurney M, Gudbjartsson D, Frigge ML, Kong A, Stefansson K, Gulcher JR. (2003) The gene encoding phosphodiesterase 4D confers risk of ischemic stroke. Nat Genet 35:131-138

Grossetete M, Rosenberg GA. (2008) Matrix metalloproteinase inhibition facilitates cell death in intracerebral hemorrhage in mouse. J Cereb Blood Flow Metab 28:752-763

Gschwendtner A, Bevan S, Cole JW, Plourde A, Matarin M, Ross-Adams H, Meitinger T, Wichmann E, Mitchell BD, Furie K, Slowik A, Rich SS, Syme PD, MacLeod MJ, Meschia JF, Rosand J, Kittner SJ, Markus HS, Müller-Myhsok B, Dichgans M, Consortium ISG. (2009) Sequence variants on chromosome 9p21.3 confer risk for atherosclerotic stroke. Ann Neurol 65:531-539

Gu Z, Cui J, Brown S, Fridman R, Mobashery S, Strongin AY, Lipton SA. (2005) A highly specific inhibitor of matrix metalloproteinase-9 rescues laminin from proteolysis and neurons from apoptosis in transient focal cerebral ischemia. J Neurosci 25:6401-6408

Gudbjartsson DF, Arnar DO, Helgadottir A, Gretarsdottir S, Holm H, Sigurdsson A, Jonasdottir A, Baker A, Thorleifsson G, Kristjansson K, Palsson A, Blondal T, Sulem P, Backman VM, Hardarson GA, Palsdottir E, Helgason A, Sigurjonsdottir R, Sverrisson JT, Kostulas K, Ng MC, Baum L, So WY, Wong KS, Chan JC, Furie KL, Greenberg SM, Sale M, Kelly P, MacRae CA, Smith EE, Rosand J, Hillert J, Ma RC, Ellinor PT, Thorgeirsson G, Gulcher JR, Kong A, Thorsteinsdottir U, Stefansson K. (2007) Variants conferring risk of atrial fibrillation on chromosome 4q25. Nature 448:353-357

Gudbjartsson DF, Holm H, Gretarsdottir S, Thorleifsson G, Walters GB, Thorgeirsson G, Gulcher J, Mathiesen EB, Njølstad I, Nyrnes A, Wilsgaard T, Hald EM, Hveem K, Stoltenberg C, Kucera G, Stubblefield T, Carter S, Roden D, Ng MC, Baum L, So WY, Wong KS, Chan JC, Gieger C,

Wichmann HE, Gschwendtner A, Dichgans M, Kuhlenbäumer G, Berger K, Ringelstein EB, Bevan S, Markus HS, Kostulas K, Hillert J, Sveinbjörnsdóttir S, Valdimarsson EM, Løchen ML, Ma RC, Darbar D, Kong A, Arnar DO, Thorsteinsdottir U, Stefansson K. (2009) A sequence variant in ZFHX3 on 16q22 associates with atrial fibrillation and ischemic stroke. Nat Genet 41:876-878

Gulcher JR, Gretarsdottir S, Helgadottir A, Stefansson K. (2005) Genes contributing to risk for common forms of stroke. Trends Mol Med 11:217-224

Guo JM, Liu AJ, Su DF. (2010) Genetics of stroke. Acta Pharmacol Sin 31:1055-1064

Hankey GJ. (2006) Potential new risk factors for ischemic stroke: what is their potential? Stroke 37:2181-2188

Hankey GJ, Jamrozik K, Broadhurst RJ, Forbes S, Anderson CS. (2002) Long-term disability after first-ever stroke and related prognostic factors in the Perth Community Stroke Study, 1989-1990. Stroke 33:1034-1040

Hansson GK, Libby P. (2006) The immune response in atherosclerosis: a double-edged sword. Nat Rev Immunol 6:508-519

Hardie K, Hankey GJ, Jamrozik K, Broadhurst RJ, Anderson C. (2004) Ten-year risk of first recurrent stroke and disability after first-ever stroke in the Perth Community Stroke Study. Stroke 35:731-735

Harnett JD, Green JS, Cramer BC, Johnson G, Chafe L, McManamon P, Farid NR, Pryse-Phillips W, Parfrey PS. (1988) The spectrum of renal disease in Laurence-Moon-Biedl syndrome. N Engl J Med 319:615-618

Hassan A, Markus HS. (2000) Genetics and ischaemic stroke. Brain 123 (Pt 9):1784-1812

Hata J, Matsuda K, Ninomiya T, Yonemoto K, Matsushita T, Ohnishi Y, Saito S, Kitazono T, Ibayashi S, Iida M, Kiyohara Y, Nakamura Y, Kubo M. (2007) Functional SNP in an Sp1-binding site of AGTRL1 gene is associated with susceptibility to brain infarction. Hum Mol Genet 16:630-639

Healy DG. (2006) Case-control studies in the genomic era: a clinician's guide. Lancet Neurol 5:701-707

Helgadottir A, Manolescu A, Thorleifsson G, Gretarsdottir S, Jonsdottir H, Thorsteinsdottir U, Samani NJ, Gudmundsson G, Grant SF, Thorgeirsson G, Sveinbjornsdottir S, Valdimarsson EM, Matthiasson SE, Johannsson H, Gudmundsdottir O, Gurney ME, Sainz J, Thorhallsdottir M, Andresdottir M, Frigge ML, Topol EJ, Kong A, Gudnason V, Hakonarson H, Gulcher JR, Stefansson K. (2004) The gene encoding 5-lipoxygenase activating protein confers risk of myocardial infarction and stroke. Nat Genet 36:233-239

Herrmann O, Tarabin V, Suzuki S, Attigah N, Coserea I, Schneider A, Vogel J, Prinz S, Schwab S, Monyer H, Brombacher F, Schwaninger M. (2003) Regulation of body temperature and neuroprotection by endogenous interleukin-6 in cerebral ischemia. J Cereb Blood Flow Metab 23:406-415

Hill WG, Goddard ME, Visscher PM. (2008) Data and theory point to mainly additive genetic variance for complex traits. PLoS Genet 4:e1000008

Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci U S A 106:9362-9367

Hingorani AD, Shah T, Kumari M, Sofat R, Smeeth L. (2010) Translating genomics into improved healthcare. BMJ 341:c5945

Horstmann S, Kalb P, Koziol J, Gardner H, Wagner S. (2003) Profiles of matrix metalloproteinases, their inhibitors, and laminin in stroke patients: influence of different therapies. Stroke 34:2165-2170

Horstmann S, Su Y, Koziol J, Meyding-Lamade U, Nagel S, Wagner S. (2006) MMP-2 and MMP-9 levels in peripheral blood after subarachnoid hemorrhage. J Neurol Sci 251:82-86

Hosmer DW, Lemeshow S. (2000) Applied Logistic Regression. New York: John Wiley & Sons, Inc.

