BIOMARKERS FOR ISCHAEMIC STROKE

NAZEEHA HASAN

A THESIS SUBMITTED FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

SUPERVISORS:
DR PANKAJ SHARMA
DR ROBERT EDWARDS

DEPARTMENT OF MEDICINE
IMPERIAL COLLEGE LONDON

DECEMBER 2013

ABSTRACT

Background: Current diagnostic, prognostic and risk stratification tools are inadequate for effective ischaemic stroke management. Hypothesising that –omics approaches can be used to detect novel candidate biomarkers for ischaemic stroke, this thesis aimed to evaluate the current status of biomarkers for ischaemic stroke and develop strategies for biomarker discovery.

Methods: Systematic literature review and individual patient data meta-analyses were performed to assess current candidate biomarkers associated with ischaemic stroke. Proteomic SELDI-TOF MS profiling was undertaken to identify novel blood-based protein biomarkers for the diagnosis of acute ischaemic stroke, consisting of a pilot study and a subsequent well-powered discovery study of 104 patients. In an integrative genomics study, transcriptomics data from carotid endarterectomy samples was combined with a genome-wide association study meta-analysis and subjected to functional enrichment analysis to detect differential gene expression or alternative splicing profiles that may be under the control of a genetic variant.

Results: Systematic review and meta-analysis concluded that no current candidate biomarkers could be recommended for routine clinical practice, supporting the pursuit of novel biomarkers for ischaemic stroke and informing the design of subsequent experimental studies. SELDI-TOF MS detected two plasma protein ions, m/z 3699 and m/z 6640, which could differentiate between acute cerebral ischaemia and stroke mimics. Protein ion m/z 6640, identified as ApoC-1, highlighted the role of lipid dysregulation and was postulated to be a novel candidate biomarker for acute cerebral ischaemia. Integrative genomics provided evidence for the genetic regulation of cytoskeletal organisation and extracellular matrix remodelling processes in carotid disease. The *LTBP4* gene was found to be a candidate risk biomarker for ischaemic stroke by predicting plaque instability and rupture.

Conclusions: This work provides workflows for successful biomarker discovery using innovative –omics approaches, highlights key pathogenic pathways and identifies novel candidate biomarkers for ischaemic stroke diagnosis and risk stratification.

The work presented in this thesis is original and my own. Information derived from other sources has been appropriately acknowledged and referenced.

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

PUBLICATIONS AND PRESENTATIONS

PUBLICATIONS

*Hasan N, Altman DG, on behalf of the SU:GAR Group. Association of Admission Glucose Levels with Symptomatic Intracerebral Haemorrhage Following Acute Stroke Thrombolysis: Quantification of Risk Using Individual Patient Data. [Manuscript in Preparation].

Cotlarciuc I, Malik R, Holliday EG, Ahmadi KR, Pare G, Psaty BM, Fornage M, Hasan N, Rinne PE, Ikram MA, Markus HS, Rosand J, Mitchell BD, Kittner SJ, Meschia JF, van Meurs JBJ, Uitterlinden AG, Worrall BB, Dichgans M, Sharma P, on behalf of METASTROKE and the International Stroke Genetics Consortium. Effect of genetic variants associated with plasma homocysteine levels on stroke risk. *Stroke*, 2014 [In Publication].

Yadav S, Cotlarciuc I, Munroe PB, Khan MS, Nalls MA, Bevan S, Cheng YC, Chen WM, Malik R, McCarthy NS, Holliday EG, Speed D, Hasan N, Pucek M, Rinne PE, Sever P, Stanton A, Shields DC, Maguire JM, McEvoy M, Scott RJ, Ferrucci L, Macleod MJ, Attia J, Markus HS, Sale MM, Worrall BB, Mitchell BD, Dichgans M, Sudlow C, Meschia JF, Rothwell PM, Caulfield M, Sharma P; International Stroke Genetics Consortium. Genome-Wide Analysis of Blood Pressure Variability and Ischemic Stroke. *Stroke.* 2013;**44**(10):2703-2709.

*Hasan N. Cardiovascular disease as a measure of sustainable development. *Journal of the Royal Society of Medicine Cardiovascular Disease.* 2013;**2**: 2048004013491731.

Yadav S, Hasan N, Marjot T, Khan MS, Prasad K, Bentley P, Sharma P. Detailed analysis of gene polymorphisms associated with ischemic stroke in South Asians. *PLoS One.* 2013; **8**(3):e57305.

*Hasan N, McColgan P, Bentley P, Edwards RJ, Sharma P. Towards the identification of blood biomarkers for acute stroke in humans: a comprehensive systematic review. *Br J Clin Pharmacol*. 2012;**74**(2):230-40.

Brooke J, Hasan N & Sharma P. Efficacy of information interventions in reducing transfer anxiety from a critical care setting to a general ward: a systematic review and meta-analysis. *Journal of Critical Care.* 2012;**27**(4): 425.e9-15.

Shalhoub J, Davies KJ, Hasan N, Thapar A, Sharma P, Davies AH. The utility of collaborative biobanks for cardiovascular research. *Angiology*. 2012;**63**(5):367-77.

Marjot T, Yadav S, Hasan N, Bentley P & Sharma P. Genes associated with adult cerebral thrombosis. *Stroke* 2011; **42:** 913-918.

ORAL AND POSTER PRESENTATIONS

***Platform Presentation:** Integrative Genomics for Symptomatic Carotid Disease. *International Stroke Genetics Consortium meeting, Lund, Sweden. November 2013.*

***Poster Presentation:** Whole Genome Exon Array of Human Carotid Plaques. *Imperial College London Department of Medicine Young Scientist Day, London, UK. April 2012.*

***Platform presentation:** Towards the Identification of Blood Biomarkers for Acute Stroke. *London Cardiovascular Society, London Medical Society, London, UK. April 2012.*

***Poster Presentation**: Whole Genome Exon Array of Human Carotid Plaques: Differential Gene Expression, Alternative Splicing and the Role of IRF5 in advanced lesion development. *British Heart Foundation Centre of Research Excellence Symposium, Cambridge, UK. June 2011.*

***Platform Presentation:** *Too Sweet to Treat? Rethinking Current Glucose Guidelines for Thrombolytic Therapy. 5th UK Stroke Forum, Glasgow UK. December 2010.*

* Publications and presentations relevant to this thesis.

ACKNOWLEDGEMENTS

First and foremost I would like to thank my supervisors: Dr Pankaj Sharma for his support and guidance throughout my PhD, and Dr Robert Edwards, with whom every conversation has been a humbling lesson in how to be a 'good' scientist.

Special thanks go to the many people who have contributed their time and expertise to the completion of this thesis. I am particularly indebted to Dr Waheedah Abdul-Salam for dedicating her time and help to the proteomics study, as well as providing constant support throughout my PhD. To Dr Paul Bentley, Dr Peter McColgan, Muhammad Saleem-Khan, Professor Doug Altman and the collaborators of the SU:GAR group for their statistical guidance and contribution to the biomarker systematic review and individual patient data meta-analysis. I would like to thank Dr Joseph Shalhoub, Professor Alun Davies, Dr Daniah Trabzuni and Dr Mina Ryten for their involvement with gene expression profiling, as well as Dr Ioana Cotlarciuc, Dr Sunaina Yadav, and Dr Doug Speed for their advice throughout the integrative genomics study.

Thank you to the members of the Imperial College Cerebrovascular Research Unit (ICCRU) for making the past four years so enjoyable. To my family and friends for their patience and support, especially Kartik, for being my rock, Madiha and Siddik for providing a home away from home, and little Samaya, for being my stressball and a much-needed source of sanity.

And finally, to my parents. None of this would have been possible without their unwavering love and encouragement.
This thesis is dedicated to them, and all those who have taught me.

CONTENTS

ABSTRACT **3**

PUBLICATIONS AND PRESENTATIONS **5**

ACKNOWLEDGEMENTS **6**

LIST OF FIGURES **12**

LIST OF TABLES **14**

ABBREVIATIONS **16**

CHAPTER ONE:

INTRODUCTION **19**

1.1 The Burden of Vascular Disease and Stroke **20**

1.2 Stroke Aetiology and Classification **21**

1.2.1 Ischaemic Stroke Subtypes 21

1.2.2 Transient Ischaemic Attack 24

1.3 Stroke Pathophysiology **25**

1.4 Risk Factors for Stroke **27**

1.4.1 Non-Modifiable Risk Factors 27

1.4.1.1 Age and Gender 27

1.4.1.2 Genetics and Family History 27

1.4.2 Modifiable Risk Factors 30

1.4.3 Clinical Risk Factors 30

1.4.3.1 Diabetes and Hypertension 30

1.4.3.2 Transient Ischaemic Attack 31

1.4.3.3 Heart Disease 31

1.4.3.4 Carotid Stenosis 31

1.5 Clinical Evaluation of Ischaemic Stroke **35**

1.5.1 Risk Stratification 35

1.5.1.1 Assessment of Carotid Stenosis and Plaque Stability 35

1.5.2 Diagnosis 37

1.5.2.1 Clinical Scales 37

1.5.2.2 Brain Imaging 38

1.5.2.3 Laboratory Tests 38

1.5.3 Prognosis 39

1.5.3.1 Outcome Scales 39

1.5.4 Prognosis after Thrombolysis 39

1.6 Use of Biomarkers for Improving Patient Management 40
1.6.1 Specimen Selection 41
1.6.2 Multi-Marker Tests 42
1.7 Approaches for Biomarker Discovery 43
1.7.1 Data Handling for Biomarker Discovery 44
1.8 Proteomics 46
1.8.1 Methods for Proteomic Analysis 47
1.8.1.1 SELDI-TOF Mass Spectrometry 48
1.8.2 Proteomics and SELDI-TOF MS for Ischaemic Stroke Biomarker Discovery 49
1.9 Transcriptomics 51
1.9.1 Methods for Transcriptomic Analysis 52
1.9.1.1 Affymetrix Exon Array 52
1.9.2 Transcriptomics for Ischaemic Stroke Biomarker Discovery 53
1.10 Genomics 57
1.10.1 Methods for GWAS Analysis 58
1.10.2 GWAS for Ischaemic Stroke Biomarker Discovery 59
1.10.3 Functional Enrichment Analysis 63
1.10.4 Expression Quantitative Trait Locus and Quantitative Trait Transcript Analysis 64
1.11 Hypothesis and Aims 66

CHAPTER TWO:

BIOMARKERS FOR ISCHAEMIC STROKE – SYSTEMATIC LITERATURE REVIEW AND META-ANALYSIS 68

2.1 Introduction 69
2.2 Methods 71
2.2.1 Literature-Based Meta-Analysis of Diagnostic and Prognostic Biomarkers for Stroke 71
2.2.1.1 Data Sources and Extraction 71
2.2.1.2 Statistical Analysis 72
2.2.2 Literature-Based Meta-Analysis of Admission Glucose Level as a Prognostic Biomarker for Post-Thrombolytic SICH 72
2.2.2.1 Data Sources and Extraction 72
2.2.2.2 Statistical Methods 73
2.2.3 Individual Patient Data Meta-Analysis of Admission Glucose Level as a Prognostic Biomarker for Post-Thrombolytic SICH 74
2.2.3.1 Study Inclusion and Data Acquisition 74
2.2.3.2 Statistical Analysis 74
2.3 Results 76

2.3.1 Literature-Based Meta-Analysis of Diagnostic and Prognostic Biomarkers for Stroke 76
2.3.1.1 Diagnostic Biomarkers of Ischaemic Stroke 78
2.3.1.2 Prognostic Biomarkers of Ischaemic Stroke 81
2.3.1.3 Prognostic Biomarkers for Thrombolysis 83
2.3.1.4 Clinical Relevance of Statistically Significant Biomarkers 84
2.3.2 Literature-Based Meta-Analysis of Admission Glucose Level as a Prognostic Biomarker for Post-Thrombolytic SICH 87
2.3.2.1 Meta-Analysis of Observational Studies 89
2.3.2.2 Meta-Analysis of Clinical Trials 90
2.3.2.3 Comment on the Evaluation of Biomarkers Using Literature Based Meta-Analysis 90
2.3.3 Individual Patient Data Meta-Analysis of Admission Glucose Level as a Prognostic Biomarker for Post-Thrombolytic SICH 93
2.3.3.1 Absolute Risk of SICH Associated with Admission Glucose Level 95
2.3.3.2 Association of SICH with Poor Functional Outcome 99
2.4 Discussion 101

CHAPTER THREE:

A PROTEOMICS APPROACH TO DIAGNOSTIC BIOMARKER DISCOVERY FOR ACUTE ISCHAEMIC STROKE 110
3.1 Introduction 111
3.2 Methods 112
3.2.1 Materials 112
3.2.2 Patient Selection and Sample Acquisition 112
3.2.2.1 Pilot Study 112
3.2.2.2 Replication: Discovery Phase Study 113
3.2.3 SELDI-TOF MS Proteomic Profiling 113
3.2.3.1 Principles of SELDI-TOF MS 113
3.2.3.2 Sample Pre-Processing 115
3.2.3.3 Target Chip Preparation and Sample Incubation 115
3.2.3.4 Data Acquisition 117
3.2.3.5 SELDI-TOF MS Protocol Optimisation 117
3.2.4 Data Interpretation and Analysis 118
3.2.5 Identification and Quantification of Candidate Biomarker 119
3.2.5.1 Immunodepletion of m/z6640 119
3.2.5.2 ApolipoproteinC-1 ELISA 119
3.3 Results 120

3.3.1 Pilot Study 120
3.3.1.1 Evaluation of Data Quality 120
3.3.1.2 Investigating Matrix and pH Binding Conditions for Biomarker Discovery 120
3.3.1.3 Preliminary Diagnostic Biomarkers 125
3.3.2 Discovery Phase Study 126
3.3.2.1 Training Set: Establishing Optimum Experimental Conditions 128
3.3.2.2 Test Set: Replication of Training Set Findings 133
3.3.2.3 Selection of a Candidate Biomarker 1397
3.3.3 Identification of Candidate Marker m/z 6640 141
3.3.3.1 Immunodepletion of m/z 6640 141
3.3.3.2 Measurement and Quantification of m/z 6640 by ELISA 143
3.4 Discussion 1453

CHAPTER FOUR:

AN INTEGRATIVE GENOMICS APPROACH TO RISK BIOMARKER DISCOVERY FOR ISCHAEMIC STROKE 153
4.1 Introduction 154
4.2 Methods 157
4.2.1 Gene Expression Profiling Study 157
4.2.1.1 Sample Selection 157
4.2.1.2 RNA Isolation 160
4.2.1.3 Expression profiling using Affymetrix GeneChip® Human Exon 1.0 ST Arrays 160
4.2.1.4 Quality Control 161
4.2.1.5 Exon Array Data Analysis 162
4.2.2 Genome-Wide Association Study (GWAS) 162
4.2.2.1 Patient Selection 162
4.2.2.2 Evaluation of Carotid Stenosis 164
4.2.2.3 Genotyping and Quality Control 164
4.2.2.4 GWAS Imputation, Association Analysis and Meta-Analysis 164
4.2.2.5 SNP to Gene Annotation 165
4.2.3 Functional Enrichment Analysis 166
4.2.3.1 Hypergeometric Testing 166
4.2.3.2 Gene Set Enrichment Analysis (GSEA) 167
4.2.4 Protein-Protein Interaction Analysis 169
4.3 Results 171
4.3.1 Analysis of Gene Expression Data 171

4.3.1.1 Sample Selection and Patient Characteristics 171
4.3.1.2 Quality Control Analysis 171
4.3.1.3 Unsupervised Analysis 175
4.3.1.4 Supervised Inter-Plaque and Intra-Plaque Analysis 176
4.3.1.5 Paired Symptomatic NST vs SP Analysis 177
4.3.2 Analysis of GWAS Data 178
4.3.2.1 Patient Characteristics 178
4.3.2.2 Quality Control Analysis 178
4.3.2.3 Genetic Loci Associated with Carotid Stenosis in the Presence of Ischaemic Stroke 180
4.3.2.4 SNP Annotation and Cis-eQTL Mapping 183
4.3.2.5 Comparison of GWAS Results with Previously Reported Associations 183
4.3.3 Functional Enrichment Analysis 185
4.3.3.1 Hypergeometric Testing 185
4.3.3.2 Gene Set Enrichment Analysis (GSEA) 187
4.3.3.3 Selection of Candidate Biomarkers 188
4.3.4 Alternative Splicing of TRIOBP 188
4.3.5 Differential Gene Expression of LTBP4 191
4.3.6 LTPB4 as a Biomarker for Symptomatic Carotid Stenosis 192
4.4 Discussion 197

CHAPTER FIVE:

FINAL DISCUSSION 206
5.1 A Journey through the Molecular Mechanisms Underlying Ischaemic Stroke 210
5.2 A Workflow for Biomarker Discovery From –Omics-Derived Data 212
5.2.1 Analysis of –Omics-Derived Data 212
5.2.2 Assessing the Significance of Candidate Biomarkers 213
5.3 Future Directions 214
5.4 Final Remarks 216

REFERENCES 217
APPENDIX 1 248
APPENDIX 2 254
APPENDIX 3 256

LIST OF FIGURES

		Page
1.1	A schematic illustration of intra- and extra-cranial arterial circulation……..	21
1.2	Flow-chart depicting the pathophysiology of ischaemic stroke……..	25
1.3	Schematic illustration of the various layers of the carotid arterial wall……..	33
1.4	Pathway of disease progression and potential impact of biomarkers……..	40
1.5	Summary of the main disciplines in systems biology……..	44
2.1	PRISMA flow diagram for systematic review and meta-analysis of diagnostic and prognostic biomarkers of ischaemic stroke……..	74
2.2	Candidate diagnostic and prognostic biomarkers of significant association with ischaemic stroke upon meta-analysis……..	75
2.3	Forest plot of candidate diagnostic biomarkers able to significantly differentiate between acute ischaemic stroke patients and healthy controls……..	78
2.4	Forest plot comparing S100B levels in acute ischaemic stroke patients and stroke mimics……..	79
2.5	Forest plot comparing MMP-9 levels in acute ischaemic stroke and haemorrhagic stroke patients……..	79
2.6	Forest plot of candidate biomarkers significantly associated with in-hospital death after ischaemic stroke……..	81
2.7	Assessing candidate biomarkers by examining distribution profiles……..	84
2.8	PRISMA flow diagram for systematic review and meta-analysis of glucose as a prognostic biomarker for post-thrombolytic SICH……..	85
2.9	Examination of publication bias in all observational studies included in literature based meta-analysis of the association of glucose with post-thrombolytic symptomatic intracerebral haemorrhage……..	90
2.10	Meta-analysis of studies contributing individual patient data……..	95
2.11	Estimated absolute risk of post-thrombolytic SICH associated with admission glucose levels……..	96
3.1	Illustration of SELDI-TOF MS technology……..	112
3.2	The dynamic interaction between the CM10 ProteinChip array surface, experimental buffers and plasma proteins……..	114
3.3	Representative SELDI-TOF MS spectra from the pilot study using CM10 ProteinChip arrays with SPA matrix……..	121