Houslay MD, Adams DR. (2003) PDE4 cAMP phosphodiesterases: modular enzymes that orchestrate signalling cross-talk, desensitization and compartmentalization. Biochem J 370:1-18

Hoy A, Leininger-Muller B, Poirier O, Siest G, Gautier M, Elbaz A, Amarenco P, Visvikis S. (2003) Myeloperoxidase polymorphisms in brain infarction. Association with infarct size and functional outcome. Atherosclerosis 167:223-230

Hsu JY, McKeon R, Goussev S, Werb Z, Lee JU, Trivedi A, Noble-Haeusslein LJ. (2006) Matrix metalloproteinase-2 facilitates wound healing events that promote functional recovery after spinal cord injury. J Neurosci 26:9841-9850

Huang PL. (1999) Neuronal and endothelial nitric oxide synthase gene knockout mice. Braz J Med Biol Res 32:1353-1359

Huang PL, Dawson TM, Bredt DS, Snyder SH, Fishman MC. (1993) Targeted disruption of the neuronal nitric oxide synthase gene. Cell 75:1273-1286

Huang Z, Huang PL, Ma J, Meng W, Ayata C, Fishman MC, Moskowitz MA. (1996) Enlarged infarcts in endothelial nitric oxide synthase knockout mice are attenuated by nitro-L-arginine. J Cereb Blood Flow Metab 16:981-987

Huang Z, Huang PL, Panahian N, Dalkara T, Fishman MC, Moskowitz MA. (1994) Effects of cerebral ischemia in mice deficient in neuronal nitric oxide synthase. Science 265:1883-1885

Humphries SE, Morgan L. (2004) Genetic risk factors for stroke and carotid atherosclerosis: insights into pathophysiology from candidate gene approaches. Lancet Neurol 3:227-235

Hurtado O, Pradillo JM, Alonso-Escolano D, Lorenzo P, Sobrino T, Castillo J, Lizasoain I, Moro MA. (2006) Neurorepair versus neuroprotection in stroke. Cerebrovasc Dis 21 Suppl 2:54-63

Hyakkoku K, Hamanaka J, Tsuruma K, Shimazawa M, Tanaka H, Uematsu S, Akira S, Inagaki N, Nagai H, Hara H. (2010) Toll-like receptor 4 (TLR4), but not TLR3 or TLR9, knock-out mice have neuroprotective effects against focal cerebral ischemia. Neuroscience 171:258-267

Ikeda N, Nonoguchi N, Zhao MZ, Watanabe T, Kajimoto Y, Furutama D, Kimura F, Dezawa M, Coffin RS, Otsuki Y, Kuroiwa T, Miyatake S. (2005) Bone marrow stromal cells that enhanced fibroblast growth factor-2 secretion by herpes simplex virus vector improve neurological outcome after transient focal cerebral ischemia in rats. Stroke 36:2725-2730

Ikram MA, Seshadri S, Bis JC, Fornage M, DeStefano AL, Aulchenko YS, Debette S, Lumley T, Folsom AR, van den Herik EG, Bos MJ, Beiser A, Cushman M, Launer LJ, Shahar E, Struchalin M, Du

Y, Glazer NL, Rosamond WD, Rivadeneira F, Kelly-Hayes M, Lopez OL, Coresh J, Hofman A, DeCarli C, Heckbert SR, Koudstaal PJ, Yang Q, Smith NL, Kase CS, Rice K, Haritunians T, Roks G, de Kort PL, Taylor KD, de Lau LM, Oostra BA, Uitterlinden AG, Rotter JI, Boerwinkle E, Psaty BM, Mosley TH, van Duijn CM, Breteler MM, Longstreth WT, Wolf PA. (2009) Genomewide association studies of stroke. N Engl J Med 360:1718-1728

Ioannidis JP. (2008) Why most discovered true associations are inflated. Epidemiology 19:640-648

Jakulin A, Bratko I. (2003) Analyzing attribute interactions. Lect. Notes Artif. Intell. 2838:229-240

Janssens ACJW, Khoury MJ. (2006) Predictive value of testing for multiple genetic variants in multifactorial diseases: implications for the discourse on ethical, legal and social issues. Italian Journal of Public Health 3:35-41

Jeffs B, Clark JS, Anderson NH, Gratton J, Brosnan MJ, Gauguier D, Reid JL, Macrae IM, Dominiczak AF. (1997) Sensitivity to cerebral ischaemic insult in a rat model of stroke is determined by a single genetic locus. Nat Genet 16:364-367

Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, Doerks T, Julien P, Roth A, Simonovic M, Bork P, von Mering C. (2009) STRING 8--a global view on proteins and their functional interactions in 630 organisms. Nucleic Acids Res 37:D412-416

Jerrard-Dunne P, Cloud G, Hassan A, Markus HS. (2003) Evaluating the genetic component of ischemic stroke subtypes: a family history study. Stroke 34:1364-1369

Ji W, Foo JN, O'Roak BJ, Zhao H, Larson MG, Simon DB, Newton-Cheh C, State MW, Levy D, Lifton RP. (2008) Rare independent mutations in renal salt handling genes contribute to blood pressure variation. Nat Genet 40:592-599

Jin K, Wang X, Xie L, Mao XO, Zhu W, Wang Y, Shen J, Mao Y, Banwait S, Greenberg DA. (2006) Evidence for stroke-induced neurogenesis in the human brain. Proc Natl Acad Sci U S A 103:13198-13202

Jood K, Ladenvall C, Rosengren A, Blomstrand C, Jern C. (2005) Family History in Ischemic Stroke Before 70 Years of Age. The Sahlgrenska Academy Study on Ischemic Stroke. Stroke 36:1383-1387

Joutel A, Corpechot C, Ducros A, Vahedi K, Chabriat H, Mouton P, Alamowitch S, Domenga V, Cécillion M, Marechal E, Maciazek J, Vayssiere C, Cruaud C, Cabanis EA, Ruchoux MM, Weissenbach J, Bach JF, Bousser MG, Tournier-Lasserve E. (1996) Notch3 mutations in CADASIL, a hereditary adult-onset condition causing stroke and dementia. Nature 383:707-710

Kaarisalo MM, Räihä I, Sivenius J, Immonen-Räihä P, Lehtonen A, Sarti C, Mähönen M, Torppa J, Tuomilehto J, Salomaa V. (2005) Diabetes worsens the outcome of acute ischemic stroke. Diabetes Res Clin Pract 69:293-298

Kagiyama S, Fukuhara M, Matsumura K, Lin Y, Fujii K, Iida M. (2005) Central and peripheral cardiovascular actions of apelin in conscious rats. Regul Pept 125:55-59

Kalafatis M, Rand MD, Mann KG. (1994) The mechanism of inactivation of human factor V and human factor Va by activated protein C. J Biol Chem 269:31869-31880

Kasner SE. (2006) Clinical interpretation and use of stroke scales. Lancet Neurol 5:603-612

Kathiresan S, Melander O, Guiducci C, Surti A, Burtt NP, Rieder MJ, Cooper GM, Roos C, Voight BF, Havulinna AS, Wahlstrand B, Hedner T, Corella D, Tai ES, Ordovas JM, Berglund G, Vartiainen E, Jousilahti P, Hedblad B, Taskinen MR, Newton-Cheh C, Salomaa V, Peltonen L, Groop L, Altshuler DM, Orho-Melander M. (2008) Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. Nat Genet 40:189-197

Kathiresan S, Voight BF, Purcell S, Musunuru K, Ardissino D, Mannucci PM, Anand S, Engert JC, Samani NJ, Schunkert H, Erdmann J, Reilly MP, Rader DJ, Morgan T, Spertus JA, Stoll M, Girelli D, McKeown PP, Patterson CC, Siscovick DS, O'Donnell CJ, Elosua R, Peltonen L, Salomaa V, Schwartz SM, Melander O, Altshuler D, Merlini PA, Berzuini C, Bernardinelli L, Peyvandi F, Tubaro M, Celli P, Ferrario M, Fetiveau R, Marziliano N, Casari G, Galli M, Ribichini F, Rossi M, Bernardi F, Zonzin P, Piazza A, Yee J, Friedlander Y, Marrugat J, Lucas G, Subirana I, Sala J, Ramos R, Meigs JB, Williams G, Nathan DM, MacRae CA, Havulinna AS, Berglund G, Hirschhorn JN, Asselta R, Duga S, Spreafico M, Daly MJ, Nemesh J, Korn JM, McCarroll SA, Surti A, Guiducci C, Gianniny L, Mirel D, Parkin M, Burtt N, Gabriel SB, Thompson JR, Braund PS, Wright BJ, Balmforth AJ, Ball SG, Hall AS, Linsel-Nitschke P, Lieb W, Ziegler A, König I, Hengstenberg C, Fischer M, Stark K, Grosshennig A, Preuss M, Wichmann HE, Schreiber S, Ouwehand W, Deloukas P, Scholz M, Cambien F, Li M, Chen Z, Wilensky R, Matthai W, Qasim A, Hakonarson HH, Devaney J, Burnett MS, et al. (2009) Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. Nat Genet 41:334-341

Kim S, Iwao H. (2000) Molecular and cellular mechanisms of angiotensin II-mediated cardiovascular and renal diseases. Pharmacol Rev 52:11-34

Kim YS, Nakanishi G, Lewandoski M, Jetten AM. (2003) GLIS3, a novel member of the GLIS subfamily of Krüppel-like zinc finger proteins with repressor and activation functions. Nucleic Acids Res 31:5513-5525

Kirshner HS. (2009) Differentiating ischemic stroke subtypes: risk factors and secondary prevention. J Neurol Sci 279:1-8

Kleindorfer D, Kissela B, Schneider A, Woo D, Khoury J, Miller R, Alwell K, Gebel J, Szaflarski J, Pancioli A, Jauch E, Moomaw C, Shukla R, Broderick JP, Institute N. (2004) Eligibility for recombinant tissue plasminogen activator in acute ischemic stroke: a population-based study. Stroke 35:e27-29

Kroymann J, Mitchell-Olds T. (2005) Epistasis and balanced polymorphism influencing complex trait variation. Nature 435:95-98

Krug T, Manso H, Gouveia L, Sobral J, Xavier JM, Albergaria I, Gaspar G, Correia M, Viana-Baptista M, Simões RM, Pinto AN, Taipa R, Ferreira C, Fontes JR, Silva MR, Gabriel JP, Matos I, Lopes G, Ferro JM, Vicente AM, Oliveira SA. (2010) Kalirin: a novel genetic risk factor for ischemic stroke. Hum Genet 127:513-523

Krupinski J, Kaluza J, Kumar P, Wang M, Kumar S. (1993) Prognostic value of blood vessel density in ischaemic stroke. Lancet 342:742

Ku CS, Loy EY, Salim A, Pawitan Y, Chia KS. (2010) The discovery of human genetic variations and their use as disease markers: past, present and future. J Hum Genet 55:403-415

Kubo M, Hata J, Ninomiya T, Matsuda K, Yonemoto K, Nakano T, Matsushita T, Yamazaki K, Ohnishi Y, Saito S, Kitazono T, Ibayashi S, Sueishi K, Iida M, Nakamura Y, Kiyohara Y. (2007) A nonsynonymous SNP in PRKCH (protein kinase C eta) increases the risk of cerebral infarction. Nat Genet 39:212-217

Kuhlencordt PJ, Gyurko R, Han F, Scherrer-Crosbie M, Aretz TH, Hajjar R, Picard MH, Huang PL. (2001) Accelerated atherosclerosis, aortic aneurysm formation, and ischemic heart disease in apolipoprotein E/endothelial nitric oxide synthase double-knockout mice. Circulation 104:448-454

Kuhlencordt PJ, Hotten S, Schodel J, Rutzel S, Hu K, Widder J, Marx A, Huang PL, Ertl G. (2006) Atheroprotective effects of neuronal nitric oxide synthase in apolipoprotein e knockout mice. Arterioscler Thromb Vasc Biol 26:1539-1544

Lakhan SE, Kirchgessner A, Hofer M. (2009) Inflammatory mechanisms in ischemic stroke: therapeutic approaches. J Transl Med 7:97

Lanktree MB, Dichgans M, Hegele RA. (2010) Advances in genomic analysis of stroke: what have we learned and where are we headed? Stroke 41:825-832

Lefkowitz DL, Roberts E, Grattendick K, Schwab C, Stuart R, Lincoln J, Allen RC, Moguilevsky N, Bollen A, Lefkowitz SS. (2000) The endothelium and cytokine secretion: the role of peroxidases as immunoregulators. Cell Immunol 202:23-30

Leker RR, Soldner F, Velasco I, Gavin DK, Androutsellis-Theotokis A, McKay RD. (2007) Long-lasting regeneration after ischemia in the cerebral cortex. Stroke 38:153-161

Levy E, Carman MD, Fernandez-Madrid IJ, Power MD, Lieberburg I, van Duinen SG, Bots GT, Luyendijk W, Frangione B. (1990) Mutation of the Alzheimer's disease amyloid gene in hereditary cerebral hemorrhage, Dutch type. Science 248:1124-1126

Li Y, Willer C, Sanna S, Abecasis G. (2009) Genotype imputation. Annu Rev Genomics Hum Genet 10:387-406

Lisabeth LD, Smith MA, Brown DL, Uchino K, Morgenstern LB. (2005) Family history and stroke outcome in a bi-ethnic, population-based stroke surveillance study. BMC Neurol 5:20

Liu J, Sun K, Bai Y, Zhang W, Wang X, Wang Y, Wang H, Chen J, Song X, Xin Y, Liu Z, Hui R. (2009) Association of three-gene interaction among MTHFR, ALOX5AP and NOTCH3 with thrombotic stroke: a multicenter case-control study. Hum Genet 125:649-656

Liu VW, Huang PL. (2008) Cardiovascular roles of nitric oxide: a review of insights from nitric oxide synthase gene disrupted mice. Cardiovasc Res 77:19-29

Liu XS, Zhang ZG, Zhang L, Morris DC, Kapke A, Lu M, Chopp M. (2006) Atorvastatin downregulates tissue plasminogen activator-aggravated genes mediating coagulation and vascular permeability in single cerebral endothelial cells captured by laser microdissection. J Cereb Blood Flow Metab 26:787-796

Liu Y, Papasian CJ, Liu J, Hamilton J, Deng H. (2008) Is Replication the Gold Standard for Validating Genome-Wide Association Findings? PLoS ONE 3:e4037

Lloyd-Jones D, Adams R, Carnethon M, De Simone G, Ferguson TB, Flegal K, Ford E, Furie K, Go A, Greenlund K, Haase N, Hailpern S, Ho M, Howard V, Kissela B, Kittner S, Lackland D, Lisabeth L, Marelli A, McDermott M, Meigs J, Mozaffarian D, Nichol G, O'Donnell C, Roger V, Rosamond W, Sacco R, Sorlie P, Stafford R, Steinberger J, Thom T, Wasserthiel-Smoller S, Wong N, Wylie-Rosett J, Hong Y, Subcommittee AHASCaSS. (2009) Heart disease and stroke statistics--2009 update: a report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Circulation 119:e21-181