3.4 The effect of matrix choice and pH on plasma protein profiles in the pilot study…. 122

3.5 Unsupervised analysis of Training Set data by PCA…… 128

3.6 Evaluating pH binding conditions in the Discovery Study Training Set to identify potential diagnostic biomarkers…….. 130

3.7 Examination of putative biomarkers identified through DFA of Training Set data…… 132

3.8 Individual scatter plots of m/z 3699 and m/z 6640…… 135

3.9 Distribution profile of m/z 6640 from Test Set ischaemic and non-ischaemic populations…… 137

3.10 Immunodepletion of m/z 6640, theorised to be apolipoproteinC-1 (ApoC-1)…….. 139

3.11 Measurement and quantification of m/z 6640 by ELISA…….. 141

3.12 Distribution profile of ApoC-1 in ischaemic and non-ischaemic populations…….. 142

4.1 Study design and methodology flowchart…….. 155

4.2 Carotid endarterectomy specimen dissection into stenosising plaque or non-stenosing tissue areas…….. 157

4.3 Plotted QC metrics to assess array data quality…….. 171

4.4 Unsupervised analysis of gene expression data by PCA…… 172

4.5 Venn diagrams depicting gene expression and alternative splicing profiles in various phenotypes…….. 174

4.6 Quantile-Quantile (Q-Q) plots of GWAS association results obtained from BRAINS and ISGS/SWISS cohorts and BRAINS and ISGS/SWISS meta-analysis…… 177

4.7 Manhattan plot of genome-wide association signals from meta-analysis of GWAS of BRAINS and ISGS/SWISS ischaemic stroke cohorts…….. 178

4.8 Exploring the alternative splicing profile of TRIOBP…….. 187

4.9 Exploring the differential expression of LTBP4…….. 190

4.10 Protein-Protein Interaction analysis of LTBP4…….. 192

4.11 Correlations in expression levels between genes involved in the LTBP4 network…… 193

4.12 Examination of LTBP4 as a biomarker for predicting carotid plaque rupture. …….. 193

LIST OF TABLES

		Page
1.1	Summary of SELDI-TOF MS proteomics investigations of acute ischaemic stroke……..	50
1.2	Summary of microarray transcriptomics investigations of carotid artery disease……..	55
1.3	Summary of GWAS of ischaemic stroke and carotid artery stenosis……..	61
2.1	Candidate diagnostic biomarkers of ischaemic stroke identified through literature review……..	77
2.2	Candidate prognostic biomarkers of ischaemic stroke identified through literature review……..	80
2.3	Candidate biomarkers predicting prognosis after thrombolysis for ischaemic stroke, as identified through literature review……	83
2.4	Characteristics of studies included in the literature based meta-analysis of the association of glucose with post-thrombolytic symptomatic intracerebral haemorrhage……..	86
2.5	Analysis of studies included in literature based meta-analysis of the association of glucose with post-thrombolytic symptomatic intracerebral haemorrhage……..	89
2.6	Characteristics and patient demographics of studies contributing individual patient data for analysis of the association of admission glucose levels with risk of post-thrombolytic SICH……	92
2.7	Analysis of studies contributing individual patient data……..	94
2.8	Investigation of functional outcome in patients with SICH……..	98
2.9	The relative merits of individual patient data meta-analysis over published literature-based meta-analysis……..	105
3.1	Pilot study patient characteristics……..	120
3.2	Preliminary diagnostic biomarkers identified from the pilot study……..	123
3.3	Discovery Study Training Set patient characteristics……..	125
3.4	Differentially expressed protein ions in the Discovery Study Training Set……..	127
3.5	Discovery Study Test Set patient characteristics……	133
3.6	Logistic regression analysis of Test Set demographic variables……..	136

4.1 Gene expression study total patient population characteristics……..…….. 156

4.2 Gene expression study patient characteristics……..…….. 170

4.3 GWAS cohort patient characteristics…….. 176

4.4 GWAS association results of carotid stenosis ≥20% in ischaemic stroke patients, after meta-analysis of BRAINS and ISGS/SWISS cohorts…… 179

4.5 Comparison of GWAS results with previously reported associations…….. 181

4.6 Functional Enrichment analysis by hypergeometric testing of gene expression and GWAS data…… 183

4.7 SNPs in the *TRIOBP* gene…….. 186

4.8 SNPs in the *LTBP4* gene…….. 188

ABBREVIATIONS

2D-PAGE	2-dimensional polyacrylamide gel electrophoresis
ACN	acetonitrile
AIS	acute ischaemic stroke
ANOVA	analysis of variance
ApoC-1	apolipoprotein C-1
AUC	area under the curve
BBB	blood brain barrier
BRAINS	Bio-Repository of DNA in Stroke
CCA	common carotid artery
CBF	cerebral blood flow
CEA	carotid endarterectomy
CHAPS	3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulphonate
CHCA	alpha-cyano-4-hydroxycinnamic acid
CI	confidence interval
cIMT	carotid intima-media thickening
CRP	C-reactive protein
CT	computed tomography
CV	coefficient of variation
CVD	cardiovascular disease
CSF	cerebral spinal fluid
Da	dalton(s)
DAPPLE	Disease Association Protein-Protein Link Evaluator
DIGE	difference gel electrophoresis
DNA	deoxyribonucleic acid
DW	diffusion weighted
ECA	external carotid artery
ECASS	European Cooperative Acute Stroke Study
ECST	European Carotid Surgery Trial
EDTA	ethylenediaminetetraacetic acid
ELISA	enzyme-linked immunosorbent assay
eQTL	expression quantitative trait locus
FDA	Food and Drug Administration
FDR	false discovery rate
GO	Go Ontology
GSEA	Gene Set Enrichment Analysis
GWAS	genome-wide association study
HDL	high density lipoprotein
HT	haemorrhagic transformation

ICA	internal carotid artery
IPD	individual patient data
ISGC	International Stroke Genetics Consortium
ISGS	Ischemic Stroke Genetics Study
KEGG	Kyoto Encyclopaedia of Genes and Genomes
LAA	large-artery atherosclerosis
LACI	lacunar infarcts
LD	linkage disequilibrium
LDL	low density lipoprotein
LBTP4	latent transforming growth factor beta binding protein 4
MAF	minor allele frequency
MALDI	matrix assisted laser desorption/ionisation
MI	myocardial infarction
MMP	matrix metallopreinase
MRI	magnetic resonance imaging
mRS	modified rankin scale
MW	molecular weight
MS	mass spectrometry
m/z	mass to charge ratio
NASCET	North American Symptomatic Carotid Endarterectomy Trial
NCD	non-communicable disease
NHS	National Health Service
NIHSS	National Institute for Health Stroke Scale
NINDS	National Institute of Neurological Disorders and Stroke
OCSP	Oxford Community Stroke Project
OR	odds ratio
OTT	onset to treatment time
PACI	partial anterior circulation infarct
PCA	principal component analysis
pI	isoelectric point
POCI	posterior circulation infarct
QC	quality control
QTT	quantitative trait transcript
RIN	RNA integrity number
RNA	ribonucleic acid
ROC	receiver operating characteristic
rt-PA	recombinant tissue plasminogen activator
SD	standard deviation
SELDI-TOF	surface enhanced laser desorption/ionisation time of flight
SICH	symptomatic intracerebral haemorrhage
SITS-ISTR	Safe Implementation of Thrombolysis in Stroke International Stroke Thrombolysis Register
SITS-MOST	Safe Implementation of Thrombolysis in Stroke-Monitoring Study

SMD	standardised mean difference
SNP	single nucleotide polymorphism
SPA	sinapic acid
SUC	stroke of undetermined cause
SWISS	Siblings with Ischemic Stroke Study
TACI	total anterior circulation infarct
TIA	transient ischaemic attack
TOAST	Trial of Org 10172 in Acute Stroke Treatment
TRIOBP	TRIO and F-actin-binding protein
WTCCC	Wellcome Trust Case-Control Consortium

CHAPTER ONE:

INTRODUCTION

1.1 The Burden of Vascular Disease and Stroke

In 2001, the United Nations Millennium Development Goals were established to tackle many of our greatest global development challenges. However it took another 10 years, following a historic United Nations High-Level Meeting in September 2011, for non-communicable diseases (NCDs) to receive international recognition as major targets for future sustainable development goals. Four main NCDs - cardiovascular disease (CVD), cancer, chronic respiratory diseases and diabetes – were responsible for two-thirds of global deaths in 2008.

The term 'cardiovascular disease' encompasses a family of vascular diseases, including cerebrovascular disease, coronary artery disease and peripheral artery disease. Stroke (cerebrovascular disease) is the second leading cause of death worldwide after ischaemic heart disease, and the major cause of disability in adults (Mukherjee & Patil 2011). With accelerating global trends in consumption, urbanisation and life expectancy, the burden of stroke is set to soar, and it is predicted that by 2030, the global prevalence of stroke will reach 77 million (Strong et al. 2007), with the financial cost of stroke projected to more than double (Ovbiagele et al. 2013). Contrary to common belief, CVDs are not only 'diseases of affluence', and over 85% of stroke deaths occurr in low and middle income countries; from 1970 to 2008, there has been a 42% decrease in stroke incidence in high-income countries, while low and middle income countries have experienced a 100% increase (Feigin et al. 2009). This population is more exposed to vascular disease risk factors, with less accessibility to treatment, diagnosis and prevention strategies than those in high income countries.

In the UK, over 150,000 people suffer from a stroke every year. Studies report an incidence of around 1.62/1000/year (Rothwell et al., 2004), costing the NHS over £2.8 billion (National Audit Office 2005). Stroke has a greater disability impact than any other chronic disease, with 20-30% of stroke patients dying within one month, another 30% left with a long-term disability, and 12% requiring institutional care within one year (Adamson et al. 2004). The Department of Health recognised the importance of developing better stroke services by launching the National Stroke Strategy in December 2007, with the aim of reducing the death rate from stroke, coronary heart disease and related diseases in people under 75 years of age by 40% by 2010 (National Stroke Strategy Report 2007). Subsequent initiatives have seen stroke incidence and mortality fall in recent years; however, morbidity and prevalence rates have increased (Lee et al. 2011). Hence there is further scope to improve risk stratification of high risk individuals and develop patient management to enhance diagnosis and prognosis following stroke.

1.2 Stroke Aetiology and Classification

According to the Multinational Monitoring of Trends and Determinants in Cardiovascular Disease (MONICA) project, stroke is defined as "rapidly developed clinical signs of focal or global disturbance of cerebral function lasting more than 24 hours (unless interrupted by surgery or death), with no apparent cause other than a vascular origin. It includes patients presenting with clinical signs and symptoms suggestive of subarachnoid haemorrhage, intracerebral haemorrhage or cerebral ischaemic necrosis. It does not include transient cerebral ischaemia or stroke events in cases of blood disease, brain tumour or brain metastases" (World Health Organisation MONICA Project 1990).

Stroke is classified on the basis of its aetiology as either ischaemic (80%) or haemorrhagic (20%). Ischaemic stroke is caused by occlusion of a cerebral artery through thrombosis (50%) or embolism (25%), while haemorrhagic stroke (intracerebral and/or subarachnoid haemorrhage) is caused by spontaneous rupture of a cerebrospinal artery (Warlow et al. 2003). Despite the two types sharing similar risk profiles (Michel & Bogousslavsky 2006), they exhibit distinct molecular mechanisms in the acute phase (Kleinig & Vink 2009; Wang et al. 2007; Wang & Doré 2007; Xi et al. 2006).

1.2.1 Ischaemic Stroke Subtypes

Ischaemic stroke is not just one disease, but rather consists of a collection of entities with a common final outcome. The heterogeneity and diversity associated with this complex syndrome can be captured by various subtype classifications. Prognosis and risk of recurrence differs between different subtypes of ischaemic stroke, hence classification according to aetiology provides an insight into the mechanisms underlying the stroke and allows informed decisions to be made about patient treatment and management.

A number of classification systems have been proposed for ischaemic stroke: the Oxford Community Stroke Project (OCSP) classification (Bamford et al. 1991), the Atherosclerosis, Small Vessel Disease, Cardiac Source, Other Cause (A-S-C-O) classification (Amarenco et al. 2009), the Trial of Org 10172 in Acute Stroke Treatment (TOAST) classification (Adams et al. 1993) and the Causative Classification of Stroke System (CCS) (Ay et al. 2007). These can be divided into systems focussing upon phenotyic subtyping (OCSP, A-S-C-O), causative subtyping (TOAST) or both (CCS).

The OCSP system is based upon clinical symptoms alone (or in combination with Computed Tomography (CT) scan findings), and hence does not rely on the availability of imaging

modalities. The classification denotes four subtypes of acute cerebral infarction: Total Anterior Circulation Infarcts (TACI), Partial Anterior Circulation Infarcts (PACI), Posterior Circulation Infarcts (POCI) and Lacunar Circulation Infarcts (LACI). Stroke subtype is further coded by adding a final letter 'I' for infarct (e.g. TACI), 'H' for haemorrhage (e.g. TACH) or 'S' for syndrome, prior to imaging (e.g. TACS). The anterior cerebral circulation supplies blood to the anterior portion of the brain, and comprises the internal carotid arteries, anterior cerebral arteries, anterior communicating artery and middle cerebral arteries. The posterior cerebral circulation supplies blood to the posterior region of the brain, including occipital lobes, cerebellum and brainstem, and comprises the vertebral arteries, basilar artery, posterior cerebral artery and posterior communicating artery. Lacunar infarcts occur in the brain's deep structures and are caused by occlusion of penetrating arteries arising from the Circle of Willis, cerebellar arteries and basilar artery (Figure 1.1).

Due to its simplicity and ability to relate the infarct to anatomical and pathophysiological processes, the OCSP system is often used in clinical practice and large-scale observational studies, in particular for the rapid identification of patients eligible for thrombolysis treatment. The four subtypes can be used to determine the area of the brain affected by stroke, and predict the extent of the stroke consequent prognosis. However, the classifications are limited by their inability to investigate causes and potential risk factors of stroke (Chen et al. 2012).

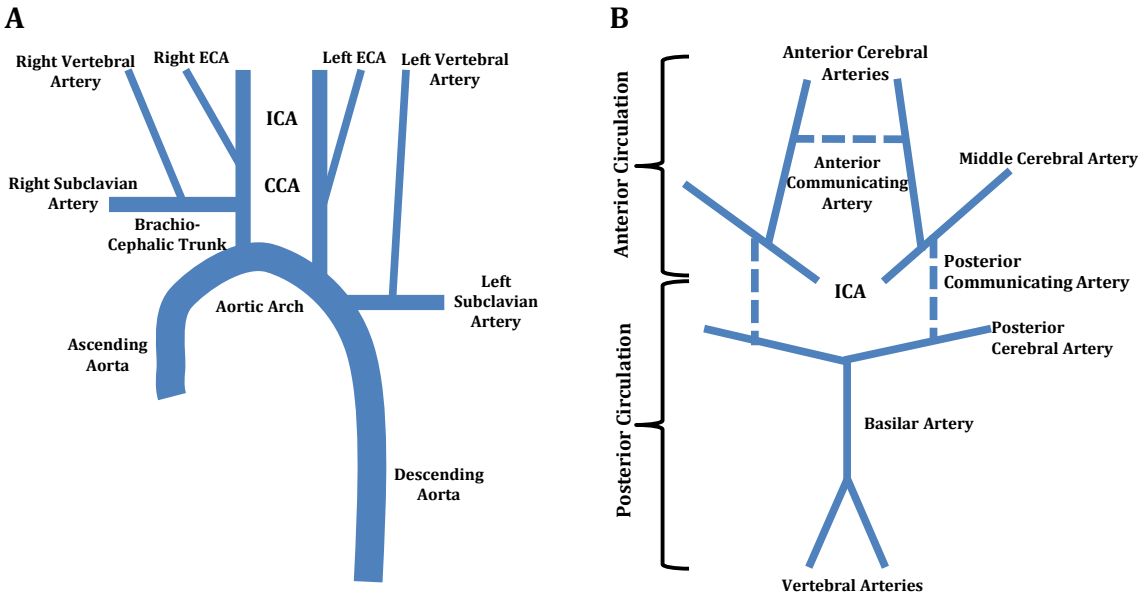


Figure 1.1: A schematic illustration of intra- and extra-cranial arterial circulation. The supra-aortic extracranial anatomy (internal and common carotid arteries) is depicted in **(A),** while the intracranial anatomy (anterior, middle and posterior cerebral arteries), containing the Circle of Willis, is depicted in **(B),** highlighting the anterior circulation derived from the internal carotid arteries and the posterior circulation derived from the vertebral arteries. Atherosclerosis of these major extracranial and intracranial arteries is the cause of Large Artery Atherosclerosis (LAA) ischaemic strokes. CCA, common carotid artery; ECA, external carotid artery; ICA, internal carotid artery. Image modified with permission from Dr. Joseph Shalhoub.

The OCSP exhibits a close relationship to the TOAST classification (Qui 2008), which has become the most widely accepted system to date. The five TOAST subtypes are based upon aetiology, as follows:

Large-Artery Atherosclerosis (LAA): Approximately 20% of all ischaemic strokes are caused by atherosclerosis affecting the major extracranial (internal and common carotid arteries) and intracranial (anterior, middle and posterior cerebral arteries) blood vessels (Figure 1.1), with the ratio of extracranial to intracranial lesions being more than 2:1. Prolonged low bloodflow due to hypotension caused by stenosis, or acute plaque rupture in these arteries causes a decrease in cerebral perfusion, resulting in ischaemia and focal neurological deficits.

Cardioembolism (CE): In a further 30% of ischaemic stroke cases, turbulent or stagnant bloodflow in the heart (e.g. due to atrial fibrillation) can result in the formation of thrombi that can dislodge and occlude blood vessels downstream, thus affecting the blood supply to the brain.

Small Artery Occlusion (SAO): Alterations in the structure of small perforating arteries in the brain (exacerbated by e.g. chronic hypertension, hyperlipidaemia, smoking and diabetes) can lead to decreased compliance, intraluminal stenosis and eventual occlusion. These arterial modifications often lead to small lacunar infarcts (<15 mm^3), typically located in deep brain structures such as the thalamus and brainstem.