Logan A, Ahmed Z, Baird A, Gonzalez AM, Berry M. (2006) Neurotrophic factor synergy is required for neuronal survival and disinhibited axon regeneration after CNS injury. Brain 129:490-502

Lucas SM, Rothwell NJ, Gibson RM. (2006) The role of inflammation in CNS injury and disease. Br J Pharmacol 147 Suppl 1:S232-240

Luheshi G, Rothwell N. (1996) Cytokines and fever. Int Arch Allergy Immunol 109:301-307

Macgregor S. (2007) Most pooling variation in array-based DNA pooling is attributable to array error rather than pool construction error. Eur J Hum Genet 15:501-504

MacLellan CL, Silasi G, Auriat AM, Colbourne F. (2010) Rodent models of intracerebral hemorrhage. Stroke 41:S95-98

Maguire J, Thakkinstian A, Levi C, Lincz L, Bisset L, Sturm J, Scott R, Whyte S, Attia J. (2010) Impact of COX-2 rs5275 and rs20417 and GPIIIa rs5918 Polymorphisms on 90-Day Ischemic Stroke Functional Outcome: A Novel Finding. J Stroke Cerebrovasc Dis

Maher B. (2008) Personal genomes: The case of the missing heritability. Nature 456:18-21

Mandriota SJ, Pepper MS. (1997) Vascular endothelial growth factor-induced in vitro angiogenesis and plasminogen activator expression are dependent on endogenous basic fibroblast growth factor. J Cell Sci 110 (Pt 18):2293-2302

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM. (2009) Finding the missing heritability of complex diseases. Nature 461:747-753

Manso H, Krug T, Sobral J, Albergaria I, Gaspar G, Ferro JM, Oliveira SA, Vicente AM. (2010) Variants of the Matrix Metalloproteinase-2 (MMP2) but not the Matrix Metalloproteinase-9 (MMP-9) genes significantly influence functional outcome after stroke. BMC Med Genet 11:40

Manso H, Krug T, Sobral J, Albergaria I, Gaspar G, Ferro JM, Oliveira SA, Vicente AM. (2011) Variants in the inflammatory *IL6* and *MPO* genes modulate stroke susceptibility through main effects and gene-gene interactions. J Cereb Blood Flow Metab [Epub ahead of print]

Marchini J, Donnelly P, Cardon LR. (2005) Genome-wide strategies for detecting multiple loci that influence complex diseases. Nat Genet 37:413-417

Markus H. (2003) Stroke Genetics. Oxford: Oxford University Press

Markus H. (2004) Genes for stroke. J Neurol Neurosurg Psychiatry 75:1229-1231

Martínez-Fernández E, Gil-Peralta A, García-Lozano R, Chinchón I, Aguilera I, Fernández-López O, Arenas J, Campos Y, Bautista J. (2001) Mitochondrial disease and stroke. Stroke 32:2507-2510

Martínez-González NA, Sudlow CL. (2006) Effects of apolipoprotein E genotype on outcome after ischaemic stroke, intracerebral haemorrhage and subarachnoid haemorrhage. J Neurol Neurosurg Psychiatry 77:1329-1335

Matarín M, Brown WM, Scholz S, Simón-Sánchez J, Fung H, Hernandez D, Gibbs JR, Vrieze FWD, Crews C, Britton A, Langefeld CD, Brott TG, Brown Jr RD, Worrall BB, Frankel M, Silliman S, Case LD, Singleton A, Hardy JA, Rich SS, Meschia JF. (2007) A genome-wide genotyping study in patients with ischaemic stroke: initial analysis and data release. Lancet Neurology 6:414-420

Matarin M, Simon-Sanchez J, Fung HC, Scholz S, Gibbs JR, Hernandez DG, Crews C, Britton A, De Vrieze FW, Brott TG, Brown RD, Jr., Worrall BB, Silliman S, Case LD, Hardy JA, Rich SS, Meschia JF, Singleton AB. (2008) Structural genomic variation in ischemic stroke. Neurogenetics 9:101-108

Matarin M, Singleton A, Hardy J, Meschia J. (2010) The genetics of ischaemic stroke. J Intern Med 267:139-155

Mattace Raso F, Rosato M, Talerico A, Cotronei P, Mattace R. (1999) Intimal-medial thickness of the common carotid arteries and lower limbs atherosclerosis in the elderly. Minerva Cardioangiol 47:321-327

Mayeux R. (2005) Mapping the new frontier: complex genetic disorders. J Clin Invest 115:1404-1407

McCarron MO, Muir KW, Nicoll JA, Stewart J, Currie Y, Brown K, Bone I. (2000) Prospective study of apolipoprotein E genotype and functional outcome following ischemic stroke. Arch Neurol 57:1480-1484

McCarron MO, Muir KW, Weir CJ, Dyker AG, Bone I, Nicoll JA, Lees KR. (1998) The apolipoprotein E epsilon4 allele and outcome in cerebrovascular disease. Stroke 29:1882-1887

McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, Hirschhorn JN. (2008) Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet 9:356-369

McColl BW, McGregor AL, Wong A, Harris JD, Amalfitano A, Magnoni S, Baker AH, Dickson G, Horsburgh K. (2007) APOE epsilon3 gene transfer attenuates brain damage after experimental stroke. J Cereb Blood Flow Metab 27:477-487

Meaburn E, Butcher LM, Schalkwyk LC, Plomin R. (2006) Genotyping pooled DNA using 100K SNP microarrays: a step towards genomewide association scans. Nucleic Acids Res 34:e27

Meaburn EL, Harlaar N, Craig IW, Schalkwyk LC, Plomin R. (2008) Quantitative trait locus association scan of early reading disability and ability using pooled DNA and 100K SNP microarrays in a sample of 5760 children. Mol Psychiatry 13:729-740

Megherbi SE, Milan C, Minier D, Couvreur G, Osseby GV, Tilling K, Di Carlo A, Inzitari D, Wolfe CD, Moreau T, Giroud M. (2003) Association between diabetes and stroke subtype on survival and functional outcome 3 months after stroke: data from the European BIOMED Stroke Project. Stroke 34:688-694

Mehta SL, Manhas N, Raghubir R. (2007) Molecular targets in cerebral ischemia for developing novel therapeutics. Brain Res Rev 54:34-66

Meschia JF. (2004) Clinically translated ischemic stroke genomics. Stroke 35:2735-2739

Metzker ML. (2010) Sequencing technologies - the next generation. Nat Rev Genet 11:31-46

Mitsios N, Gaffney J, Kumar P, Krupinski J, Kumar S, Slevin M. (2006) Pathophysiology of acute ischaemic stroke: an analysis of common signalling mechanisms and identification of new molecular targets. Pathobiology 73:159-175

Mommersteeg MT, Hoogaars WM, Prall OW, de Gier-de Vries C, Wiese C, Clout DE, Papaioannou VE, Brown NA, Harvey RP, Moorman AF, Christoffels VM. (2007) Molecular pathway for the localized formation of the sinoatrial node. Circ Res 100:354-362

Moore JH. (2003) The ubiquitous nature of epistasis in determining susceptibility to common human diseases. Hum Hered 56:73-82

Moore JH, Gilbert JC, Tsai CT, Chiang FT, Holden T, Barney N, White BC. (2006) A flexible computational framework for detecting, characterizing, and interpreting statistical patterns of epistasis in genetic studies of human disease susceptibility. J Theor Biol 241:252-261

Moore JH, Williams SM. (2005) Traversing the conceptual divide between biological and statistical epistasis: systems biology and a more modern synthesis. Bioessays 27:637-646