Stroke of Other Determined Cause (SOC): While the majority of ischaemic strokes can be classified into the three categories above, some cases require further investigation, especially in young patients and those who present with no apparent risk factors for ischaemic stroke. These rarer causes include genetic disorders (e.g. CADASIL), coagulopathies and metabolic disorders.

Stroke of Undetermined Cause (SUC): In just under 40% of ischaemic stroke cases, there is no obvious cause for stroke, despite extensive investigation. Comprehensive diagnosis and classification of these so called 'cryptogenic' strokes would greatly aid clinical evaluation of their management and prognosis.

The criteria used in TOAST have been recently upgraded in the CCS classification, in which cardioembolic, lacunar and cryptogenic strokes have been further subcategorised (Arsava et al. 2010). CCS and TOAST classifications demonstrate very high agreement with each other, indicating that either system is valuable for stroke subtyping for epidemiological and genetic studies (Lanfranconi & Markus 2012). Nonetheless, CCS has the advantage of being able to

provide both phenotypic and causative information by relying on five sources of data: clinical evaluation, brain imaging, extracranial and intracranial vascular survey, heart evaluations, and work-up for uncommon causes of stroke. Thus, advances in technology and the availability of these technologies within a clinical environment have driven ever more precise and refined diagnosis and classifications of ischaemic stroke subytypes.

1.2.2 Transient Ischaemic Attack

Studies suggest that 7-40% of patients presenting with ischaemic stroke have a pre-warning in the form of a transient ischaemic attack (TIA). TIAs are acute episodes of neurological dysfunction resulting from focal cerebral ischemia, with symptoms lasting less than 24 hours (Easton et al. 2009). Accurate estimates of TIA prevalence and incidence are difficult to determine due to varying criteria used to identify TIA and a general lack of recognition due to challenges in diagnosis. TIAs are often mistaken as stroke mimics (e.g. focal seizures) and diagnosis depends largely on reliable patient history rather than symptoms or brain imaging (Warlow et al. 2003).

TIAs are defined under the assumption that the ischaemic event occurs rapidly enough to cause only transient symptoms that resolve within 24 hours and leave no permanent brain injury. This 24-hour criterion was based on the fact that a syndrome lasting longer than this would be detectable by microscopy, in an era when almost no neuro-diagnostic techniques or treatment for brain ischaemia were available. However, subsequent clinical, experimental and imaging data have suggested that the 24-hour definition of TIA is outdated, does not consider the possibility of brain injury and fails to inform clinicians on appropriate patient management. In light of this a tissue-based, rather than time-based, definition has been proposed, with which the difference between TIA and ischaemic stroke becomes similar to the distinction made between angina (which may be brief or prolonged) and myocardial infarction (Albers 2002). Nonetheless, consensus for a new definition is still pending, and the detection and diagnosis of TIA remains unreliable.

1.3 Stroke Pathophysiology

The brain's response to acute ischaemia is dependent upon the severity and duration of cerebral blood flow (CBF) interruption. The region with CBF lower than 10ml/100g of tissue/min is referred to as the ischaemic core, and these cells are assumed to die within minutes of stroke onset. The region with CBF of up to 10ml/100g of tissue/min is termed the ischaemic penumbra. This area consists of marginal perfusion and potentially salvageable tissue, and is therefore the target for therapeutic reversal of the acute neurological symptomology arising from stroke.

On the molecular level, decrease in CBF causes depletion of oxygen and glucose and subsequent energy failure, leading to dysregulation of membrane Na^+, Ca^{2+}, Cl^- and K^+ ion gradients and neuron depolarisation. The resulting increase in intracellular Ca^{2+} drives the upregulation of kinases, proteases and lipases. Subsequent release of glutamate, dopamine and free radicals causes lipid peroxidation, membrane disruption, DNA damage and inflammation, eventually leading to neuronal death (Muthaian et al. 2012; Dirnagl et al. 1999). These processes are illustrated in Figure 1.2.

The blood brain barrier (BBB) plays a key role in stroke pathophysiology, and is considered by many to be an important therapeutic target (Borlongan et al. 2012). The neurovascular unit of the BBB, consisting of the basal lamina, endothelial cells, pericytes and astrocytic endfeet, contributes to the selective permeability of the barrier. During an acute ischaemic insult the neurovascular units (and endothelial tight junctions in particular) are disrupted through a number of mechanisms including oxidative stress and a cytokine-mediated inflammatory response (Ronaldson & Davis 2012), resulting in a release of brain-derived proteins across the BBB and into the bloodstream (Borlongan et al. 2012).

Although the ischaemic penumbra can be observed for up to 16 hours after stroke onset (Furlan et al. 1996), safe and effective thrombolysis treatment is still limited to a 4.5 hour time window (Del Zoppo et al. 2009). Given that loss of BBB integrity can be considered a precursor to haemorrhagic transformation (Kastrup et al. 2008), understanding BBB damage occurring within the approved thrombolysis time window can carry substantial prognostic information and help determine the decision to treat or not (Liu et al. 2012).

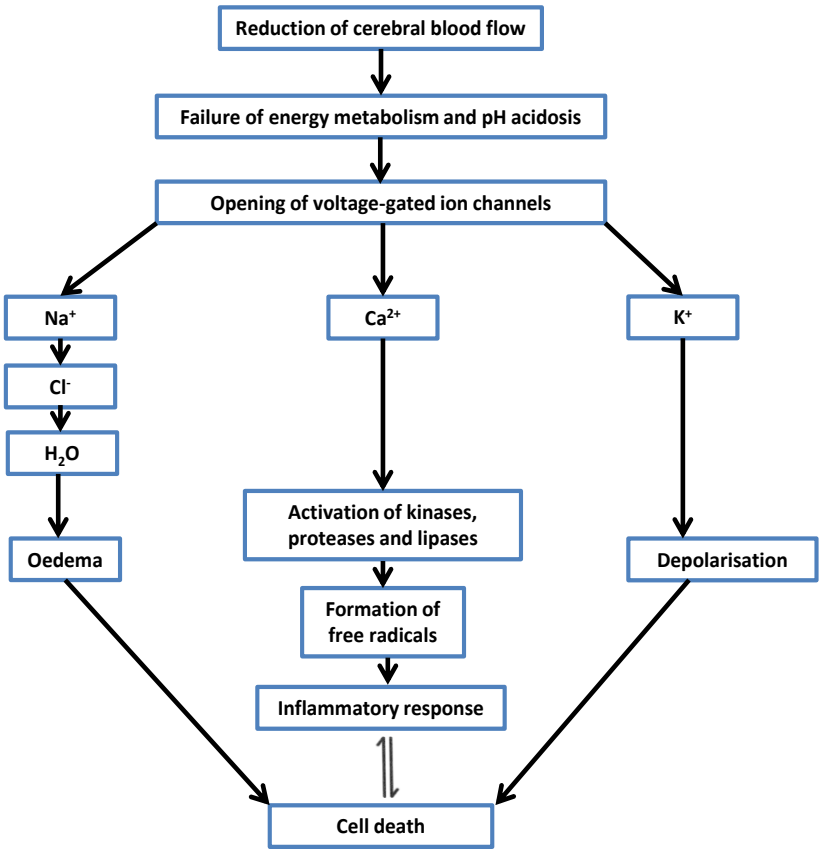


Figure 1.2: Flow-chart depicting the pathophysiology of ischaemic stroke. Energy failure and dysregulation of ion gradients leads to neuronal depolarisation, oedema and inflammation, eventually causing cell death.

1.4 Risk Factors for Stroke

A number of factors have been shown to be associated with an increased risk of ischaemic stroke, supported by many epidemiological and experimental studies. The INTERSTROKE study, consisting of 3000 stroke cases and 3000 controls from 22 countries worldwide, reported that 5 factors (hypertension, current smoking, abdominal obesity, diet, and physical activity) accounted for more than 80% of the global risk of stroke (ischaemic and haemorrhagic). The addition of another five factors (diabetes, heart disease, alcohol consumption, high cholesterol and stress or depressive symptoms) increased the population-attributable risk to 90.3% (O'Donnell et al. 2010). It is also important to appreciate the complex interaction between classic risk factors and socioeconomic factors, which can account for as much as 68% of the variation in standardised stroke mortality rates (Asplund 2005).

Risk factors are often classified as non-modifiable (e.g. age, gender, race, family history) or modifiable (e.g. diet, exercise). Additionally, some clinical conditions have been linked to a higher risk of stroke.

1.4.1 Non-Modifiable Risk Factors

Age, gender, race and family history are non-modifiable factors that have become important for identifying individuals at greater risk at stroke, thus allowing a more targeted approach for addressing modifiable risk factors in this high risk population.

1.4.1.1 Age and Gender

Approximately a quarter of stroke victims in the UK are under the age of 65 (Intercollegiate Stroke Working Party 2012). However, stroke incidence rises rapidly with increasing age, and the Framingham Study found that stroke rate more than doubles in both men and women for each successive decade after age 55 (Brown et al. 1996). Stroke incidence is approximately 25% higher in men than in women. Nonetheless, stroke mortality rate is higher in women than men (10% compared to 7%) (Townsend 2012).

1.4.1.2 Genetics and Family History

Stroke is sometimes overlooked as a genetic disease, often considered to be a sporadic event. Clinical practice traditionally focuses upon modifiable risk factors, which account for a substantial proportion of stroke risk. However, much of stroke risk remains unexplained and it is still unclear why some individuals with risk factors develop stroke while others with a similar risk profile do not. Thus it is likely that there is an intricate network of interaction between these clinical aspects and genetic risk factors.

The heritability of a trait refers to the proportion of total phenotypic variation within a population that can be attributed to genetic factors. Using genome-wide complex trait analysis (GCTA) to estimate heritability by assessing the proportion of variation in case–control status explained by genotyped single nucleotide polymorphisms (SNPs), the genetic heritability of ischaemic stroke has been recently estimated to be 37.9%, with heritability varying by stroke subtype (40.3% for large-vessel disease, 16.1% for small-vessel disease) (Bevan et al. 2012). Intermediate phenotypes of ischaemic stroke also exhibit substantial heritability: the heritability of carotid intima-media thicknening (cIMT) is thought to be greater than for carotid stenosis (h^2=0.61 versus h^2=0.47 respectively), and considering cIMT as a covariate for stenosis results in a heritability estimate of h^2=0.26, suggesting that genes affecting stenosis and cIMT regulate about half of the heritability of stenosis, while independent genes may regulate the other half (Moskau et al. 2005).

A small proportion of ischaemic strokes are monogenic, and identifying mutations in these causal genes (e.g. the *NOTCH3* gene in CADASIL disease) provides diagnostic and prognostic information, specific treatment and enables counselling of other family members (Markus 2011). The vast majority of ischaemic strokes seem to be polygenic, with multiple genes conferring small risk of disease. Identification of these genetic factors could improve risk profiling in a non-symptomatic population. An integrated model combining common genetic variants and traditional risk factors could more effectively discriminate patients likely to develop ischaemic events (Pezzini et al. 2009).

A five-fold increase in stroke prevalence between monozygotic and dizygotic twins (Brass et al. 1992) and the impact of parental history on progeny developing stroke (Seshadri et al. 2010) lends convincing evidence for the role of genetic susceptibility in stroke onset (Markus 2011). Linkage analysis has been a very popular method for detecting genes of major effect in monogenic disorders, with these genes being relatively rare, relatively isolated to the disease population, and inherited in a Mendelian fashion. Linkage studies commonly examine large families where the disease affects individuals in several generations, with the aim of identifying a genetic marker that is always inherited by family members with the disease but not by those who do not have the disease.

In complex traits, linkage analysis can be more problematic because it may be difficult to develop a model that adequately explains the inheritance pattern. In these situations, the inheritance patterns of groups of loci (rather than individual loci) can be examined, and linkage can be established on a subset of a pedigree (rather than a pedigree as a whole)(Lander & Schork 1994). Factors such as complex inheritance, incomplete penetrance, locus heterogeneity

can be incorporated into an optimised model to improve statistical power and the quality of a study (Bush & Haines 2010). Linkage studies investigating the co-segregation of ischaemic stroke with genetic markers within family members have largely been unsuccessful for elucidating the genetic determinants of complex traits such as ischaemic stroke. Geneological data available from the deCODE group in Iceland implicated two genes for ischaemic stroke susceptibility: *STRK1*, coding for phosphodiesterase 4D, and *ALOX5P,* encoding 5-lipoqxygenase activating protein (Gretarsdottir et al. 2003; Helgadottir et al. 2004). However, these findings have failed to be replicated independently (Markus 2011). Importantly, they also indicate that the ischaemic stroke is a complex genetic disorder, rather than a monogenic disease (although monogenic forms e.g. CADASIL, Fabry Disease, do exist), requiring the examination of many genes that make small contributions to disease aetiology.

Given the limited success of linkage studies, until recently the mainstay for stroke genetics research, as with many complex diseases, has been the candidate gene association approach. These association studies test whether a particular genetic marker and trait occur more often than expected by chance alone, and are ideal for investigating complex disorders such as ischaemic stroke, which do not have a clear pattern of inheritance and are therefore not amenable to linkage analysis to determine their genetic basis. Therefore a number of genetic loci, each with small overall effect, can be explored. Association studies also have the advantage of relying on historical recombination events which can be examined in whole populations, as compared to linkage studies that are dependent upon pedigree information. As association analysis assumes that individuals are unrelated or have little population structure, any significant marker-trait associations can be mapped to fine genetic resolution within the kilobase range, rather than in the centimorgan range, as with linkage analysis.

To date, candidate genes have been investigated in pathways involved primarily in inflammation, endothelial function, homocysteine metabolism, hyperlidaemia and coagulation pathways. The most consistent associations found include methylenetetrahydrofolate reductase (*MTHFR*), angiotensin-converting enzyme (*ACE*), prothrombin and factor V Leiden (Bentley et al. 2010), but results from most candidate gene studies have been difficult to replicate due to small sample sizes, phenotype heterogeneity and publication bias in literature-based meta-analyses. Genome wide association studies (GWAS) represent a hypothesis-free approach that uses markers across the entire genome to examine association. Compared to linkage studies and candidate gene-based association studies, GWAS provide a more powerful strategy for detecting common variants with relatively small effects sizes by investigating very large population cohorts. The GWAS strategy is discussed in more detail in Section 1.10.

1.4.2 Modifiable Risk Factors

Modifiable lifestyle risk factors have become key targets for primary and secondary stroke prevention.

Smoking doubles the risk of stroke, and is particularly dangerous in combination with hypertension. It has been shown that smoking causes an immediate rise in blood pressure, while nicotine and other toxic by-products found in tobacco cause irreversible blood vessel damage (Shinton & Beevers 1989).

High-lipid diets can increase blood cholesterol levels and obesity can increase risk of stroke by 64% (Strazzullo et al. 2010), with central obesity (abdominal deposition of fat) being particularly related to the occurrence of atherosclerotic disease. Low density lipoproteins (LDL) and triglycerides can cause atherosclerotic plaque build-up, while high-density lipoproteins (HDL) can have a protective influence on extracranial carotid atherosclerosis (Heiss et al. 1991). Additionally, the INTERSTROKE study recently examined the association of apolipoproteins with stroke risk, concluding that the reduction in risk of ischaemic stroke conferred by HDL and apolipoprotein A-1 (ApoA-1) was larger than the risk associated with increased levels of apolipoprotein B (ApoB) or non-HDL cholesterol (O'Donnell et al. 2010). Exercise exerts a beneficial influence on risk factors for the atherosclerotic disease by raising HDL cholesterol and lowering LDL cholesterol and reducing blood pressure and weight. It has also been shown to delay the onset of diabetes by altering DNA methylation and increasing the expression of genes involved in glucose regulation (Barres et al. 2012).

1.4.3 Clinical Risk Factors

1.4.3.1 Diabetes and Hypertension

Diabetics have double the risk of stroke than non-diabetics, in part due to the fact that diabetes is also associated with the presence of other risk factors such as obesity, high cholesterol levels and hypertension (Emerging Risk Factors Collaboration et al. 2010). Hypertension is the most common modifiable risk factor for stroke, and both incidence and mortality increase at a blood pressure of >100/75mmHg. Furthermore, visit-to-visit variability in systolic blood pressure is a strong predictor of ischemic stroke independent of mean blood pressure (Rothwell 2010). The presence of hypertension has been found to predict the severity of carotid atherosclerosis through mechanisms including endothelial dysfunction, shear stress and increased smooth muscle proliferation, as well as exposing plaques to a greater risk of rupture (Beaussier et al.

2008). Patients with well controlled blood pressure (<120/80 mm Hg) have approximately half the lifetime risk for stroke in comparison to patients with hypertension (Mancia 2010).

1.4.3.2 Transient Ischaemic Attack

Population studies and randomised trials have demonstrated that 17% of TIAs occur on the day of stroke, 9% on the previous day and 43% at some point 7 days prior to stroke presentation (Rothwell & Warlow 2005). Given the high early risk of ischaemic stroke associated with TIA, tools such as the ABCD and subsequent ABCD^2 score have been developed for use in primary and emergency care settings to aid the prediction of stroke risk in the acute phase of TIA. The ABCD^2 score assesses patients based upon the following criteria: age (≥60yrs), blood pressure (≥140/90mm Hg on first evaluation), clinical symptoms (speech disturbance with/without unilateral weakness), duration (10 to 59 minutes, >60 minutes) and presence of diabetes. Incorporation of brain imaging findings into this model, in an updated ABCD^{2}I score, increases its predictive power, improving the area under the curve (AUC) from 0.66 with the ABCD^2 score alone to 0.78 in the combined model (Giles et al. 2010). Thus there is growing appreciation that effective diagnosis and management of TIA offers valuable opportunity to initiate treatment and stall the probable onset of ischaemic stroke and the associated brain injury and disability.

1.4.3.3 Heart Disease

Atherosclerosis is typically a systematic disease, and similar mechanisms underlie the formation of atheromatous lesions throughout the arterial tree. Stroke and myocardial infarction (MI) share a number of risk factors, and it has been shown that a family of stroke or MI is associated with a considerably increased prevalence of many risk factors in the proband (Kennedy et al. 2012). Subsequently, the same nine risk factors assessed in the INTERHEART study for risk of MI also predict risk of stroke in the INTERSTROKE study, despite a few differences (e.g. the ratio of apolipoprotein B to apolipoprotein A-1 was found to be the most important risk factor for MI, while hypertension was found to be the most important cause of stroke; Yusuf et al. 2004; O'Donnell et al. 2010).

1.4.3.4 Carotid Stenosis

Large artery stenosis accounts for around 20% of all ischaemic strokes (Grau et al. 2001). Carotid atherosclerosis leading to amaurosis fugax, TIA or ischaemic stroke is termed symptomatic, whereas atherosclerotic arterial narrowing without an ischaemic event is termed asymptomatic carotid stenosis.