Moutsopoulos NM, Madianos PN. (2006) Low-grade inflammation in chronic infectious diseases: paradigm of periodontal infections. Ann N Y Acad Sci 1088:251-264

Munafò MR, Flint J. (2004) Meta-analysis of genetic association studies. Trends Genet 20:439-444

Nakata S, Tsutsui M, Shimokawa H, Yamashita T, Tanimoto A, Tasaki H, Ozumi K, Sabanai K, Morishita T, Suda O, Hirano H, Sasaguri Y, Nakashima Y, Yanagihara N. (2007) Statin treatment upregulates vascular neuronal nitric oxide synthase through Akt/NF-kappaB pathway. Arterioscler Thromb Vasc Biol 27:92-98

Nakatomi H, Kuriu T, Okabe S, Yamamoto S, Hatano O, Kawahara N, Tamura A, Kirino T, Nakafuku M. (2002) Regeneration of hippocampal pyramidal neurons after ischemic brain injury by recruitment of endogenous neural progenitors. Cell 110:429-441

Nejentsev S, Walker N, Riches D, Egholm M, Todd JA. (2009) Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. Science 324:387-389

Neuman RJ, Wasson J, Atzmon G, Wainstein J, Yerushalmi Y, Cohen J, Barzilai N, Blech I, Glaser B, Permutt MA. (2010) Gene-gene interactions lead to higher risk for development of type 2 diabetes in an Ashkenazi Jewish population. PLoS One 5:e9903

Nilsson-Ardnor S, Janunger T, Wiklund PG, Lackovic K, Nilsson AK, Lindgren P, Escher SA, Stegmayr B, Asplund K, Holmberg D. (2007) Genome-wide linkage scan of common stroke in families from northern Sweden. Stroke 38:34-40

NINDS. (2004) Stroke: Hope Through Research. NIH Publication 99-2222

Nishimura DY, Swiderski RE, Searby CC, Berg EM, Ferguson AL, Hennekam R, Merin S, Weleber RG, Biesecker LG, Stone EM, Sheffield VC. (2005) Comparative genomics and gene expression analysis identifies BBS9, a new Bardet-Biedl syndrome gene. Am J Hum Genet 77:1021-1033

Nygren J, Kokaia M, Wieloch T. (2006) Decreased expression of brain-derived neurotrophic factor in BDNF(+/-) mice is associated with enhanced recovery of motor performance and increased neuroblast number following experimental stroke. J Neurosci Res 84:626-631

Nyholt DR. (2004) A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. Am. J. Hum. Genet. 74:765-769

Orr N, Chanock S. (2008) Common genetic variation and human disease. Adv Genet 62:1-32

Palmer CN, Kimber CH, Doney AS, Proia AS, Morris AD, Gaetani E, Quarta M, Smith RC, Pola R. (2010) Combined effect of inflammatory gene polymorphisms and the risk of ischemic stroke in a prospective cohort of subjects with type 2 diabetes: a Go-DARTS study. Diabetes 59:2945-2948

Palmer LJ, Cardon LR. (2005) Shaking the tree: mapping complex disease genes with linkage disequilibrium. Lancet 366:1223-1234

Paternoster L, Chen W, Sudlow CL. (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 40:2020-2026

Pearson TA, Manolio TA. (2008) How to interpret a genome-wide association study. JAMA 299:1335-1344

Peddareddygari LR, Dutra AV, Levenstien MA, Sen S, Grewal RP. (2009) An analysis of methylenetetrahydrofolate reductase and glutathione S-transferase omega-1 genes as modifiers of the cerebral response to ischemia. BMC Neurol 9:37

Pereira TV, Rudnicki M, Franco RF, Pereira AC, Krieger JE. (2007) Effect of the G-308A polymorphism of the tumor necrosis factor alpha gene on the risk of ischemic heart disease and ischemic stroke: a meta-analysis. Am Heart J 153:821-830

Peters N, Opherk C, Bergmann T, Castro M, Herzog J, Dichgans M. (2005) Spectrum of mutations in biopsy-proven CADASIL: implications for diagnostic strategies. Arch Neurol 62:1091-1094

Polychronopoulos P, Gioldasis G, Ellul J, Metallinos IC, Lekka NP, Paschalis C, Papapetropoulos T. (2002) Family history of stroke in stroke types and subtypes. J Neurol Sci 195:117-122

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559-575

R: A language and Environment for Statistical Computing. (2004). Vienna

Revesz T, Holton JL, Lashley T, Plant G, Rostagno A, Ghiso J, Frangione B. (2002) Sporadic and familial cerebral amyloid angiopathies. Brain Pathol 12:343-357

Rice TK, Schork NJ, Rao DC. (2008) Methods for handling multiple testing. Adv Genet 60:293-308

Rinaldo A, Bacanu SA, Devlin B, Sonpar V, Wasserman L, Roeder K. (2005) Characterization of multilocus linkage disequilibrium. Genet Epidemiol 28:193-206

Ritchie MD. (2011) Using biological knowledge to uncover the mystery in the search for epistasis in genome-wide association studies. Ann Hum Genet 75:172-182

Ritchie MD, Hahn LW, Moore JH. (2003) Power of multifactor dimensionality reduction for detecting gene-gene interactions in the presence of genotyping error, missing data, phenocopy, and genetic heterogeneity. Genet Epidemiol 24:150-157

Ritchie MD, Hahn LW, Roodi N, Bailey LR, Dupont WD, Parl FF, Moore JH. (2001) Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. Am J Hum Genet 69:138-147

Rodríguez-Yáñez M, Castillo J. (2008) Role of inflammatory markers in brain ischemia. Curr Opin Neurol 21:353-357

Rosell A, Lo EH. (2008) Multiphasic roles for matrix metalloproteinases after stroke. Curr Opin Pharmacol 8:82-89

Saenger AK, Christenson RH. (2010) Stroke biomarkers: progress and challenges for diagnosis, prognosis, differentiation, and treatment. Clin Chem 56:21-33

Sahni R, Weinberger J. (2007) Management of intracerebral hemorrhage. Vasc Health Risk Manag 3:701-709

Sarzynska-Dlugosz I, Gromadzka G, Baranska-Gieruszczak M, Ciesielska A, Czlonkowska A. (2007) APOE does not predict poor outcome 1 year after ischemic stroke. Neurol Res 29:64-69

Schäbitz WR, Steigleder T, Cooper-Kuhn CM, Schwab S, Sommer C, Schneider A, Kuhn HG. (2007) Intravenous brain-derived neurotrophic factor enhances poststroke sensorimotor recovery and stimulates neurogenesis. Stroke 38:2165-2172

Schieffer B, Schieffer E, Hilfiker-Kleiner D, Hilfiker A, Kovanen PT, Kaartinen M, Nussberger J, Harringer W, Drexler H. (2000) Expression of angiotensin II and interleukin 6 in human coronary atherosclerotic plaques: potential implications for inflammation and plaque instability. Circulation 101:1372-1378

Schindhelm RK, van der Zwan LP, Teerlink T, Scheffer PG. (2009) Myeloperoxidase: a useful biomarker for cardiovascular disease risk stratification? Clin Chem 55:1462-1470

Schork NJ, Murray SS, Frazer KA, Topol EJ. (2009) Common vs. rare allele hypotheses for complex diseases. Curr Opin Genet Dev 19:212-219

Schulz UG, Flossmann E, Rothwell PM. (2004) Heritability of ischemic stroke in relation to age, vascular risk factors, and subtypes of incident stroke in population-based studies. Stroke 35:819-824

Senée V, Chelala C, Duchatelet S, Feng D, Blanc H, Cossec JC, Charon C, Nicolino M, Boileau P, Cavener DR, Bougnères P, Taha D, Julier C. (2006) Mutations in GLIS3 are responsible for a rare syndrome with neonatal diabetes mellitus and congenital hypothyroidism. Nat Genet 38:682-687

Service S, DeYoung J, Karayiorgou M, Roos JL, Pretorious H, Bedoya G, Ospina J, Ruiz-Linares A, Macedo A, Palha JA, Heutink P, Aulchenko Y, Oostra B, van Duijn C, Jarvelin MR, Varilo T, Peddle L, Rahman P, Piras G, Monne M, Murray S, Galver L, Peltonen L, Sabatti C, Collins A, Freimer N. (2006) Magnitude and distribution of linkage disequilibrium in population isolates and implications for genome-wide association studies. Nat Genet 38:556-560

Servin B, Stephens M. (2007) Imputation-based analysis of association studies: candidate regions and quantitative traits. PLoS Genet 3:e114

Seyedabadi M, Goodchild AK, Pilowsky PM. (2002) Site-specific effects of apelin-13 in the rat medulla oblongata on arterial pressure and respiration. Auton Neurosci 101:32-38

Shao H, Burrage LC, Sinasac DS, Hill AE, Ernest SR, O'Brien W, Courtland HW, Jepsen KJ, Kirby A, Kulbokas EJ, Daly MJ, Broman KW, Lander ES, Nadeau JH. (2008) Genetic architecture of complex traits: large phenotypic effects and pervasive epistasis. Proc Natl Acad Sci U S A 105:19910-19914

Sharma P, Carter ND, Barley J, Lunt R, Seymour CA, Brown MM. (1994) Polymorphisms in the gene encoding angiotensin 1-converting enzyme and relationship to its post-translational product in cerebral infarction. J Hum Hypertens 8:633-634

Shen CD, Zhang WL, Sun K, Wang YB, Zhen YS, Hui RT. (2007) Interaction of genetic risk factors confers higher risk for thrombotic stroke in male Chinese: a multicenter case-control study. Ann Hum Genet 71:620-629

Shendure J, Ji H. (2008) Next-generation DNA sequencing. Nat Biotechnol 26:1135-1145

Shesely EG, Maeda N, Kim HS, Desai KM, Krege JH, Laubach VE, Sherman PA, Sessa WC, Smithies O. (1996) Elevated blood pressures in mice lacking endothelial nitric oxide synthase. Proc Natl Acad Sci U S A 93:13176-13181

Shifman S, Bhomra A, Smiley S, Wray NR, James MR, Martin NG, Hettema JM, An SS, Neale MC, van den Oord EJ, Kendler KS, Chen X, Boomsma DI, Middeldorp CM, Hottenga JJ, Slagboom PE, Flint J. (2008a) A whole genome association study of neuroticism using DNA pooling. Mol Psychiatry 13:302-312

Shifman S, Johannesson M, Bronstein M, Chen SX, Collier DA, Craddock NJ, Kendler KS, Li T, O'Donovan M, O'Neill FA, Owen MJ, Walsh D, Weinberger DR, Sun C, Flint J, Darvasi A. (2008b) Genome-wide association identifies a common variant in the reelin gene that increases the risk of schizophrenia only in women. PLoS Genet 4:e28

Skol AD, Scott LJ, Abecasis GR, Boehnke M. (2006) Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. Nature Genetics 38:209-213

Slevin M, Kumar P, Gaffney J, Kumar S, Krupinski J. (2006) Can angiogenesis be exploited to improve stroke outcome? Mechanisms and therapeutic potential. Clin Sci (Lond) 111:171-183

Small DL, Buchan AM. (2000) Animal models. Br Med Bull 56:307-317

Smith CJ, Emsley HC, Gavin CM, Georgiou RF, Vail A, Barberan EM, del Zoppo GJ, Hallenbeck JM, Rothwell NJ, Hopkins SJ, Tyrrell PJ. (2004) Peak plasma interleukin-6 and other peripheral markers of inflammation in the first week of ischaemic stroke correlate with brain infarct volume, stroke severity and long-term outcome. BMC Neurol 4:2

Sood RR, Taheri S, Candelario-Jalil E, Estrada EY, Rosenberg GA. (2008) Early beneficial effect of matrix metalloproteinase inhibition on blood-brain barrier permeability as measured by magnetic resonance imaging countered by impaired long-term recovery after stroke in rat brain. J Cereb Blood Flow Metab 28:431-438

Sotgiu S, Zanda B, Marchetti B, Fois ML, Arru G, Pes GM, Salaris FS, Arru A, Pirisi A, Rosati G. (2006) Inflammatory biomarkers in blood of patients with acute brain ischemia. Eur J Neurol 13:505-513

Spanbroek R, Grabner R, Lotzer K, Hildner M, Urbach A, Ruhling K, Moos MP, Kaiser B, Cohnert TU, Wahlers T, Zieske A, Plenz G, Robenek H, Salbach P, Kuhn H, Radmark O, Samuelsson B,

Habenicht AJ. (2003) Expanding expression of the 5-lipoxygenase pathway within the arterial wall during human atherogenesis. Proc Natl Acad Sci U S A 100:1238-1243

Steinberg MH, Adewoye AH. (2006) Modifier genes and sickle cell anemia. Curr Opin Hematol 13:131-136

Storey JD, Tibshirani R. (2003) Statistical significance for genomewide studies. Proc Natl Acad Sci U S A 100:9440-9445

Sudlow C, Martínez González NA, Kim J, Clark C. (2006) Does apolipoprotein E genotype influence the risk of ischemic stroke, intracerebral hemorrhage, or subarachnoid hemorrhage? Systematic review and meta-analyses of 31 studies among 5961 cases and 17,965 controls. Stroke 37:364-370

Sugiyama S, Okada Y, Sukhova GK, Virmani R, Heinecke JW, Libby P. (2001) Macrophage myeloperoxidase regulation by granulocyte macrophage colony-stimulating factor in human atherosclerosis and implications in acute coronary syndromes. Am J Pathol 158:879-891

Sun Y, Jin K, Xie L, Childs J, Mao XO, Logvinova A, Greenberg DA. (2003) VEGF-induced neuroprotection, neurogenesis, and angiogenesis after focal cerebral ischemia. J Clin Invest 111:1843-1851

Svedin P, Hagberg H, Savman K, Zhu C, Mallard C. (2007) Matrix metalloproteinase-9 gene knock-out protects the immature brain after cerebral hypoxia-ischemia. J Neurosci 27:1511-1518

Switzer JA, Hess DC, Nichols FT, Adams RJ. (2006) Pathophysiology and treatment of stroke in sickle-cell disease: present and future. Lancet Neurol 5:501-512

Talman WT, Nitschke Dragon D. (2007) Neuronal nitric oxide mediates cerebral vasodilatation during acute hypertension. Brain Res 1139:126-132

Thomas D. (2010) Gene-environment-wide association studies: emerging approaches. Nat Rev Genet 11:259-272

Tilling K, Sterne JA, Rudd AG, Glass TA, Wityk RJ, Wolfe CD. (2001) A new method for predicting recovery after stroke. Stroke 32:2867-2873