Carotid artery atherosclerosis is a chronic process that shares similar risk factors to ischaemic stroke, including age, diabetes, hyperlipidaemia, hypertension and history of smoking. Meta-

analysis on the prevalence of carotid artery stenosis in the general population established that around 2% of individuals aged 60-69 years have asymptomatic carotid stenosis in which the degree of lumen narrowing exceeds 50%; this proportion increases to 7.5% in men and 5.0% in women aged of 80 years or older (de Weerd et al. 2010). In asymptomatic patients with >50% carotid artery stenosis, the annual risk of stroke has been estimated to be 1-4.3% (Inzitari et al. 2000; Nicolaides et al. 2005). Risk of stroke increases in those with progressive carotid artery stenosis (Inzitari et al. 2000), and progression to 80-99% stenosis or occlusion is more likely in patients whose initial stenosis is >50% (Johnson et al. 1995).

Pathophysiology of Carotid Atherosclerosis

The carotid artery is composed of three layers: the tunica intima, tunica media and tunica adventitia (Figure 1.3). The endothelial cells of the tunica intima play a key role in vascular permeability, platelet aggregation and resistance to thrombosis, and the ability of the endothelium to adapt and repair itself has a substantial impact upon the development of atherosclerosis. In response to physical or metabolic stress, changes in the expression of cell adhesion molecules (e.g. VCAM-1, ICAM-1) and other surface receptors causes endothelial cytoskeletal rearrangement and an increase in cell permeability. Subsequent platelet adhesion and migration of circulating cells beyond the endothelium induces smooth muscle cell proliferation within the tunica media as well as alterations of the extracellular matrix composition and organisation within this layer. The secretion of pro-inflammatory cytokines from the endothelium and formerly circulating cells attracts monocytes, T cells and macrophages (Hall & Bassiouny 2012). In particular, the aggregation of foam cells (lipoprotein-rich macrophages) beneath the endothelial layer results in 'fatty streaks', which are the first visible lesions in the development of atherosclerosis. Further thickening of the tunica intima and tunica media is termed carotid intima-media thickening (cIMT), and mean cIMT (although not cIMT progression) has been shown to be associated with an increased risk of MI and ischaemic stroke (Lorenz et al. 2012).

The progression of these early stages of carotid atherosclerosis into atheromatous plaque depends upon a number of metabolic, haemodynamic, environmental and genetic risk factors. Smooth muscle cell migration into the tunica intima and perpetuation of an immunological response within the arterial wall leads to the formation of a fibrous cap and the development of an atheroma, which is an active lesion producing cytokines and undergoing constant remodelling. Growth of the atheroma into the lumen initially leads to dilation of the artery to compensate for the reduction in luminal diameter (Glagovian remodelling) (Hall & Bassiouny 2012). However, this adaptation fails with continued encroachment of the lumen beyond 40-

50%, and further progression of the plaque can lead to focal arterial stenosis and potential occlusion.

The Vulnerable Plaque

It has been shown that incidence of cerebral ischaemic events is correlated with plaque morphology and composition, regardless of the degree of luminal stenosis (Yuan et al. 2001; Rothwell et al. 1996), thus promoting the concept of the 'vulnerable' or 'high-risk' plaque. Plaque vulnerability was first mentioned by Muller & Tofler (1992) to describe an atherosclerotic plaque that is vulnerable to rupture and subsequently cause an ischaemic event. Following the pioneering works of Virmani et al. (2000), Redgrave et al. (2008) and Naghavi et al. (2003a, 2003b), vulnerable carotid plaques are commonly characterised as follows: a thin fibrous cap of <200μm with an extensive loss of smooth muscle cells and extracellular matrix within the cap, overlying a large necrotic/lipid core that often comprises intraplaque haemorrhage, calcification and neovasculisation (Pelisek et al. 2012).

Inflammation has been noted as a crucial driving force behind plaque progression and instability. The majority of plaques are vascularised by a network of small vessels originating from the vaso vasorum of the tunica adventitia (Figure 1.3). Along with the formation of these neovessels (angiogenesis), the growth and widening of existing vessels (arteriogenesis) and progressive permeability of these vessels can allow inflammatory cells to pass the endothelium and accumulate within the plaque. A developing plaque contains mainly monocytes and macrophages, as well as other inflammatory cells such as T and B cells that secrete a plethora of inflammatory cytokines. Additionally, the presence of neutrophils within unstable lesions has been correlated with the site of plaque rupture. The secretion of proteolytic enzymes from these cells degrades the extracellular matrix and modulates inflammatory responses within the plaque (Naruko et al. 2002). Degradation of the extracellular matrix, in particular in the fibrous area overlying the lipid core (precipitating a thin fibrous cap), is the key mechanism propagating plaque destabilisation and lesion vulnerability. The process is largely mediated by matrix metalloproteinases (MMPs), which at first instigate intimal thickening but later lead to the production of new extracellular matrix products and extracellular matrix degradation (Pelisek et al. 2012).

While local factors such as sheer stress and vessel or plaque anatomy play an important role in plaque stability, there is strong evidence to suggest that systemic processes such as infection, autoimmunity and genetics are involved. Examination of patients with symptomatic carotid stenosis in the European Carotid Surgery trial (ECST) found that patients with plaque surface irregularity in one carotid artery were more likely than would be expected by chance to have an

irregular plaque in the other carotid artery and more likely than patients with smooth plaques to die as a consequence of myocardial infarction and other non-stroke vascular disorders (Rothwell et al. 2000). The grouping of unstable plaques within certain individuals reinforces the hypothesis that genetic factors can influence plaque progression and instability.

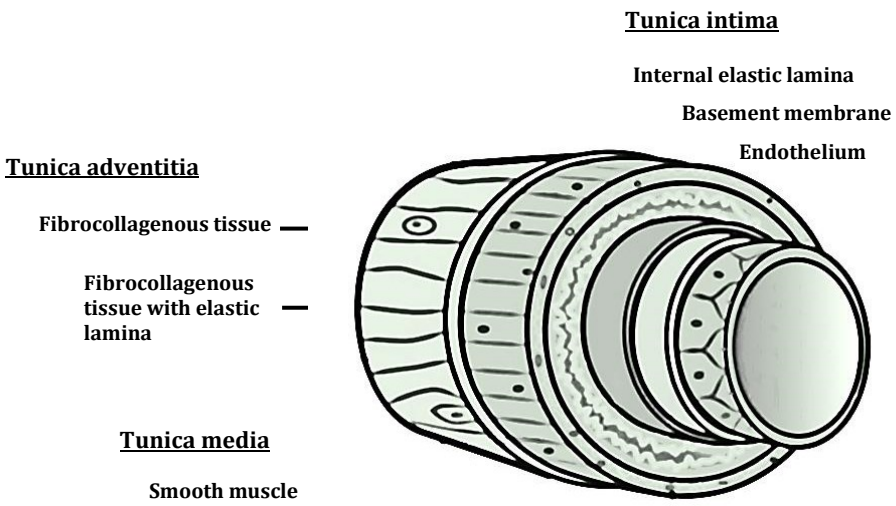


Figure 1.3: Schematic illustration of the various layers of the carotid arterial wall. The tunica intima is comprised of a monolayer of endothelial cells. Smooth muscle cells are found in the tunica media, and a network of small blood vessels comprising the vaso vasorum are located in the tunica adventitia.

1.5 Clinical Evaluation of Ischaemic Stroke

Risk prediction for stroke and other cardiovascular diseases has recently become a priority in public health policy, and future screening strategies will play an important role for primary prevention (discussed in Hasan 2013). Furthermore, a correct diagnosis of stroke impacts the admittance of a patient to a medical ward or stroke unit and the decision to administer thrombolysis, while assessment of prognosis is important for deciding whether to start the patient on medication to reduce the risk of future CVD events, and whether to consider nursing home care or a period of rehabilitation.

The clinical evaluation of ischaemic stroke risk, diagnosis and prognosis, and the associated challenges, are outlined below.

1.5.1 Risk Stratification

Risk stratification and prevention can be implemented on either a primary or secondary level. Primary prevention aims to prevent a first event of stroke or TIA by detecting and managing people with risk factors but no recognised disease. For those with known cerebrovascular disease, secondary prevention aims to reduce the risk for subsequent vascular events by treating CVD and the associated risk factors.

A recent statement from the American Heart Association and American Stroke Association emphasised the necessity for new policies for stroke to focus on primary prevention (Ovbiagele et al. 2013). In the UK, the NHS Health Check Putting Prevention First programme was introduced in 2009 as a high-risk screening policy for preventing CVD. This specified that all adults aged 40-74 years old should be screened for CVD risk, and those individuals found to exceed a 20% risk of a vascular event within the next decade should be treated with education on lifestyle choices as well as medication (anti-platelets such as aspirin, anti-coagulants such as warfarin, cholesterol reducing drugs such as statins or anti-hypertensives to lower blood pressure) as appropriate..

1.5.1.1 Assessment of Carotid Stenosis and Plaque Stability

In addition to age, life expectancy and quality of life, both primary and secondary risk stratification of ischaemic stroke has focussed upon the assessment of carotid artery disease.

A multicentre natural history study, the Asymptomatic Carotid Stenosis and Risk of Stroke (ACRS) study, was conducted to identify patients with asymptomatic carotid stenosis at high risk of stroke, and confirmed the relationship between cerebral ischaemic events and the degree

of carotid stenosis (Nicolaides et al. 2005). While treatment of carotid artery stenosis by stenting or endarterectomy (CEA) substantially reduces stroke risk, it is also accompanied by surgery-related morbidity and mortality (Doig & Brown 2012). As the Asymptomatic Carotid Surgery Trial (ACST) highlighted, there is a need to identify a high-risk subgroup of asymptomatic individuals that would benefit from intervention (Halliday et al. 2010).

The identification of vulnerable carotid plaques that can predict symptomatic cerebrovascular events has been the subject of much research and accurate, non-invasive methods that predict plaque progression and instability remain an unmet clinical need. Effective prediction of individuals at high risk of ischaemic stroke or TIA could offer personalised patient management, aid in the decision to intervene via carotid stenting or endarterectomy (Doig & Brown 2012) and provide a new surrogate endpoint for plaque stabilising therapies. Risk stratification will most likely rely on a combination of clinical parameters, carotid artery imaging and molecular tests to provide comprehensive assessment of both stenosis and plaque stability.

The main structural predictor of ischaemic stroke is the degree of luminal stenosis, as assessed in the ECST and NASCET trials (European Carotid Surgery Trialists' Collaborative Group [ECST] 1991; North American Symptomatic Carotid Endarterectomy Trial Collaborators [NASCET] 1991). Both trials imaged carotid arteries using conventional angiography, first with conventional radiographic film and later with digital subtraction angiography. However, substantial differences have been reported between the NASCET and ECST angiographic methods, such that 50% stenosis by NASCET is equivalent to approximately 70% stenosis by ECST. Less invasive carotid-imaging techniques, such as CT and MR angiography, and Doppler ultrasound are now more commonly used. Doppler ultrasound records sound waves to assess speed of blood flow through the artery. Peak systolic velocity and the ratio between internal carotid artery peak systolic velocity and common carotid artery end diastolic velocity are used to estimate % stenosis per NASCET method, although Doppler velocity measurements are known to overestimate or underestimate stenosis in certain situations (Oates et al. 2009).

The lack of reliable correlation between the widely used non-invasive carotid artery imaging modalities and the invasive catheter angiography used in NASCET and ECST upon which degrees of stenosis are defined makes assessment of carotid stenosis alone inadequate for determining surgical intervention. Furthermore, it belittles the role of plaque characteristics and stability in causing the ischaemic event (Ahmed et al. 2010). Plaque imaging techniques including ultrasound, positron emission tomography, MRI and multidetector CT angiography can characterise vulnerable carotid plaque features (inflammation, intraplaque haemorrhage, fibrous cap thickness, large lipid-rich necrotic core) as well as luminal stenosis. In addition,

diagnostic methods such as thermography, spectroscopy or molecular imaging, capable of detecting plaque physiology, would help to identify lesions of instability based upon activity, rather than morphology and stenosis alone (Naghavi et al. 2003a). However, at present the efficacy and reliability of these modalities remains unclear.

1.5.2 Diagnosis

Stroke diagnosis and its likely prognosis are commonly determined by clinicians with an assessment of the clinical features of the presenting syndrome, the patient's history and brain imaging. The rapid development of pathophysiology means that prompt diagnosis of ischaemic stroke is critical for preventing large-scale brain damage. Early intervention is crucial, with approximately 2 million neurons lost per minute after stroke onset (Saver 2006).

A diagnosis of ischaemic, rather than haemorrhagic, stroke appropriates the administration of thrombolysis, anticoagulants such as heparin to prevent recurrent cardiac embolism in patients with atrial fibrillation, anti-platelets such as aspirin to prevent recurrent vascular events and carotid artery imaging to examine stenosis. In an age of interventional stroke therapies, identifying stroke mimics (disorders that have a presentation resembling AIS or TIA) is also necessary to avoid potential adverse effects. Common stroke mimics include seizures, sepsis, syncope, brain tumours and metabolic disorders, and can account for up to a third of patients with stroke-like symptoms (Hand et al. 2006). Differentiating between stroke mimics and ischaemic events is becoming increasingly important, especially between stroke mimics and TIAs, which carry their own diagnostic challenges, as discussed earlier

1.5.2.1 Clinical Scales

At present, stroke diagnosis is made upon probability rather than certainty, and initial assessment of stroke patients is mostly performed by paramedics or triage nurses rather than stroke specialists who could make a more accurate analysis of symptoms and imaging results. With this in mind, a number of clinical scales have been developed to aid non-specialists. In the UK, the Stroke Association's Act FAST (Face Arm Speech Test) is reported to have a positive predictive value (the likelihood that a patient testing positive actually has stroke) of 78% when used by ambulance staff (Harbison et al. 2003). The ROSIER (Recognition of Stroke in the Emergency Room Scale) scale, designed for emergency department staff, consists of a 7 item scoring system based on asymmetrical facial, arm or leg weakness, speech disturbance, visual field defect, evidence of syncope and seizure.

1.5.2.2 Brain Imaging

Brain imaging has become a fundamental tool in stroke diagnosis. Due to its accessibility and speed, CT has been the mainstay of stroke imaging. Although CT imaging is able to exclude haemorrhagic stroke with near 100% sensitivity, less experienced observers are liable to overlook haemorrhage or mistake non-pathological calcification as blood (Schriger et al. 1998). Furthermore, CT imaging can often present as normal in acute stages of AIS, and 50% of infarcts never become visible on CT (Warlow et al. 2003). Thus it is assumed by deduction, rather than certainty, that if a stroke symptom is not caused by haemorrhage it is likely to be caused by ischaemic stroke, even if the CT does not detect a lesion. A recent meta-analysis of seven studies estimated that CT had a sensitivity of 39% (95% CI, 16% - 69%) and specificity of 100% (95% CI, 94% – 100%) for AIS. In contrast, the summary estimates for diffusion weighted magnetic resonance imaging (DW-MRI) showed a sensitivity of 99% (95% CI, 23% – 100%) and specificity of 92% (95% CI, 83% – 97%) (Brazzelli et al. 2009). Thus DW-MRI is a more reliable tool for diagnosis, but has the disadvantage of being expensive and unavailable in most emergency departments. Furthermore, there are many cases where the clinical diagnosis is certain and the MRI is negative, as well as the converse, so MRI is an inadequate diagnostic gold standard for acute ischaemic stroke (Jensen et al. 2008). Both CT and MRI are disadvantaged by the fact that their implementation and interpretation expends considerable time and effort. Whichever imaging method is used, radiologists require knowledge of the time and date of stroke in order to interpret the images correctly.

1.5.2.3 Laboratory Tests

In patients presenting with sudden onset focal neurological deficit, point of care testing consists of a basic metabolic panel test (glucose, calcium, sodium, potassium, chloride, bicarbonate, blood urea nitrogen and creatinine), complete blood count, cardiac enzymes and coagulation studies. Once the patient has stablilised, lipid panel and haemoglobin A_{1C} testing can be used to identify risk factors. Further testing e.g. to assess for a hypercoagulable state or genetic, metabolic or inflammatory disturbances may be performed if aetiology remains uncertain.

1.5.3 Prognosis

Around 30% of stroke patients die within a year of stroke, and nearly half of survivors are left dependent (Warlow et al. 2003). Reliable predictions about stroke outcome can guide decisions about medical treatment and patient management as well as allowing assessment of the performance of stroke services.

Just as cardiologists have learnt that the most important criteria for determining the prognosis of myocardial infarction are the nature, location and severity of coronary artery disease, so the cause of brain ischaemia (cardiac, small or large vessel) determines stroke outcome and treatment. In clinical practice, the evaluation of stroke prognosis is usually based upon the judgement of the assessing doctor. However, evidence has shown that clinicians are often over-optimistic about recovery, with only 65% of patients who are predicted to recover gaining independence by one year after stroke (Counsell et al. 2004).

1.5.3.1 Outcome Scales

In order to provide clinicians with more accurate prognostic information, a number of outcome scales have been devised. The most widely used clinical scale is the National Institute of Health Stroke Scale (NIHSS), consisting of a 15-item assessment of consciousness, gaze, vision, facial palsy, arm and leg strength, limb ataxia, sensory loss, neglect, dysarthria, and aphasia. The modified Rankin Scale (mRS) is commonly used to assess disability and dependence in the months following stroke, consisting of a 6-point scale from 0-5: patients with no impairment receive a score of 0, while patients with severe disability who are bed ridden and require constant nursing care and attention receive a score of 5 and death is rated as 6. Similarly, the Barthel index scale measures performance in activities of daily living, using 10 items including feeding, bathing and dressing and mobility to score a patient's independence out of a total of 100 points. Models based upon a number of variables such as age, arm power, independence before stroke, verbal ability and mobility can also provide reliable estimates of stroke prognosis (Counsell et al. 2002).

1.5.4 Prognosis after Thrombolysis

Prognostic tools can also be used to identify patients most likely to benefit from early therapeutic intervention. The aim of thrombolytic therapy is to break down the thrombus associated with AIS, providing reperfusion before permanent tissue damage. Intravenous recombinant tissue plasminogen activator (rt-PA) has revolutionised acute stroke management since approval by the US Food and Drug Administration (FDA) in 1996, and is the standard thrombolysing agent used in the US, Canada, Australia and European Union. Its arrival has

demonstrated that acute ischaemic stroke is treatable, and furthermore that reperfusion can dramatically improve patient prognosis.

Although rt-PA is the only FDA-approved therapy specifically directed at acute ischemic stroke, there are a number of issues and controversies surrounding its use (Riggs 1996). Whilst 50% of patients receiving intravenous rt-PA achieve complete or partial recanalisation of the middle cerebral artery, a third of these patients also experience neurological worsening or delayed ischaemic injury (Grotta et al. 2001; Kidwell et al. 2002; Moldes et al. 2008). Additionally, intra-cerebral haemorrhage (ICH) represents one of the most dangerous complications associated with thrombolytic therapy. These risks have greatly narrowed the number of patients being administered rtPA, therefore prediction of outcome following thrombolysis will aid clinicians to administer therapy successfully and greatly improve the number of patients receiving treatment.

1.6 Use of Biomarkers for Improving Patient Management

A biomarker has been formally defined as "a biological characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention" (Biomarkers Definitions Working Group 2001). The FDA states that an ideal biomarker should be highly sensitive (very few false negatives) and specific (very few false positives), predictive, able to bridge preclinical and clinical trials and be easily measurable. The interest in the applications of biomarkers has increased greatly in recent years and they are now considered to be one of the cornerstones of the modern drug discovery process in both preclinical and clinical studies (Feuerstein & Chavez 2009).

In several disorders, the delineation of biomarkers from normal biology causes them to be either newly expressed or increase or decrease in response to pathological stimuli. Thus, biomarkers can be effective indicators of disease mechanisms and progression. They have the potential to serve as diagnostic tools, prognostic aids for risk stratification and therapeutic intervention targets (Figure 1.4).

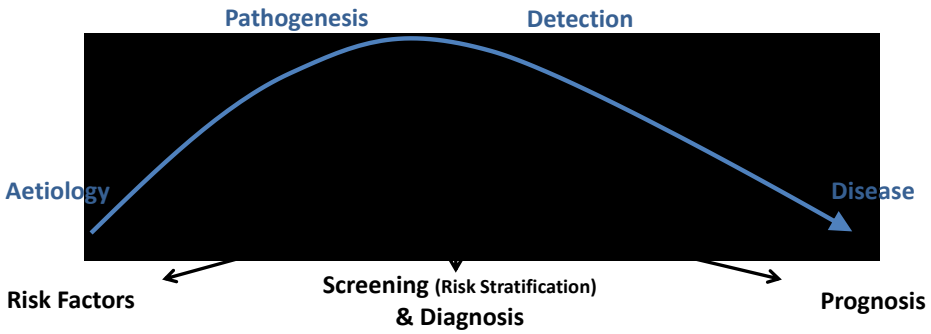


Figure 1.4: Pathway of disease progression and potential impact of biomarkers. Biomarkers can be utilised at many different stages of disease progression, to aid with risk stratification, diagnosis and prognosis. Image modified from Mayeux (2004).

1.6.1 Specimen Selection

Biomarker measurements can be made directly on biological samples (e.g. blood, tissue, urine) or via imaging to detect changes in the brain. Specimen selection is a critical process in biomarker discovery and is largely dependent upon the aim of the study in question. If the aim is to discover circulating biomarkers of disease, blood samples can be studied; if the aim is to elucidate biomarkers for the mechanisms of disease progression, tissue extracts may be preferred. Microdialysis for continuous sampling of the interstitial fluid of tissues is another technique that has predominantly been used in neurointensive care for monitoring the progress of brain ischaemia.

Plasma and serum profiles are very interesting from a clinical point of view as they are able provide data about the physiological state of different tissues within the body, their relative merits being subject to debate (Rai et al. 2005). Plasma is obtained from the centrifugation of blood collected in an anticoagulant solution. On the contrary, serum is obtained after coagulation. The formation of a clot of activated leukocytes and platelets causes the release of proteins that can lead to considerable differences between plasma and serum profiles (Blanco-Colio et al. 2006). The complexity and large concentration ranges of proteins within these samples present substantial analytical challenges.

Disruption of the blood-brain barrier during stroke causes the release of brain-derived molecules into the bloodstream, but the variability of perfusion through the disrupted area could mean that biomarker levels become substantially altered in plasma and serum levels long after management decisions need to be made. The cerebral spinal fluid (CSF) is derived from the

choroid plexus, blood vessels of the arachnoid and the ependymal lining of the ventricular system and glial membrane (You et al. 2005), and is therefore a good reflection of biochemical changes occurring within the nervous system.

However, CSF sample collection by lumbar puncture is invasive and entails a number of complications, with potential blood contamination during collection altering protein profiles and hindering the accurate identification of biomarkers (You et al. 2005). Furthermore, the composition of normal CSF differs considerably from the composition of extracellular fluids in the brain, including microdialysate (Hamberger et al. 1990). A number of studies have shown a lack of correlation between biomarker levels in the CSF and serum (Petzold et al. 2008). This could be due to the time taken for molecules to enter the blood following stroke onset, or differences in the biological half-life of some proteins and metabolites between CSF and blood. CSF proteins are generally less abundant than their corresponding plasma and serum counterparts, necessitating the use of more sensitive analytical techniques. Consequently, potential biomarkers identified from either fluid need to be treated with caution until their validity can be confirmed in the other.

In the context of stroke biomarker research, brain tissue or atherosclerotic plaque samples arguably have the advantage of being the most useful for preliminary detection of potential disease-specific markers (Blanco-Colio et al. 2006). Tissue samples represent an ideal medium for investigating the tissue-specific mechanisms underlying ischaemic stroke e.g. for risk biomarkers of carotid plaque rupture, or mechanistic biomarkers of ischaemia-related changes within the brain. However the collection of these samples is difficult and obtaining samples from a healthy comparison group is often impractical. In contrast, plasma and serum samples are more convenient to collect and can be used practically for non-invasive, point-of-care diagnostic biomarker tests.

1.6.2 Multi-Marker Tests

The clinical utility of a biomarker is defined by its ability for classification, namely its sensitivity and specificity. Disorders such as stroke are complex and multi-factorial, therefore a single biomarker is may not have adequate sensitivity and specificity for a clinical test. Combination of markers in a 'signature fingerprint' for disease, constituting a multi-marker test, has shown to increase the power of discrimination between study populations (Gillette et al. 2005; Woolas et al. 1995). A number of studies have attempted this approach in stroke biomarker research (e.g. Laskowitz et al. 2005; Montaner et al. 2005; Reynolds et al. 2003; Tuttolomondo et al. 2008), but thus far none have been suitably successful.

1.7 Approaches for Biomarker Discovery

Biomedical research to date has been largely candidate-driven, with experiments being designed to test whether a hypothesis is true or false. These hypothesis-based, directed studies have been successful in identifying a number of important biomarkers for disease, for example troponins for the detection of cardiac injury (Babuin & Jaffe 2005), and CD340 as a prognostic biomarker and anti-cancer therapy target in breast cancer (Luftner et al. 2003; Baselga et al. 1998). The majority of biomarkers implicated in stroke thus far have also been identified using these candidate-based approaches.

Although candidate-based approaches have yielded much information about individual components of a system, they have been less successful at providing an integrated understanding of the structure and dynamics of the complete gene or protein system being studied, and the compensatory responses of the system to perturbations (Alawieh et al. 2012). In 1928, the physicist Niels Bohr suggested that a particular experiment can only reveal a certain kind of information; other information that is equally important cannot be measured simultaneously and is lost. His student Max Delbrück applied this 'principle of complementarity' to biology, proposing that the function of molecules such as proteins can only be fully understood when they are studied at a whole system level, taking into account their interaction with other biological molecules.

The sequencing of the human genome and recent advances in high-throughput technology have created a paradigm shift in molecular biology and laid the foundations for the development of -omics approaches that can interrogate entire gene, RNA or protein systems. This new approach aims to better map the functionality of complex biological systems for insights into the pathophysiology of disease, increasing diagnostic, prognostic and disease-monitoring potential for clinical applications (Dominiczak et al. 2010). Thus proteomic, transcriptomic and genomic technologies provide new platforms for unbiased hypothesis testing in biomarker discovery (Figure 1.5).

1.7.1 Data Handling for Biomarker Discovery

While –omics-based approaches have significant scientific potential, the difficulty lies in extracting valuable biological information from the mass of data generated, and translating this into clinically useful biomarker tests. The experiments collect large amounts of data in the absence of hypotheses concerning particular candidate molecules, therefore, they must be well designed in order to take advantage of statistics for data interpretation and maximise the likelihood of new discoveries and new testable hypotheses. Indeed, the results of any -omics study is largely dependent upon the quality of the data used for analysis and the statistical approaches employed (Tworoger & Hankinson 2006).

Dimensionality reduction is often performed in order to reduce 'noise' that can confuse downstream analysis, and facilitates the use of conventional statistical approaches that are often limited by the number of variables that can be examined. This can be performed by selecting variables that provide the most statistically significant differences between study groups (through use of e.g. Student's t-test or analysis of variance), or by clustering variables that correlate with each other (through use of e.g. principal components analysis).

The association between a biomarker and disease is traditionally assessed by epidemiologists using regression analysis to calculate effect size (e.g. odds ratio or standardised mean difference). While individual primary studies provide information about the research question being studied, results may be conflicting and therefore it becomes difficult to draw firm conclusions about the scientific progress made. A systematic literature review (qualitative analysis) can be conducted to gather all empirical evidence fitting pre-specified eligibility criteria in order to answer a specific research question e.g. the association of a biomarker with disease. Subsequent meta-analysis can be conducted to statistically combine results from two or more independent studies by calculating pooled effect sizes (e.g. pooled odds ratio, pooled standardised mean difference) to obtain a quantitative association between a biomarker and disease phenotype. Thus meta-analysis can establish the direction and size of the effect of association through increasing power and precision by combining all primary studies, answer questions not posed by individual primary studies, address controversies arising from conflicting results of individual studies and generate new hypotheses. By using clear, systematic methods to minimise bias, literature review and meta-analysis therefore present a balanced summary of results to inform clinical practice and policy development (Higgins & Green 2011).

While measures of association using odds ratios and standardised mean differences are useful, biomarkers proposed for discrimination of individual subjects can also be assessed for their

sensitivity and specificity. An ideal biomarker classification model yields 100% discrimination between study groups, but this is often unachievable in reality. The receiver operating characteristic (ROC) curve is able to describe the whole set of potential combinations of sensitivity and specificity for a continuous biomarker, thus enabling the selection of an optimum threshold value for discrimination between groups. While effect sizes calculated via regression analysis are interpreted per unit increase in the biomarker and therefore may not be comparable between biomarkers measured in different units, the ROC curve provides a natural common scale for comparing different biomarkers, regardless of the units of measurement (Cook 2007). The area under the curve (AUC) generated by ROC analysis is often used to assess the accuracy of a biomarker model, where an AUC of 1 represents 100% discrimination and values of 0.8 and above are deemed good (Pepe 2004).

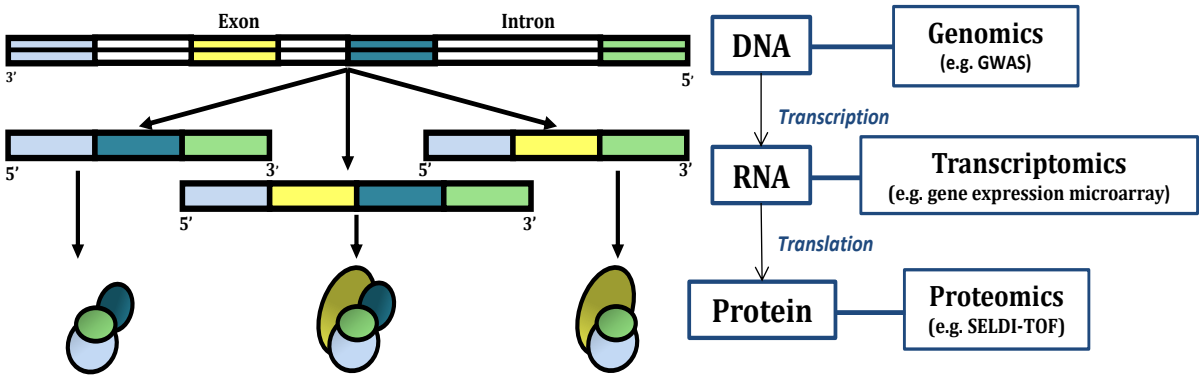


Figure 1.5: Summary of the main disciplines in systems biology. The genetic information contained within a genome can be investigated using genomics approaches e.g. GWAS. DNA sequences are converted into RNA via a process of transcription, with introns being spliced out and exons retained, in various combinations (alternative splicing). RNA gene expression and alternative splicing profiles can be interrogated using transcriptomics approaches. RNA sequences are converted into proteins via a process of translation. Quantification and characterisation of proteins can be achieved by proteomics approaches e.g. SELDI-TOF MS. GWAS, genome wide association study; SELDI-TOF MS, surface enhanced laser desorption ionisation time of flight mass spectrometry.

1.8 Proteomics

The completion of the Human Genome Project showed that given the relatively small increase in genes through evolution, diversity in organisms is largely determined by protein regulation. While genomic and transcriptomic techniques have been successful in the discovery of countless biomarkers, they are limited in the amount of information they can provide about the molecular dynamics associated with acute neurological insults.

The proteome can be defined as all the proteins present in a cell or organism at a given time, including those translated directly from genetic material as well as a spectrum of proteins arising from alternative splicing of RNA transcripts and post-translational processing (Arrell et al. 2001). Thus the major advantage of proteomics over gene based technologies is that it directly investigates the functional molecules and not the source code. Proteins commonly have short half-lives and are constantly synthesised and degraded as required, altering physiology via their interactions with protein and non-protein substrates and thus can providing information about the molecular dynamics associated with acute insults.

Due to the dynamic nature of protein regulation, the study of the proteome can be technically very challenging. Proteomics aims to identify, characterise and understand these protein changes. Proteomic investigations can be divided into three distinct categories, depending upon the aim of the study. Structural proteomics aims to provide 3-dimensional information for all proteins within a cell or organelle. Expressional proteomics compares relative protein quantities between samples from different physiological states e.g. normal versus diseased, while functional proteomics aims to classify protein changes such as post-translational modification or protein-protein interaction occurring during certain biological conditions (Karvunidis et al. 2009).

The presence of a large number of proteins in blood makes human plasma an excellent material for discovering biomarkers for potential clinical diagnostics and therapeutics. However, the complexity and large concentration ranges of proteins within plasma samples present major challenges for their analysis. A handful of highly abundant proteins, such as albumin, represent over 90% of the plasma proteome and can mask the detection of low-abundance proteins that are likely to be more valuable as clinical biomarkers (Anderson & Anderson 2002). Furthermore, the complete measured dynamic range of proteins spans more than 10 logs of molar abundance (Hortin & Sviridov 2010). As a result, affinity chromatography or membrane filtration is often performed before sample analysis, in order to deplete high-abundance proteins. Additionally, samples can undergo multiple fractionation steps in order to simplify the proteome and maximise protein separation and resolution. Physicochemical approaches to

fractionation include alcohol precipitation and ultracentrifugation, while methods exploiting specific protein characteristics, such as their inherent chemical properties e.g biospecificity, hydrophobicity or charge can also be utilised. Sample preparation is critical for accurate proteomics analysis (Tam et al. 2004).

1.8.1 Methods for Proteomic Analysis

With recent advances in separation techniques and mass spectrometry (MS), proteomic approaches have flourished and have allowed the unequivocal identification and quantification of molecules that could not have been detected by classical techniques such as immunoassays, making proteomics a promising approach for biomarker identification.

Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) has played an invaluable role in proteomics and has been a key method of protein identification in the majority of early proteomics investigations. Samples are spotted on separate gels, and proteins are resolved in two sequential steps: in the first dimension separation involves isoelectric focusing (IEF), by which proteins are focused by relative charge according to their inherent charge, or by isoelectric point (pI). Proteins are then resolved in the second dimension by their molecular weight. After separation, staining (e.g by Coomassie, zinc, silver staining, radiolabelling or fluorescence via 2D-DIGE) is carried out to analyse protein expression. Protein levels are compared between samples, then spots of interest can be excised, enzymatically digested, and the peptides are analysed using MS. Although 2D-PAGE technology has greatly improved in recent years, limitations of the technique still include a lack of sensitivity and dynamic detection range, slow throughput and low reproducibility.

The substitution of 2D-PAGE with liquid chromatography based high-resolution MS has made proteomics a convincing approach to biomarker eludcidation. Mass spectrometry is now at the heart of nearly all proteomics experiments and offers unparalleled sensitivity for the investigation of protein profiles and provides the possibility of rapid detection of conditions such as acute ischaemic stroke, in which early diagnosis and intervention determines the prognosis of clinical outcome (White et al. 2006).

MS consists of an ion source, a mass analyser that measures the mass to charge ratio (*m/z*) of the ionised analytes, and a detector that registers the number of ions at each *m/z* value. Gel-free techniques, unlike 2D-PAGE, allow the direct coupling of the LC system to the mass spectrometer so that the effluent from the column (separated species) can be detected and characterized by MS/MS. In the ‘top-down’ approach, intact protein isoforms can be isolated

before MS identification; in the 'bottom-up' approach, protein mixtures are enzymatically digested into complex peptide mixtures before being subjected to tandem MS. Electron spray ionisation (ESI) ionises the analytes out of a solution and is therefore easily coupled to liquid-based separation tools such as high pressure liquid chromatography (HPLC), in which peptides are separated according to hydrophobicity. Subsequently, within the MS, ions with specific *m/z* ratios can be selected in a 'trap' for collision induced fragmentation. Fragmentation of the peptide ion, commonly at the peptide bond, produces the tandem mass spectrum from which an amino acid sequence for the peptide can be deduced (Guerrera & Kleiner 2005).

Four basic types of mass analyser are currently used in proteomic research, each with their own advantages and disadvantages: the ion trap, the quadrupole, the Fourier transform and the time-of-flight (TOF) analysers (Aebersold & Mann 2003). Matrix-assisted laser desorption/ionisation (MALDI) and surface-enhanced laser desorption/ionisation (SELDI) technologies are both typically coupled to TOF analysers.

1.8.1.1 SELDI-TOF Mass Spectrometry

SELDI-TOF MS is an affinity-based technology first described by Hutchens and Yip in 1993 and developed into the ProteinChip system by Ciphergen Biosystem Inc (Fremont, CA, USA). It couples chromatographic separation using reactive surfaces with high-throughput mass spectrometry, encompassing two key subsets of MS technology: surface enhanced neat desorption and surface enhanced affinity capture (Hutchens & Yip 1993).

The ProteinChip array distinguishes SELDI-TOF technology from other proteomics platforms, with chromatographic capture of proteins occurring on the same support as that used for laser desorption, hence simplifying the SELDI-TOF procedure and assisting automated analysis to increase reproducibility (Ciphergen Biosystems).