Tiret L, Rigat B, Visvikis S, Breda C, Corvol P, Cambien F, Soubrier F. (1992) Evidence, from combined segregation and linkage analysis, that a variant of the angiotensin I-converting enzyme (ACE) gene controls plasma ACE levels. Am J Hum Genet 51:197-205

Toda N, Ayajiki K, Okamura T. (2009a) Cerebral blood flow regulation by nitric oxide: recent advances. Pharmacol Rev 61:62-97

Toda N, Ayajiki K, Okamura T. (2009b) Control of systemic and pulmonary blood pressure by nitric oxide formed through neuronal nitric oxide synthase. J Hypertens 27:1929-1940

Treger I, Froom P, Ring H, Friedman G. (2003) Association between apolipoprotein E4 and rehabilitation outcome in hospitalized ischemic stroke patients. Arch Phys Med Rehabil 84:973-976

Truelsen T, Ekman M, Boysen G. (2005) Cost of stroke in Europe. Eur J Neurol 12 Suppl 1:78-84

Tso AR, Merino JG, Warach S. (2007) Interleukin-6 174G/C polymorphism and ischemic stroke: a systematic review. Stroke 38:3070-3075

Tsuji K, Aoki T, Tejima E, Arai K, Lee SR, Atochin DN, Huang PL, Wang X, Montaner J, Lo EH. (2005) Tissue plasminogen activator promotes matrix metalloproteinase-9 upregulation after focal cerebral ischemia. Stroke 36:1954-1959

Tucker T, Marra M, Friedman JM. (2009) Massively parallel sequencing: the next big thing in genetic medicine. Am J Hum Genet 85:142-154

Turner ST, Jack CR, Fornage M, Mosley TH, Boerwinkle E, de Andrade M. (2004) Heritability of leukoaraiosis in hypertensive sibships. Hypertension 43:483-487

Tyler AL, Asselbergs FW, Williams SM, Moore JH. (2009) Shadows of complexity: what biological networks reveal about epistasis and pleiotropy. Bioessays 31:220-227

van Swieten JC, Koudstaal PJ, Visser MC, Schouten HJ, van Gijn J. (1988) Interobserver agreement for the assessment of handicap in stroke patients. Stroke 19:604-607

Wald DS, Law M, Morris JK. (2002) Homocysteine and cardiovascular disease: evidence on causality from a meta-analysis. BMJ 325:1202

Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, Cooper GM, Nord AS, Kusenda M, Malhotra D, Bhandari A, Stray SM, Rippey CF, Roccanova P, Makarov V, Lakshmi B, Findling RL, Sikich L, Stromberg T, Merriman B, Gogtay N, Butler P, Eckstrand K, Noory L, Gochman P, Long R, Chen Z, Davis S, Baker C, Eichler EE, Meltzer PS, Nelson SF, Singleton AB, Lee MK, Rapoport JL, King MC, Sebat J. (2008) Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. Science 320:539-543

Wang Q, Tang XN, Yenari MA. (2007) The inflammatory response in stroke. J Neuroimmunol 184:53-68

Wang X, Cheng S, Brophy VH, Erlich HA, Mannhalter C, Berger K, Lalouschek W, Browner WS, Shi Y, Ringelstein EB, Kessler C, Luedemann J, Lindpaintner K, Liu L, Ridker PM, Zee RY, Cook NR, Consortium RMSSSNP. (2009) A meta-analysis of candidate gene polymorphisms and ischemic stroke in 6 study populations: association of lymphotoxin-alpha in nonhypertensive patients. Stroke 40:683-695

Warlow C, Sudlow C, Dennis M, Wardlaw J, Sandercock P. (2003) Stroke. Lancet 362:1211-1224

Weimar C, Roth M, Willig V, Kostopoulos P, Benemann J, Diener HC. (2006) Development and validation of a prognostic model to predict recovery following intracerebral hemorrhage. J Neurol 253:788-793

Weimar C, Ziegler A, Konig IR, Diener HC. (2002) Predicting functional outcome and survival after acute ischemic stroke. J Neurol 249:888-895

Weisscher N, Vermeulen M, Roos YB, De Haan RJ. (2008) What should be defined as good outcome in stroke trials; a modified Rankin score of 0-1 or 0-2? J Neurol 255:867-874

Whisnant JP. (1997) Modeling of risk factors for ischemic stroke. The Willis Lecture. Stroke 28:1840-1844

WHO. (1988) The World Health Organization MONICA Project (monitoring trends and determinants in cardiovascular disease): a major international collaboration. WHO MONICA Project Principal Investigators. J Clin Epidemiol 41:105-114

Wong AA, Davis JP, Schluter PJ, Henderson RD, O'Sullivan JD, Read SJ. (2005) The effect of admission physiological variables on 30 day outcome after stroke. J Clin Neurosci 12:905-910

Wu TC, Grotta JC. (2010) Stroke treatment and prevention: five new things. Neurology 75:S16-21

Xi G, Keep RF, Hoff JT. (2006) Mechanisms of brain injury after intracerebral haemorrhage. Lancet Neurol 5:53-63

Yamada Y, Fuku N, Tanaka M, Aoyagi Y, Sawabe M, Metoki N, Yoshida H, Satoh K, Kato K, Watanabe S, Nozawa Y, Hasegawa A, Kojima T. (2009) Identification of CELSR1 as a susceptibility gene for ischemic stroke in Japanese individuals by a genome-wide association study. Atherosclerosis 207:144-149

Yamada Y, Metoki N, Yoshida H, Satoh K, Ichihara S, Kato K, Kameyama T, Yokoi K, Matsuo H, Segawa T, Watanabe S, Nozawa Y. (2006) Genetic risk for ischemic and hemorrhagic stroke. Arterioscler Thromb Vasc Biol 26:1920-1925

Zhao BQ, Wang S, Kim HY, Storrie H, Rosen BR, Mooney DJ, Wang X, Lo EH. (2006) Role of matrix metalloproteinases in delayed cortical responses after stroke. Nat Med 12:441-445

Zhou L, Zhu DY. (2009) Neuronal nitric oxide synthase: structure, subcellular localization, regulation, and clinical implications. Nitric Oxide 20:223-230

Zintzaras E, Rodopoulou P, Sakellaridis N. (2009) Variants of the arachidonate 5-lipoxygenase-activating protein (ALOX5AP) gene and risk of stroke: a HuGE gene-disease association review and meta-analysis. Am J Epidemiol 169:523-532

Zondervan KT, Cardon LR. (2007) Designing candidate gene and genome-wide case-control association studies. Nat Protoc 2:2492-2501

Appendix I. Other manuscript contributions

I1. Rosa A, Fonseca BV, Krug T, **Manso H**, Gouveia L, Albergaria I, Gaspar G, Correia M, Baptista MV, Simões RM, Pinto AN, Taipa R, Ferreira C, Fontes JR, Silva MR, Gabriel JP, Matos I, Lopes G, Ferro JM, Vicente AM, Oliveira SA (2008). Mitochondrial haplogroup H1 is protective for ischemic stroke in Portuguese patients. BMC Medical Genetics, 9:57.

Abstract

Background – The genetic contribution to stroke is well established but it has proven difficult to identify the genes and the disease-associated alleles mediating this effect, possibly because only nuclear genes have been intensely investigated so far. Mitochondrial DNA (mtDNA) has been implicated in several disorders having stroke as one of its clinical manifestations. The aim of this case-control study was to assess the contribution of mtDNA polymorphisms and haplogroups to ischemic stroke risk.