SELDI-TOF MS is a variation of MALDI-TOF MS, with target surfaces to which the proteins and matrices are applied to being modified for various conditions such as anion exchange, cation exchange and pH. Thus classes of proteins can be fractionated on the chip prior to analysis. The combination of selective protein binding with quantitative mass detection makes SELDI-TOF MS a simple and attractive method for comparative protein analysis.

1.8.2 Proteomics and SELDI-TOF MS for Ischaemic Stroke Biomarker Discovery

While several studies have employed immunoassays for the detection of diagnostic biomarkers for acute ischaemic stroke (Reynolds et al. 2003; Lynch et al. 2004; Laskowitz et al. 2009; Montaner et al. 2012), the use of proteomics technology in stroke research is relatively uncommon. To date, proteomics technology has been used in ischaemic stroke research to investigate brain tissue and cells (Cuadrado et al. 2009; Cuadrado et al. 2010; Brea et al. 2011), urine (Dawson et al. 2012), CSF (Zimmermann-Ivol et al. 2004; Lescuyer et al. 2004; Dayon et al. 2008; Allard et al. 2005), microdialysate (Maurer et al. 2003; Dayon et al. 2011), plasma (Ning et al. 2010; Lopez et al. 2012) and serum (Huang et al. 2009). These studies have predominantly focussed upon diagnosis of acute ischaemic stroke against haemorrhagic stroke or healthy controls. Many investigations for stroke biomarkers have focussed upon CSF and microdialysate as these biofluids are considered to be less complex to work with; studies investigating blood have commonly employed have typically employed technically demanding methods of analysis, based upon fractionation, multiple stages of separation and mass spectrometry in order to expose meaningful proteomic data from the plasma or serum samples (Kodali et al. 2012; Lopez et al. 2012; Brea et al. 2009).

Preliminary investigations into acute and chronic pathologies by the Lescuyer group identified 13 potential proteins in post-mortem CSF for the diagnosis of ischaemia and neurodegeneration (Lescuyer et al. 2004). Subsequent studies confirmed that levels of PARK7, Nucleoside Diphosphate Kinase A (Allard et al. 2005) and Brain-Type and Heart-Type Fatty-Acid Binding Proteins (Wunderlich et al. 2005) are elevated in early serum and plasma samples from patients with stroke. Plasma validation of proteins expressed in post-mortem CSF demonstrated that ubiquitin fusion degradation protein, RNA-binding protein and nucleoside diphosphate kinase A have high specificity and sensitivity for ischaemic stroke in a cohort of over 600 patients (Allard 2004).

The data obtained from these studies demonstrate that proteins involved in ischaemic stroke pathogenesis can be detected using proteomic techniques. However, they are largely limited to media that are invasive and not amenable to rapid point of care testing (e.g. CSF, microdialysate), and complicated experimental techniques that can result in loss of proteomic information at each level of fractionation, separation and analysis.

By combining fractionation and mass spectrometry in one platform, SELDI-TOF MS provides an ideal method for high-throughput analysis of complex plasma samples. Employing a proteomic approach using SELDI-TOF, 1D gel electrophoresis, nano LC-ESI-MS/MS and subsequent ELISA, Allard et al. (2004) categorised plasma proteins that could distinguish between acute ischaemic

and haemorrhagic stroke. The aim of the study was to identify a biomarker which could improve stroke management, and ideally fulfill the following criteria: be a) brain-specific, b) a plasmatic marker as these samples are easiest to obtain c) detectable as early as possible after symptom onset d) expressed proportionally to the extent of the lesion e) beneficial for prognosis f) able to distinguish ischaemic from haemorrhagic stroke. Analysis revealed that apolipoprotein C1 (ApoC-I) and apolipoprotein CIII (ApoC-III) could provide good predictive information of ischaemic versus haemorrhagic stroke (94% sensitivity and 73% specificity for ApoC-I and 94% sensitivity and 87% specificity for ApoC-III). Using SELDI-TOF MS and CM10 ProteinChips to analyse plasma samples from acute ischaemic stroke patients and hospitalised controls, Zhang et al. (2008) detected 13 protein ions of significance, 10 of which were down-regulated and 3 up-regulated in patients with acute ischaemic stroke. The authors concluded that a diagnostic test for would likely consist of a combination of biomarkers in a multi-marker model, reflecting the multi-factorial roles of inflammation, atherosclerosis, thrombosis, ischaemia and neurodegeneration in the pathogenesis of ischaemic stroke. In another study, Staals et al. (2008) investigated protein profiles of lacunar stroke patients with a single symptomatic lacunar lesion compared with those with multiple 'silent' lacunar lesions and extensive white matter lesions. They found that a 16kDa protein ion upon SELDI-TOF MS, identified as the alpha-2-chain of haptoglobin, was over-expressed in single symptomatic lacunar strokes, and speculated that the differences in alpha-1 or alpha-2 protein expression could be due to polymorphic differences between the study groups. Table 1.1 summarises the findings from these studies.

To date none of the candidate biomarkers identified through SELDI-TOF MS analysis of blood samples have entered clinical practice for the diagnosis of ischaemic stroke. Nonetheless, the studies demonstrate that SELDI-TOF MS is an effective strategy for high-throughput blood proteome analysis.

Table 1.1: Summary of SELDI-TOF MS proteomics investigations of acute ischaemic stroke.

Study	Disease State	Biological Sample	Differentially Expressed Protein Ions
Staals et al 2008	First-Ever Lacunar Stroke: Single Symptomatic Lacunar vs. Multiple Additional 'Silent' Lacunar Lesions	Serum	Alpha-2-chain of Haptoglobin
Zhang et al 2008	AIS vs. Hospitalised Controls (Otorhinolaryngology Department)	Plasma	2.71kDa, 3.42kDa, 4.29kDa, 4.36kDa, 6.71kDa, 6.99kDa, 8.46kDa, 8.53kDa, 8.56kDa, 8.64kDa, 9.13kDa, 22.80kDa, 24.36kDa
Allard et al 2004	AIS vs. ICH and Hospitalised Controls (Orthopedic Department)	Plasma	Apolipoprotein CI, Apolipoprotein CIII, Serum Amyloid A, Antithrombin-III Fragment

1.9 Transcriptomics

While most cells in the body contain the same genome, regardless of cell type or environmental conditions, the transcriptome (the complete set of RNA transcripts produced by the genome at any one time) is highly dynamic and varies under disease conditions or environmental stress. Transcriptomics, the study of the whole transcriptome, provides a strategy for exploring gene expression and the genetic mechanisms underlying disease pathogenesis.

Transcription describes the creation of a complementary RNA molecule from a DNA sequence, and transcription of genes to produce messenger RNA (mRNA) is a primary step in gene expression. RNA is synthesised in a 5' to 3' direction by the enzyme RNA polymerase. The process is naturally modulated by the binding of regulatory proteins to DNA binding sites, as well as epigenetic regulation e.g. via DNA methylation, which affects chromatin folding and thus the accessibility of genes for transcription. There is also growing evidence that a vast proportion of the RNA manufactured from the human genome does not code for proteins; rather, many types of RNA molecules exist that are not intermediates between DNA and protein, as classically thought. In particular, microRNAs are non-coding RNAs known to control transcription and translation to decide a tissue-specific protein repertoire to actively regulate homeostasis, and have been implicated in plaque instability and stroke pathogenesis (Bidzhekov et al. 2012; Haver et al. 2010; Vemuganti 2010).

Following transcription and prior to being capped with a poly(A) tail at the 3' end, eukaryotic RNA is commonly spliced to remove non-protein coding introns and stitch together exons in various combinations, creating splice variants that code for diverse protein isoforms (Figure 1.5). While early estimates stated that around 100,000 genes were required to code a mammalian genome, the actual number is less than a quarter of this. Alternative splicing is now known to account for a much of this 'missing information' and plays a key role in regulating gene expression (Nilsen & Graveley 2010). It is estimated that over 90% of human genes express multiple RNA transcripts due to alternative RNA splicing (Wang et al. 2008), and these transcriptional isoforms can alter protein products associated with the manifestation of disease (Cline et al. 2005).

1.9.1 Methods for Transcriptomic Analysis

Gene expression can be measured by quantifying mRNA abundance. Northern blotting was first used to directly quantify steady-state levels of mRNA by electrophoresis and subsequent incubation with specific probes. Real-time quantitative polymerase chain reaction (qPCR) remains the most accurate method for quantifying specific nucleic acid targets and is commonly employed following reverse transcription of mRNA to complementary DNA (cDNA).

Following the development of DNA microarrays in the mid-1990s, it is now possible to examine the genes spanning the entire human genome at once and investigate global gene expression patterns. A range of microarrays are available, including oligonucleotide and cDNA arrays, which predominantly utilise a 3' hybridisation method.

1.9.1.1 Affymetrix Human Exon Array 1.0 ST

While standard 3' microarrays measure only gene expression, recently introduced exon arrays can measure the expression of single exons within genes. The Affymetrix Human Exon 1.0 ST array allows the highest resolution of expression analysis, consisting of ~5.5 million probes clustered into 1.4 million probe sets and covering around 80% of exons within the human genome. Approximately 4 probes per exon enable 'exon-level' analysis to distinguish between different isoforms of a gene and detection of alterations in exon usage (alternative splicing events). Additionally, the summarisation of probes (around 40 per gene) into an expression value for all transcripts allows 'gene-level' expression analysis.

Input sequences and annotations to design the probes for exon arrays are selected from a variety of sources including the high quality human genome assembly, annotations inferred from human cDNAs and from a number of publicly available gene prediction sets e.g. Ensembl,

GENSCAN and Vega. This information is consolidated onto the genome and divided into probe selection regions that are contiguous and do not overlap in genomic space (Affymetrix 2006). Probes on the array are divided into three categories, according to their reliability: 1) 'core' probes derived from RefSeq transcripts or full-length mRNAs 2) 'extended' probes, derived from expressed sequence tag evidence and 3) 'full' probes, derived from computationally predicted exons.

To estimate a reliable background signal, the exon array design consists of 40,000 mismatch probes (MM) for comparison with perfect match (PM) probes, binned according to their GC content in order to account for the effects of GC-richness on hybridization strengths. The background signal for a PM probe is estimated by calculating the median intensity of all background probes with the same GC content as the PM probe (Zimmermann & Leser 2010).

The Affymetrix exon array differs from 3' expression arrays in a number of ways to allow enhanced gene expression profiling. A higher coverage of more evenly distributed probes provides a more accurate measure of gene expression. While 3' array probes hybridise to an anti-sense strand (commonly RNA), the Affymetrix exon array consists of a 'Sense Target' (ST) hybridisation design using random primers that attach to a cDNA sense strand. This avoids hybridisation bias resulting from targeting only the 3' end of transcripts, many of which may lack poly-A tails and therefore do not bind to a 3' array (Zimmermann & Leser 2010).

1.9.2 Transcriptomics for Ischaemic Stroke Biomarker Discovery

Several studies have employed large-scale gene expression platforms to determine genes associated with carotid stenosis and plaque instability (summarised in Table 1.2), and have implicated the involvement of inflammation, angiogenesis, deranged extracellular matrix and lipid metabolism, and iron haemostasis in carotid disease progression, as well as RNA splicing and processing factors. Hence tissue-derived biomarkers based upon gene expression data can provide novel insights into the molecular mechanisms underlying carotid disease and plaque rupture.

Comparing carotid plaques from the Biobank of Karolinska Endarterectomies (BiKE) to normal iliac artery tissue, Wang et al. (2012) found an up-regulation of IGFBP-1 and down-regulation of IGF-1, IGF-2 and IGFBP-3 to IGFBP-6 in carotid plaque samples. The expression IGFBP-1, involved in cell migration and proliferation, was co-localised with smooth muscle cells and macrophages and correlated with a number of inflammatory genes. Di Taranto et al. (2012) also demonstrated a key role for inflammation, showing that genes involved in the eicosanoid

synthesis pathway were differentially expressed in carotid plaques as compared to normal artery samples. Investigating gene expression differences between unstable carotid plaques and normal mammary arteries, Martinet et al. (2002) used microarray analysis, immunohistochemistry and cell culture and found that LDL aggregation in unstable carotid plaques caused overexpression of DAP, a mediator of apoptotic cell death, in smooth muscle cells. This demonstrates the role of lipid-induced changes within the carotid plaque, causing instability and rupture. Korol et al. (2011) further implicated role of expression differences in genes involved with extracellular matrix remodelling in the loss of plaque stability.

The studies above have investigated gene expression differences between carotid plaques and normal artery tissue. A number of studies have also explored differences between asymptomatic (stable) and symptomatic (unstable) carotid plaques, and have also found differential gene expression of genes involved in lipid metabolism, inflammation, extra-cellular matrix remodelling and RNA splicing (Vemuganti & Dempsey 2006; Saksi et al. 2011). Taking four symptomatic and four asymptomatic carotid plaque specimens, Dahl et al. (2007) revealed significant differential expression of 137 genes with fold difference of >2, and in a replication cohort of 21 more plaque samples the investigators demonstrate an increase in visfatin RNA and protein levels in symptomatic plaques, a finding that was further replicated at the site of coronary plaque rupture in samples from patients with acute coronary artery disease. More recently, Perisic et al. (2013) have used a novel approach using transcriptomic and tissue microarray profiling to demonstrate an increased expression of PCSK6 (proprotein convertase) in symptomatic carotid plaques that could be associated with inflammatory and extracellular matrix remodelling processes that lead to plaque rupture. In order to minimise inter-patient gene expression differences, studies have also investigated paired stable and unstable carotid plaque regions taken from the same patient. For example, Papaspyridonos et al. (2006) took paired stable and unstable plaque regions from 3 symptomatic individuals, and replicated 27 genes found to be differentially expressed through microarray in a further set of 46 unstable and stable plaque regions. These investigations establish the utility of gene microarray technology for hypothesis-free interrogation of carotid plaque gene expression. Such studies are making progress at uncovering a 'genetic fingerprint' of vulnerable plaques, which can comprise a multi-marker risk-stratifying test for ischaemic stroke.

Alternative splicing mechanisms have previously been implicated in carotid artery disease. Berglund et al. (2012) found that splicing of the *AIF-1* gene results in the expression of two different proteins – AIF-1, whose overexpression leads to proliferation of vascular smooth muscle cells and a pro-inflammatory, unstable carotid plaque profile, and IRT-1, which plays an

opposite role. Smooth muscle cell proliferation has also been demonstrated to be promoted by expression of alternatively spliced *ID3* gene (Forrest et al. 2004).

These studies have used targeted approaches for understanding transcript diversity. In contrast, the Affymetrix Human Exon 1.0 ST Array provides a global view of both gene and exon expression profiles. In the field of stroke, to date this platform has only been utilised in humans to study angiographic vasospasm in patients with aneurysmal subarachnoid haemorrhage (Ko et al. 2011). Investigation of global gene expression and splicing processes using the Affymetrix Human Exon 1.0 ST Array could reveal novel mechanisms underlying carotid plaque progression and biomarkers for predicting risk of ischaemic stroke.

Table 1.2: Summary of microarray transcriptomics investigations of carotid artery disease.

Study	Disease State	Platform Used	Summary of Findings
Di Taranto et al 2012	Carotid Stenosing Plaque vs Non-Stenosing Plaque Region vs Healthy Control	TaqMan low-density array human inflammation panel	Differential expression of 44 genes between diseased and non-diseased tissue, mostly involved in the eicosanoid synthesis pathway.
Wang et al 2012	Carotid Stenosing Plaque vs Healthy Control	Affymetrix HG U133A 2.0 Array	Upregulation of IGFBP-1 in carotid plaques compared with normal arteries, and downregulation of IGF-1, IGF-2 and IGFBP-3 to IGFBP-6. No significant difference of the gene expression level of IGFBP-1 between symptomatic and asymptomatic plaques.
Korol et al 2011	Unstable Carotid Plaque vs Healthy Control	Affymetrix HG U133A v2 Array	Upegulation of collagen I, III and IV, lysyl oxidase and MMP-1, 9 and 12 in unstable plaque
Martinet et al 2002	Carotid Stenosing Plaque vs Healthy Control	ClonTech Human Apoptosis Array	Differential expression of apoptosis-related and housekeeping genes, in particular death-associated protein (DAP) kinase
Perisic et al 2013	Symptomatic vs Asymptomatic Carotid Plaque	Affymetrix HGU133plus 2.0 Array	Differential expression of genes involved in immune response, cell proliferation, adhesion and apoptosis, lipid metabolism, glucose transport and proteases (e.g. PCSK6)
Saksi et al 2011	Symptomatic vs Asymptomatic Carotid Plaque	Affymetrix HG U133A Array	Differential expression of genes involved in inflammatory responses, intracellular lipid and carbohydrate metabolism and transport, lysosomal activity, and reorganization of the extracellular matrix.
Agardh et al 2011	Symptomatic vs Asymptomatic Carotid Plaque	Affymetrix HG U133A Array	Upregulation of FABP4/aP2 in unstable plaques, correlated with differential expression of CD36, CD68, CD163, CD52 and adipophilin
Dahl et al 2007	Symptomatic vs Asymptomatic Carotid Plaque	Affymetrix HG U133A Array	Differential expression of 136 genes, including visfatin
Vemuganti et al 2006	Symptomatic vs Asymptomatic Carotid Plaque	Affymetrix HG U133A Array	Upregulation in symptomatic plaques of genes involved in ionic homeostasis, signal transduction and lipid metabolism, as well as angiogenesis promoters, transcription factors, and RNA splicing and processing factors
Woodside et al 2003	Paired Stenosing vs Non-Stenosising Carotid Plaque	Atlas Human Apoptosis cDNA Expression Arrays	Differential expression of genes involved in signalling pathways that alter cellular apoptotic patterns
Ijas et al 2007	Paired Symptomatic vs Asymptomatic Carotid Plaque	Affymetrix GeneChip U95A v2 Array	Differential expression of genes particularly involved in the homeostasis of iron and heme, especially HO-1 and CD163
Papaspyridonos et al 2006	Paired Symptomatic vs Asymptomatic Carotid Plaque	Affymetrix HG U133A Array	Differential expression of genes known to be associated with the recruitment of cells to sites of injury or inflammation, lipid metabolism and activation of MMPs and cathepsins

1.10 Genomics

While genetics examines specific genes, genomics is the study of all the genetic information in a genome to determine biological markers that predispose an individual to a disease phenotype. The diploid human genome contains ~6 billion nucleotide bases, of which ~0.5% can differ between any two individuals (Lanktree et al. 2009). These differences include single nucleotide polymorphisms (SNPs), whereby a given base (A, T, G or C) differs between two alleles.

With the ease and affordability of genotyping, as well as the cataloguing of common alleles and population variation by the International HapMap Consortium and the 1000 Genomes project, genome-wide association studies (GWAS) have become a popular strategy for investigating the association of SNPs to a disease phenotype. These studies contrast with candidate gene-based association approaches in that no *a priori* knowledge of possible causative mechanisms is required, as with most -omics study designs therefore allowing novel associations can be detected. The GWAS approach proposes a 'common disease, common variant' hypothesis, whereby common diseases are likely influenced by genetic variation that is also common within a general population. Hence compared to linkage analysis and candidate gene-based association analysis, GWAS provides a useful strategy for polygenic traits such as ischaemic stroke, with the implication that a number of common alleles confer small genetic effects (i.e. have low penetrance).

Biomarkers identified through GWAS could prove useful for patient risk stratification or prognostic categorisation following medical therapy. Most importantly though, these genomic markers can provide distinctive mechanistic information about disease pathogenesis and therefore deliver new targets for drug development (Kingsmore et al. 2007). Previous GWAS have successfully identified the role of complement factor H in age-related macular degeneration (Klein et al. 2005), and a number of genes associated with pancreatic islet beta cell function and insulin biosynthesis in type 2 diabetes (Sladek et al. 2007) that can act as therapeutic targets. More recently, novel genetic loci associated with established tau and phosphorylated tau protein biomarkers for Alzheimer's Disease demonstrate the ability of GWAS to detect new genetic variants that could drive the dysregulation of known molecular pathways involved in disease progression (Cruchaga et al. 2013). Despite their small effects, alleles identified through GWAS could have a substantial impact on human health due to their high incidence within a population. Thus identification of SNPs associated with ischaemic stroke could improve risk profiling, although multiple markers of small effect, rather than a single SNP, would likely be required.

1.10.1 Methods for GWAS Analysis

The GWAS strategy is built upon knowledge that the relatively brief history of most human populations has impeded the number of generations required to create recombination events (or mutations) between closely linked markers, thus 80% of the human genome is comprised of ~10 kilobase regions, termed haplotype blocks, that do not show recombination and are in linkage disequilibrium (LD) (International HapMap Consortium 2005).

The fundamental requirement of a genotyping array used for GWAS is the ability to interrogate the genetic loci associated with a trait, so that allele frequency differences between 'disease' and 'control' populations can be measured. Direct locus interrogation involves the genotyping of SNPs directly responsible for the observed phenotype. More commonly, interrogation is indirect: genotyped SNPs are not necessarily causal, but in close proximity to functional mutations, and are therefore associated with a trait due to LD – the non-random association of alleles at two or more genomic loci.

A number of arrays are available for GWAS, which can be compared by their global coverage and platform design. Global coverage is defined as the fraction of common SNPs (minor allele frequency (MAF) ≥0.05) across the genome that are tagged by, or in LD with, the SNPs on the array (typically at a threshold of r^2 ≥0.8). Analysis of data from the HapMap project estimates that >80% of commonly occurring SNPs in a CEU population (individuals of northern and western European descent) can be captured using a subset of 500,000 to 1 million SNPs across the genome (Li et al. 2008). Assay designs vary across platforms, for example Illumina arrays employ a tag SNP approach, whereby SNPs on the platform are chosen based upon LD patterns from the HapMap data set, and have been shown to demonstrate unparalleled and coverage for efficient association discovery (Delano et al.2010). The Illumina Human 1M BeadChip consists of 900,000 tag SNPs from the HapMap dataset as well as 100,000 non-HapMap SNPs, allowing coverage of 93% of common SNPs in the CEU population. The Illumina 610 Quad chip and the Illumina HumanHap550k chip consist of 550,000 SNPs (with the 610 Quad chip containing an extra ~60,000 copy number variants), and provide 87% coverage (Spencer et al. 2009). In contrast to the Illumina design, other arrays, such as the Affymetrix SNP Array 6.0, use an unbiased selection of SNPs from various sources (recombination hotspots, mitochondria etc.) and therefore do not employ a tag SNP strategy.

As with all –omics-based approaches, GWAS faces the challenge of 'multiple testing'. When multiple statistical tests are conducted for a single hypothesis (i.e. over 500,000 SNPs are tested for association with a trait), it is likely that a proportion display association by chance alone, hence stringent adjustment is required to limit the number of type I errors (false positive associations). The Wellcome Trust Case Control Consortium proposed both frequentist and

Bayesian approaches to multiple testing adjustment (Wellcome Trust Case Control Consortium 2007). Current GWAS have settled on a frequentist-based threshold of $p<5 \times 10^{-8}$ for assigning 'genome-wide' significance in populations of European ancestry, based upon a significance level (α) of 0.05, adjusted for an estimated 1 million independent SNPs in the genome (n) by the Bonferroni method, as described in the equation below (Johnson et al. 2010):

$$p \text{ (Study-wide Type I error)} = 1 - (1-\alpha)^n$$

The Bonferonni correction method is often used due to its simplilcity, but it is conservative and inappropriately assumes independence of SNPs on a SNP array. Therefore, others have proposed using weaker significance levels in order to account for underlying SNP LD (Ziegler et al. 2008). Monte-Carlo, permutation methods and False Discovery Rate (FDR) have also been advocated (Bush & Moore 2012).

1.10.2 GWAS for Ischaemic Stroke Biomarker Discovery

Early stroke studies attempted to replicate associations detected by GWAS of other cardiovascular diseases such as atrial fibrillation and coronary artery disease. The first GWAS of ischaemic stroke analysed 278 cases and 275 neurologically normal controls, finding no genetic loci reaching genome-wide significance after adjustment for conventional risk factors (Matarin et al. 2007). Nonetheless, replication of 3 of the 25 top associated SNPs in a Han Chinese population (Ding et al. 2010) provided encouragement for the utility of the GWAS strategy for examining genetic variants in ischaemic stroke. Genotyping of 1,661 Icelandic ischaemic stroke patients and 10,815 control subjects in another meta-analysis found that SNP rs2200733, in the gene *PITX2*, was uniquely associated with cardioembolic stroke subtype. This gene has also previously been found to be associated with atrial fibrillation, thus providing evidence for the function of distinct genetic variants in the risk of different stroke subtypes (Gretarsdottir et al. 2008).

Initial GWAS investigation of ischaemic stroke was based on small cohort sizes. Due to the large number of SNPs studied in GWAS, and the small associations of each SNP with the phenotype studied (typically OR 1.1-1.3), large sample sizes are required for statistically meaningful associations to be derived. GWAS cohorts examining the same trait can be grouped in a meta-analysis to increase the power of detecting significant and reproducible results (Thompson et al. 2011). Genotype imputation is usually performed before meta-analysis to facilitate the combination of studies. This involves using reference data from the HapMap project or 1000 Genomes project to infer missing genotypes and untyped SNPs based upon genomic LD structure (Marchini & Howie 2010). Recent GWAS have developed a meta-analysis approach, either for primary analysis or in the replication phase of the study (International Stroke

Genetics Consortium [ISGC] & Wellcome Trust Case-Control Consortium [WTCCC2] 2010; Meschia et al. 2011; ISGC & WTCCC2 2012; Holliday et al. 2012a; Traylor et al. 2012).

A GWAS of four large cohorts (Caucasian and black) from the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium found an association of a locus on *12p13* (*NINJ2*) with ischaemic stroke (Ikram et al. 2009), but this failed to be replicated in a larger case-control study of 9000 ischaemic stroke individuals (ISGC & WTCCC2 2010). A subsequent GWAS meta-analysis of 3 cohorts, comprising 1464 ischemic stroke cases and 1932 controls also failed to identify any SNPs reaching genome-wide significance of $p=5x10^{-8}$. However, all SNPs reaching suggestive significance of $p=1x10^{-5}$ were incorporated into a multivariate risk profile model and revealed that genetic loading of alleles was related to an earlier age of ischaemic stroke onset (Meschia et al. 2011).

Although these studies have been largely disappointing for identifying novel genetic associations in ischaemic stroke, recent GWAS studies have proved more fruitful. The 2012 WTCCC2 GWAS investigated ischaemic stroke and its 3 major subtypes (cardioembolic, lacunar and large artery) in 3548 cases and 5972 controls of European ancestry, and replicated suggestive findings in a further 5859 cases and 6281 controls. Previously reported associations of *PITX2* and *ZFHX3* for cardioembolic stroke and *9p21* locus for large artery stroke were replicated, and a novel association of *HDAC9* with large artery stroke was also discovered. This association was additionally replicated in 735 cases with large artery stroke. A recent experimental study has shown the HDAC9, a protein that regulates chromatin structure and gene transcription by histone deacetylation, is expressed in endothelium and smooth muscles of carotid arteries, supporting its role in carotid disease and ischaemic stroke (Markus et al. 2013).

A GWAS of subclinical carotid atherosclerosis (common cIMT, internal cIMT and carotid plaque), the only one of its kind, has recently been performed on 31,211 individuals from 9 study cohorts from the CHARGE Consortium, with replication of initial findings in an additional 11,273 individuals from seven more studies (Bis et al. 2011). Three genomic loci were identified for association with common cIMT (*ZHX2, APOC-1, PINX1*) and two regions (*PIK3CG, EDNRA*) were associated with the presence of carotid plaque, indicating that genes related to cellular signalling, lipid metabolism and blood pressure regulation are involved in the risk of carotid disease.

The largest GWAS for ischaemic stroke published to date has combined 15 studies comprising 12,389 ischaemic stroke cases and 62,004 controls, and conducted a follow-up study of 16 cross-sectional and two population-based studies with 13,347 cases and 29,083 controls (Traylor et al. 2012). The four previous reported loci, *PITX2*, *ZFHX3*, *9p21*and *HDAC9*, were

replicated, with *PITX2* and *ZFHX3* being associated with cardioembolic stroke and 9p21and *HDAC9* being associated with large artery stroke. A further 12 novel loci were identified but could not be replicated in the follow-up study. Another GWAS of ischaemic strokes in an Australian population of European ancestry confirmed that genetic factors explain a substantial proportion of the ischaemic stroke phenotype, with the highest genetic load observed for large artery stroke as compared to other stroke subtypes. A novel association of the *6p21.1* locus with large artery stroke was also observed (Holliday et al. 2012).

Table 1.3 summarises GWAS investigations of ischaemic stroke and carotid disease in European cohorts, with details of associations reaching high significance. The studies establish the value of GWAS for determining novel genetic mechanisms involved in ischaemic stroke risk. Importantly, they indicate that stroke subtypes have different underlying pathophysiological mechanisms and could therefore require distinct treatment profiles. Given that genetic variants commonly have association ORs of ~1.2, it has been calculated that up to 300 variants could explain the combined risk of complex polygenic disorders such as ischaemic stroke (Kraft & Hunter 2009; Hirschhorn 2009). Therefore further studies are required in order to investigate the novel genetic variants and biological pathways conferring risk of ischaemic stroke subtypes.

Table 1.3: Summary of GWAS of ischaemic stroke and carotid artery stenosis. SNPs reaching genome wide significance of $<10^{-6}$ are listed, along with their respective associated genes. **ISGS,** International Stroke Genetics Consortium; **NS,** none significant; **WTCCC2,** Wellcome Trust Case–Control Consortium 2.

Study	Trait	SNP	Region	Reported Gene(s)
Traylor et al 2012	Ischaemic Stroke	rs2107595	7p21.1	HDAC9
		rs6843082	4q25	PITX2
		rs879324	16q22.3	ZFHX3
		rs13407662	2p16.2	Intergenic
		rs16851055	3q23	SPSB4
		rs2238151	12q24.12	ALDH2
Holliday et al 2012	Ischaemic Stroke	rs556621	6p21.1	CDC5L, SUPT3H
ISGC & WTCCC2 2012	Ischaemic Stroke	rs11984041	7p21.1	HDAC9
Meschia et al 2011	Ischaemic Stroke	NS	NS	NS
ISGC & WTCCC2 2010	Ischaemic Stroke	NS	NS	NS
Ikram et al 2009	Ischaemic Stroke	rs12425791	12p13.33	NINJ2
Gretarsdottir et al 2008	Ischaemic Stroke	rs2200733	4q25	PITX2 -RPL36AP23
Matarin et al 2007	Ischaemic Stroke	rs7506045	18p11.21	IMPA2
		rs9536591	13q14.3	Intergenic
		rs10486776	7p21.2	Intergenic
		rs783396	6q21	AIM1
Bis et al 2011	Carotid intima media hickening and carotid plaque	rs17398575	7q22.3	PIK3CG
		rs1878406	4q31.22	EDNRA
		rs445925	19q13.32	APOC-1
		rs11781551	8q24.13	ZHX2
		rs6601530	8p23.1	PINX1
		rs4712972	6p22.2	SLC17A4
		rs6511720	19p13.2	LDLR
		rs17045031	3p14.1	LRIG1

1.10.3 Functional Enrichment Analysis

Adjusting for multiple testing in GWAS by using a stringent significance threshold reduces the number of type I errors at the expense of type II errors (false negatives). Thus many true associations may be concealed within the $p=10^{-7}$ to $p=0.05$ region, and the huge amount of genotypic information obtained from GWAS may not be optimally used. Furthermore, it is often unclear how associated loci exert an effect at both a gene and pathway level, which is essential for understanding disease pathogenesis.

Functional enrichment analysis provides a method for multiple SNP-disease testing that may provide better grasp of disease mechanisms, using prior biological knowledge of gene function and interaction to leverage deeper mining of genotypic data (Wang et al. 2007; Elbers et al. 2009). This approach is likely to be more powerful than single SNP analysis when applied to small studies, as well as being more robust to the effects of genetic heterogeneity between individuals, ethnic backgrounds, genotyping platforms and study design (Holmans et al. 2009). Functional enrichment and pathway analysis can also add power to GWAS meta-analyses based upon only *p*-values across studies (Cantor et al. 2010).

Three main strategies exist for functional enrichment analysis: 1) the 'hypothesis-driven' approach 2) the 'hypothesis-generating' approach and 3) the 'hypothesis-free' approach. Using the 'hypothesis-driven' approach, specific pathways are examined based upon presumed biological processes involved in the disease being studied. The 'hypothesis-generating' approach involves taking a subset of significant SNPs from a GWAS and examining the enrichment of these SNPs in functional categories, classified, for example, by Gene Ontology (GO). The 'hypothesis-free' approach takes SNPs that may only have reached nominal significance (e.g. $p<10^{-4}$), or all SNPs tested, and investigates pathway associations with no *a priori* speculation.

Over-representation analysis can be used in 'hypothesis-generating' and 'hypothesis-free' approaches to identify biologically-related genes and group them into functional pathways. These methods are commonly based upon a hypergeometric distribution, with the assumption that if a biological process is abnormal in a given study, the co-functioning genes are more likely to be selected as a relevant group by high-throughput screening than by chance alone. A number of knowledge databases have been developed to aid this process, for example Ingenuity Pathway Analysis (Ingenuity®Systems, www.ingenuity.com), gene ontology (GO) (Ashburner et al. 2000) and Kyoto Encyclopedia of Genes and Genomes **(**KEGG) (Kanehisa & Goto 2000) and protein protein interaction databases such as STRING (Jensen et al. 2009) and InWeb (Lage et al. 2007). The over-representation method has been widely used in gene expression studies to investigate the molecular pathways underlying atherosclerosis (Sluimer et al. 2007; Diez et al.

2010), and it has recently shown to be a promising approach to GWAS analysis as well (Wang et al. 2007; Morris et al. 2012).

One of the biggest limitations of over-representation analysis is that it requires an arbitrary level of significance to be set (e.g. by FDR or Bonferroni correction) on the data obtained from high-throughput screening. It also only uses gene numbers while ignoring the strength of association between the gene and phenotype being studied. In contrast, functional class scoring methods use all the genes obtained from a high-throughput experiment and compute enrichment scores for each pathway by considering the genes' association statistics (e.g. p-value, fold difference). As above, these pathways are generated based upon functional knowledge databases. Gene Set Enrichment Analysis (GSEA) is a popular functional class scoring approach that calculates pathway enrichment scores by increasing a running-sum statistic when a gene is in a gene set and decreasing it when it is not, with the size of the increase or decrease determined by the level of association between the gene and phenotype (Aravind Subramanian et al. 2005). Gene set enrichment approaches have previously been employed in gene expression analyses of carotid atherosclerosis (Sluimer et al. 2007; Cagnin et al. 2009) and more recently to interrogate data from GWAS of complex diseases such as type 2 diabetes (Voight et al. 2010) and coronary heart disease (de las Fuentes et al. 2012).

1.10.4 Expression Quantitative Trait Locus and Quantitative Trait Transcript Analysis

GWAS aims to identify loci in the genome associated with a phenotypic trait, with the assumption that a functional genetic variant underlies phenotypic variation across a population. Continuous traits, such as degree of carotid stenosis, are phenotypes that vary in degree and can be attributed to polygenic effects. Consequently, a quantitative trait locus (QTL) is a genetic variant (e.g. SNP) that underlies a continuous trait. Gene expression data can be used as an intermediate between genotype and a continuous phenotype by performing expression quantitative trait locus (eQTL) analysis to provide a better understanding of the genetic regulation underlying gene expression (Veyrieras et al. 2008; Nicolae et al. 2010). Two types of eQTLs exist: *cis*-eQTLs describe loci in 'local' proximity to the transcript they are mapped to, with proximities ranging from 100kb (Cookson et al. 2009) to 10MB (Hubner et al. 2005); in contrast *trans*-eQTLs describe loci that are in a distant region, or different chromosome, to the transcript they are mapped to. Some transcripts can be regulated by both *cis*- and *trans*- eQTLs (Petretto et al. 2010).

Similarly, the relationship between gene expression and phenotypic variation can be examined by quantitative trait transcript (QTT) analysis (Passador-Gurgel et al. 2007). In this method, genome-wide gene expression data, or a subset of *cis*-eQTL genes, are used to investigate the

correlation between gene expression level and degree of phenotypic variation, with significant correlations inferring a functional relationship between the gene and the phenotype (Morrissey et al. 2011).

Although GWAS of ischaemic stroke and its intermediate phenotypes (e.g. carotid disease in LAA stroke) have advocated the examination of distinct stroke subtypes and identified new genetic associations of disease, at present the findings do not hold any direct implications for ischaemic stroke risk stratification and the pathways underlying ischaemic stroke subtypes are unclear. Functional enrichment analysis could be used to better understand the biological context of genetic variants identified through GWAS and deduce molecular networks, rather than single genes, causing ischaemic stroke. Further eQTL and QTT analysis can establish the role of SNPs and genes identified through enrichment analysis at the functional level, hence helping to establish a causal link between the SNP, expression of its related gene, and its importance within a pathogenic molecular pathway. This integrative genomics approach can identify novel genetic biomarkers and mechanisms involved in carotid plaque pathogenesis and risk of ischaemic stroke. To date, this approach has not been attempted in the field of ischaemic stroke research.