Methods – We genotyped 19 mtDNA single nucleotide polymorphisms (SNPs) defining the major European haplogroups in 534 ischemic stroke patients and 499 controls collected in Portugal, and tested their allelic and haplogroup association with ischemic stroke risk.

Results – Haplogroup H1 was found to be significantly less frequent in stroke patients than in controls (OR=0.61, 95% CI=[0.45-0.83], *P*=0.001), when comparing each clade against all other haplogroups pooled together. Conversely, the pre-HV/HV and U mtDNA lineages emerge as potential genetic factors conferring risk for stroke (OR=3.14, 95%CI=[1.41-7.01], *P*=0.003, and OR=2.87, 95%CI=[1.13-7.28], *P*=0.021, respectively). SNPs m.3010G>A, m.7028C>T and m.11719G>A strongly influence ischemic stroke risk, their allelic state in haplogroup H1 corroborating its protective effect.

Conclusion – Our data suggests that mitochondrial haplogroup H1 has an impact on ischemic stroke risk in a Portuguese sample.

I2. Domingues-Montanari S, Fernández-Cadenas I, del Rio-Espinola A, Corbeto N, Krug T, **Manso H**, Gouveia L, Sobral J, Mendioroz M, Fernández-Morales J, Alvarez-Sabin J, Ribó M, Rubiera M, Obach V, Martí-Fàbregas J, Freijo M, Serena J, Ferro JM, Vicente AM, Oliveira SA, Montaner J (2010). Association of a genetic variant in the *ALOX5AP* gene with higher risk of ischemic stroke – a case-control, meta-analysis and functional study. Cerebrovasc Disease, 29:528-537.

Abstract

Background – Variants in the 5-lipoxygenase-activating protein (*ALOX5AP*) and phosphodiesterase 4D (*PDE4D*) genes have first been associated with ischemic stroke (IS) through whole-genome linkage screens. However, association studies obtained conflicting results. We aimed to investigate the contribution of selected single nucleotide polymorphisms (SNPs) in these genes for the first time in a large Iberian population.

Methods – A case-control design was used to analyze one SNP in *ALOX5AP* and five SNPs in *PDE4D* in a total of 1,092 IS patients and 781 healthy controls of two different subsets from Spain and Portugal. The analysis was adjusted for confounding variables and the results were integrated in a meta-analysis of all case-control studies. In addition, *ALOX5AP* gene expression levels were determined in controls and IS cases.

Results – A first meta-analysis of both subsets showed that the T allele of the SG13S114 SNP in *ALOX5AP* was a risk factor for IS after Bonferroni correction [OR = 1.22 (1.06–1.40); *P*=0.006]. A second meta-analysis of white populations confirmed these results [OR = 1.18 (1.07–1.31); *P*=0.001]. *ALOX5AP* gene expression analysis in a subset of controls and cases revealed that the SG13S114 genotypes modulate mRNA levels of *ALOX5AP* (*P*=0.001) and mRNA levels were higher in IS cases (2.8±2.4%) than in controls (1.4±1.3%; *P* =0.003). No association of the variants in *PDE4D* with IS was observed in our study.

Conclusions – The *ALOX5AP* SG13S114 variant is an independent risk factor for IS in the Iberian population and is associated with *ALOX5AP* expression levels. The role of this gene in stroke merits further investigation.

I3. Krug T, **Manso H**, Gouveia L, Sobral J, Xavier JM, Albergaria I, Gaspar G, Correia M, Baptista MV, Simões RM, Pinto AN, Taipa R, Ferreira C, Fontes JR, Silva MR, Gabriel JP, Matos I, Lopes G, Ferro JM, Vicente AM, Oliveira SA (2010). Kalirin: a novel genetic risk factor for ischemic stroke. Hum Genet 127:513-523.

Abstract

Cerebrovascular and cardiovascular diseases are the leading causes of death and disability worldwide. They are complex disorders resulting from the interplay of genetic and environmental factors, and may share several susceptibility genes. Several recent studies have implicated variants of the Kalirin (*KALRN)* gene with susceptibility to cardiovascular and metabolic phenotypes, but no studies have yet been performed in stroke patients. KALRN is involved, among others, in the inhibition of inducible nitric oxide synthase, in the regulation of ischemic signal transduction, and in neuronal morphogenesis, plasticity, and stability. The goal of the present study was to determine whether SNPs in the *KALRN* region on 3q13, which includes the Ropporin gene (*ROPN1*), predispose to ischemic stroke (IS) in a cohort of Portuguese patients and controls. We genotyped 34 tagging SNPs in the *KALRN* and *ROPN1* chromosomal region on 565 IS patients and 517 unrelated controls, and performed genotype imputation for 405 markers on chromosome 3. We tested the single marker association of these SNPs with IS. One SNP (rs4499545) in the *ROPN1–KALRN* intergenic region and two SNPs in *KALRN* (rs17286604 and rs11712619) showed significant ($P< 0.05$) allelic and genotypic (unadjusted and adjusted for hypertension, diabetes, and ever smoking) association with IS risk. Thirty-two imputed SNPs also showed an association at $P<0.05$, and actual genotyping of three of these polymorphisms (rs7620580, rs6438833, and rs11712039) validated their association. Furthermore, rs11712039 was associated with IS ($0.001<P<0.01$) in a recent well-powered genome-wide association study [Ikram et al. 2009]. These studies suggest that variants in the *KALRN* gene region constitute risk factors for stroke and that *KALRN* may represent a common risk factor for vascular diseases.

I4. Krug T, Gabriel JP, Taipa R, Gouveia L, Fonseca BV, **Manso H**, Albergaria I, Gaspar G, Ferro JM, Vicente AM, Silva MR, Matos I, Lopes G, Oliveira SA (2011) Tetratricopeptide repeat domain 7B emerges as a novel risk factor for ischemic stroke following a multifactorial approach. (Submitted)

Abstract

Stroke is a complex disorder resulting from the interplay of several genetic and environmental factors. Numerous attempts to elucidate its genetic underpinnings have led to conflicting results. We hereby propose a novel approach to the identification of ischemic stroke (IS) susceptibility genes by converging data from several unbiased genetic and genomic tools. Namely, we tested the association with IS of genes that: i. were differentially expressed between IS cases and controls; ii. mapped to previously reported linkage peaks; and iii. were nominally associated with stroke in published genomewide association studies (GWAS).

We performed gene expression profiling in peripheral blood mononuclear cells of twenty IS cases and twenty age- and sex-matched controls. Sixteen of the differentially expressed genes map to previously reported whole-genome linkage peaks, and one of these, the *TTC7B* gene, has been associated (P=5.23x10^{-5}) with major cardiovascular disease in Framingham Heart Study 100K GWAS [Larson et al. 2007]. Forty six tagging single nucleotide polymorphisms (SNPs) in *TTC7B* or its 10 kb flanking regions were genotyped and tested for association in 565 Portuguese IS cases and 520 controls. SNPs nominaly associated in at least one of the tests performed and SNPs defining associated haplotypes were then tested in 570 IS Spanish cases and 390 controls. Several SNPs and haplotypes in the intron 5 – intron 6 region of *TTC7B* were also found associated with IS risk in the Spanish and combined datasets.

Multiple independent lines of evidence therefore support the role of *TTC7B* in stroke susceptibility, but further work is warranted to pinpoint the exact risk variant and to elucidate its pathogenic potential.