1.11 Hypothesis and Aims

The rising global burden of ischaemic stroke necessitates effective disease management. However, current diagnostic, prognostic and risk stratification tools have a number of limitations. Molecular biomarkers can be effective indicators of disease mechanisms and progression, and proteomic, transcriptomic and genomic approaches represent unbiased, system-wide strategies for biomarker discovery in polygenic, complex disorders such as ischaemic stroke.

The hypothesis underlying this thesis is that -omics approaches can be employed to identify novel candidate biomarkers that could aid diagnosis and disease monitoring for ischaemic stroke. The subsequent chapters aim to evaluate the current status of biomarkers for ischaemic stroke and develop workflows for novel biomarker discovery, as follows:

Chapter Two: To undertake a comprehensive systematic review and meta-analysis of literature to identify current blood-based biomarkers associated with ischaemic stroke, and assess the implications of these biomarkers in clinical practice.

Chapter Three: To use proteomic profiling to identify novel blood-based protein biomarkers for the diagnosis of acute ischaemic stroke.

Chapter Four: To use transcriptomic and genomic profiling to characterise the molecular mechanisms underlying carotid stenosis and plaque instability, and identify tissue-based biomarkers that could aid risk stratification of ischaemic stroke.

CHAPTER TWO:

BIOMARKERS FOR ISCHAEMIC STROKE – SYSTEMATIC LITERATURE REVIEW AND META-ANALYSIS

2.1 Introduction

An accurate diagnosis of ischaemic stroke and its likely prognosis holds many important implications, for example, whether to admit a patient to a hyperacute stroke unit or medical ward, administer medications to reduce the risk of future vascular events, or to consider appropriate avenues for rehabilitation. Prompt diagnosis is required for the administration of intra-venous recombinant tissue plasminogen activator (rt-PA) thrombolytic therapy within the first few hours of ischaemic stroke onset. Furthermore, up to 8% of high risk patients suffer a recurrent stroke within the first 2 days (Johnston et al. 2007), hence rapid initiation of secondary preventative treatments is key. However, the diagnosis of ischaemic stroke and prediction of outcome is difficult to make. The probability of a stroke diagnosis and its likely prognosis is usually determined by an assessment of clinical symptoms, brain imaging and past history, but the accuracy of this evaluation is largely dependent upon the expertise of an experienced clinician.

Point of care testing for ischaemic stroke holds many advantages, including speed and ease of use even by those with little training. As well as improving clinical efficiency, it could reduce costs by shortening hospital stay and decreasing the requirement for other diagnostic tools. Circulating molecules in the bloodstream can provide a means to monitor physiological processes following acute ischaemic stroke, and these blood-based biomarkers are routinely used in the management of many disorders. For example, raised cardiac troponin levels can differentiate myocardial infarction from angina or other causes of chest pain (Brott et al. 1989), and the incorporation of brain natriuretic peptide into the GRACIE score increases prediction of outcome after myocardial infarction (Lorgis et al. 2009). Blood-based biomarker tests could increase diagnostic certainty by helping to distinguish cerebral ischaemia from haemorrhagic stroke and common stroke mimics, while also identifying patients at high risk of developing future vascular complications. Prognostic biomarkers will be valuable in stratifying a patient population by potential outcomes, as well identifying patients most likely to benefit from early therapeutic intervention.

While many protein and metabolic blood biomarkers have been implicated in the diagnosis or prognosis of ischaemic stroke, the successful development of blood-based tests through biomarker identification has proved to be a challenge. Studies investigating potential biomarkers have yielded inconsistent results and hence the significance of many proposed biomarkers remains unclear.

Aims of the Study

In an attempt to better understand the relevance of currently proposed biomarkers for stroke before undertaking research for novel biomarker discovery, this study aimed to:

- Perform a comprehensive systematic review of published literature reporting blood-based biomarkers for the diagnosis and prognosis of ischaemic stroke.
- Assess the statistical association, if any, between reported biomarkers and the diagnosis or prognosis of ischaemic stroke.

2.2 Methods

2.2.1 Literature-Based Meta-Analysis of Diagnostic and Prognostic Biomarkers for Stroke

2.2.1.1 Data Sources and Extraction

Electronic databases MEDLINE, EMBASE and Google Scholar were searched up to January 1st 2013 for all studies detailing the use of blood biomarkers for stroke in humans. Search retrieval was maximised by using 14 terms for ischaemic stroke (ischaemic, ischemic, cerebral, intracranial, parenchymal, intraventricular, infratentorial, supratentorial, cerebrovascular, infarction, stroke, thrombus, occlusion and embolus), 4 for thrombolysis (thrombolysis, thrombolytic therapy, plasminogen activator and alteplase), 3 for generic biomarkers (biological, biochemical marker and biomarker), and 780 specific biomarker terms with 'and/or' used as a Boolean operator (Appendix 1). References from all relevant papers were hand searched for additional studies, and the MEDLINE 'Related Articles' option was used to identify further articles. The search was restricted to the English language, but both unpublished and published articles were included, providing the full text was available. While the introduction of unpublished literature has raised concerns, it is widely acknowledged that if this 'grey' literature undergoes strict quality control (as performed here, with three independent reviewers), it can improve the reliability of a meta-analysis by mitigating publication bias (Bossuyt & Leeflang 2008).

Inclusion criteria were set following recommendations from the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy (Bossuyt & Leeflang 2008). For the evaluation of diagnostic biomarkers, studies were required to have reported candidates that discriminated between ischaemic stroke and non-stroke subjects (healthy, stroke mimics or other neurological diseases), or between ischaemic and haemorrhagic stroke, with blood being drawn within 24 hours of stroke onset. Where more than one cohort was examined, data for each population were extracted separately. Studies reporting the prognostic value of biomarkers were eligible for inclusion if blood had been drawn within 24 hours of stroke onset, and outcome was measured using death, disability or handicap scales. No minimal sample size was set for study inclusion, but only papers reporting continuous measures of candidate biomarker activity or concentration (mean/median and standard deviation/range) were chosen for meta-analysis. Papers reporting only the risk of subsequent stroke or risk of stroke in a non-stroke population were not included. Papers examining subarachnoid haemorrhage were excluded.

2.2.1.2 Statistical Analysis

For each candidate biomarker for which data were available for more than two studies, a meta-analysis was performed. Where values were reported as medians and ranges, the mean and standard deviation were estimated using a variety of models dependent upon sample size, as described by Hozo et al. (2005). Data were analysed using the Review Manager v5.0 software. Standardised Mean Differences (SMDs) were calculated for each candidate biomarker using the DerSimonian-Laird random-effects model (DerSimonian & Laird 1986), as well as 95% confidence intervals (CI) to measure the strength of association. SMD was calculated as follows:

$$\text{SMD} = \frac{\text{Difference in mean outcome between groups}}{\text{Standard deviation of outcome among participants}}$$

Heterogeneity was tested with the chi-square (χ^2) test and assessed using I^2 ($p<0.05$) to describe the percentage of variability in point estimates that was due to heterogeneity rather than sampling error (Deeks & Altman 2001). The possibility of publication bias was not formally assessed as it has been noted that test power is usually too low to distinguish chance from real asymmetry when considering fewer than ten studies (Sterne et al. 2011).

2.2.2 Literature-Based Meta-Analysis of Admission Glucose Level as a Prognostic Biomarker for Post-Thrombolytic SICH

2.2.2.1 Data Sources and Extraction

Electronic databases MEDLINE, EMBASE and Google Scholar were searched up to January 1st 2013 for all studies examining blood glucose levels in thrombolysed acute ischaemic stroke patients who developed SICH post-treatment. Papers were identified using the following key words: 'symptomatic', 'intracerebral', 'haemorrhage', 'stroke' and 'thrombolytic'. Studies were eligible for inclusion if blood glucose levels were measured within 24 hours of stroke onset, and primary outcome measures included SICH (of any classification) and an assessment of functional outcome (mRS). The search was limited to those patients treated only with intravenous rt-PA alteplase. Reference lists of all relevant papers were hand-searched for additional studies and the MEDLINE 'Related Articles' option was used to identify further articles. The search was not restricted by language, and both unpublished and published articles were included, providing the full text was available. No minimum sample size was set for study inclusion.

2.2.2.2 Statistical Methods

The association of glucose levels with the presence of post-thrombolysis SICH was analysed against the absence of SICH. In instances when glucose level was reported in mg/dL, the units were converted to mmol/L for consistency.

Studies reporting the odds ratio (OR) associated with developing SICH at specific incremental changes of glucose were pooled for analysis. For each of these studies, ORs and CIs were standardised to 1mmol/L increase of glucose, using the following formula (Allison 1999):

OR per 1mmol/L increase in glucose = OR per θ mmol/L ^ (1/ θ) **[A]**

For studies reporting continuous data (means and standard deviations for their SICH and non-SICH populations) SMDs were calculated. This firstly involved re-expressing the difference in mean glucose between SICH and non-SICH groups as a multiple of the pooled standard deviation (S_{pooled}), which is a weighted average of the standard deviations (SDs) from the SICH (t) and non-SICH (c) group in each data set:

$$s_{pooled} = \sqrt{\frac{(n_t - 1)s_t^{\,2} + (n_c - 1)s_c^{\,2}}{n_t + n_c}}$$

where n = number of subjects

SMD was then calculated as:

SMD = mean difference/S_{pooled}

The SMD was converted into a logOR by multiplying by 1.81 ($\frac{\pi}{\sqrt{3}}$) (Chinn 2000) and then converted to an OR. The calculation of OR for each continuous data set and thus be summarised as follows:

OR = exp [1.81*mean diff/S_{pooled}]

Multiplying the SMD by standard deviation (in this case S_{pooled}) gives the amount of glucose which changes outcome from non-SICH to SICH (Higgins & Green 2011). Having calculated the OR associated with SICH, and already knowing the level of glucose increase responsible for this outcome, an OR per 1mmol/L increase in glucose was calculated, again using formula [A] above. This subsequently allowed studies reporting ORs and studies reporting continuous data to be pooled together in one meta-analysis. Funnel plot and two-tailed Egger regression test were used to assess publication bias using Comprehensive Meta-Analysis v2.0 software (Egger et al. 1997).

2.2.3 Individual Patient Data Meta-Analysis of Admission Glucose Level as a Prognostic Biomarker for Post-Thrombolytic SICH

2.2.3.1 Study Inclusion and Data Acquisition

Studies were identified through literature search, as detailed in Section 2.2.2 above. Inclusion was restricted to studies with a sample size of more than 50 patients to ensure that the logistical effort of obtaining and organising the data was commensurate with the contribution of the data to the analysis (Riley et al. 2010). In an attempt to limit heterogeneity due to study design, randomised trials were excluded and only observational studies were considered.

Authors and principal investigators of eligible studies were invited to provide individual patient data (IPD). Primary data of interest included patient admission glucose level and presence/absence of SICH following thrombolysis, with a clear explanation of how SICH was defined. When possible, the following patient information was also obtained: age, sex, presence of diabetes and hypertension, onset to treatment time (OTT), baseline NIHSS score and functional outcome (mRS).

2.2.3.2 Statistical Analysis

The primary analysis used the IPD to estimate the absolute risk of SICH in relation to admission glucose level. The largest study cohort (SITS-ISTR, >30,000 patients, (Mazya et al. 2012) was firstly used to examine the relationship between glucose levels and SICH and enable key modelling decisions to be made. Preliminary analysis was performed on continuous variables to assess the normality of their distribution. All available baseline variables (age, sex, onset to time of treatment, admission NIHSS, hypertension, diabetes, smoking, atrial fibrillation, hyperlipidaemmia, previous stroke) were analysed and differences between continuous and categorical variables in the SICH and non-SICH groups was assessed using the Mann-Whitney *U* test and the Pearson χ^2 method respectively. Variables showing an association with all three definitions of SICH (NINDS, ECASS and SITS-MOST) at $p<0.05$ in univariate analysis and available across all datasets were included as potential predictors into multivariate logistic regression models.

It is essential to maintain the clustering of patients within studies when performing individual patient data meta-analyses (Abo-Zaid et al. 2013), therefore a two-step analysis approach was taken. Individual datasets were first subjected to logistic regression analysis to calculate a regression coefficient (log odds ratio) and 95% confidence interval (95% CI). Weighted average log odds ratios were then calculated using a weight of $1/SE^2$ for each study, and these were combined in a meta-analysis to produce a pooled odds ratio (OR) and 95% CI. To test

robustness, both random effects and fixed effects models were used. Meta-analyses were performed separately for each classification of SICH (NINDS, ECASS and SITS-MOST). Heterogeneity was examined using the I^2 statistic to describe the percentage of variability in point estimates due to heterogeneity rather than sampling variability (Deeks & Altman 2001). Publication bias was not formally assessed as there were fewer than ten studies in the meta-analysis (Sterne et al. 2011).

The pooled ORs calculated from the meta-analyses related to the risk of SICH associated with each 1mmol/L increase in admission glucose level. The absolute risk of SICH associated with each increment of 1mmol/L in admission glucose level was calculated as:

$$A \times (B^{(C)})$$

where 'A' is the known % risk of SICH within the pooled population, 'B' is the estimated OR per unit increase of $\log_{10}$glucose and 'C' represents the difference, in $\log_{10}$ units, between the $\log_{10}$ target value and anchor (mean) glucose value of the pooled population. For rare events, with risks less than 10%, the relative risk (RR) and OR are very closely similar, therefore the step required to converted OR to RR could be omitted (Professor Doug Altman, personal communication). All analyses were performed using PASW Statistics v18 and Stata v11.

2.3 Results

2.3.1 Literature-Based Meta-Analysis of Diagnostic and Prognostic Biomarkers for Stroke

The primary search identified 3974 papers reporting candidate diagnostic and prognostic biomarkers for ischaemic stroke and candidate prognostic biomarker for thrombolysis. Of these, 703 studies met criteria for inclusion into the qualitative analysis. Following screening of full text articles, 65 studies for diagnostic biomarkers for ischaemic stroke, 40 studies for prognostic biomarkers for ischaemic stroke and 39 studies for biomarkers of prognosis following thrombolysis were incorporated into meta-analyses (Figure 2.1).

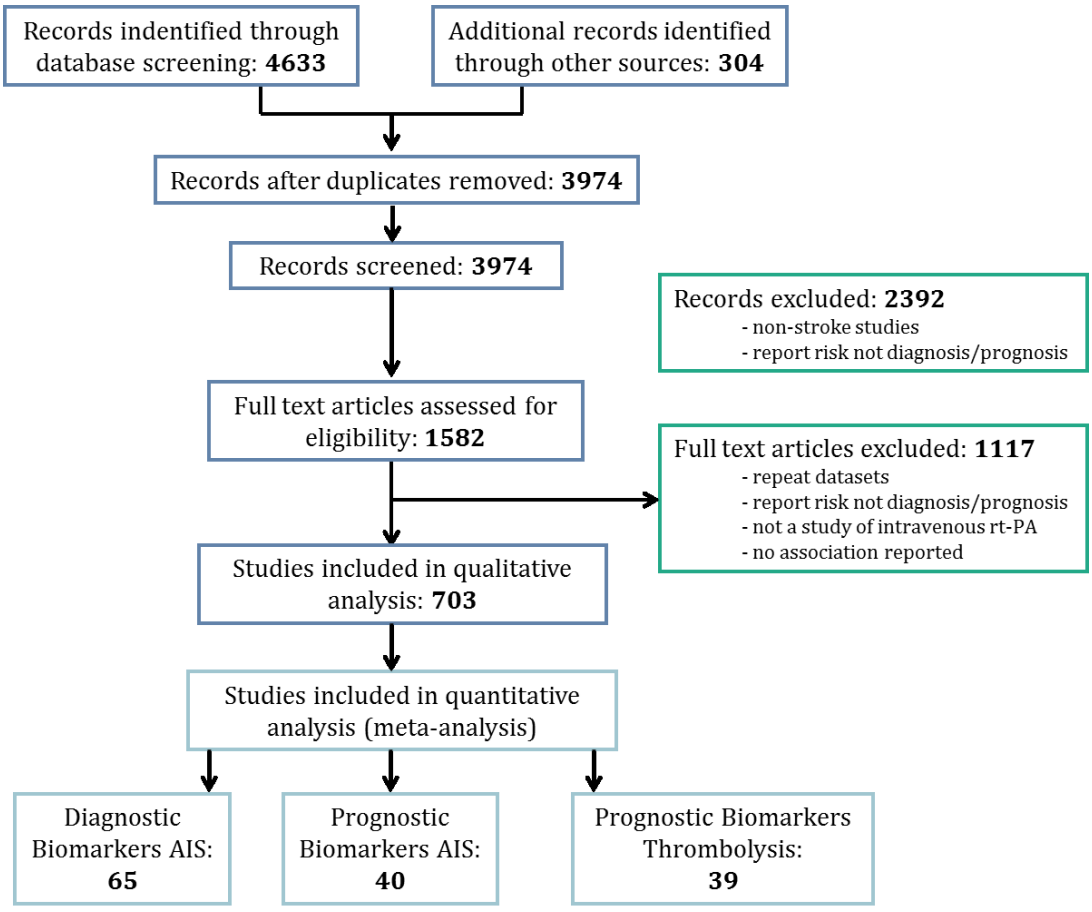


Figure 2.1 PRISMA flow diagram for systematic review and meta-analysis of diagnostic and prognostic biomarkers of ischaemic stroke. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram illustrates the search strategy and number of studies included in the systematic review (qualitative analysis) and meta-analysis (quantitative analysis). **AIS**, acute ischaemic stroke; **rt-PA**, recombinant tissue plasminogen activator.

A total of 137 different candidate biomarkers were identified. A review of the papers reporting diagnostic and prognostic biomarkers for ischaemic stroke revealed that most candidates were only investigated in a single study and had not been validated in large patient populations.

Candidate biomarkers implicated for ischaemic stroke could be classified into functional pathways including lipid metabolism (cholesterol, triglyceride, apolipoproteins), coagulation (fibrinogen, D-dimer, PAI-1) and inflammation (CRP, TNF-α, interleukins), metabolites (creatinine, glutamate, glucose), brain derived molecules (B-type neurotrophic growth factor, S100B, neurone specific enolase, GFAP) and proteins indicating endothelial processes (MMP-9, ICAM-1, P-Selectin). A summary of candidate biomarkers with significant association, as identified through meta-analysis, is illustrated in Figure 2.2.

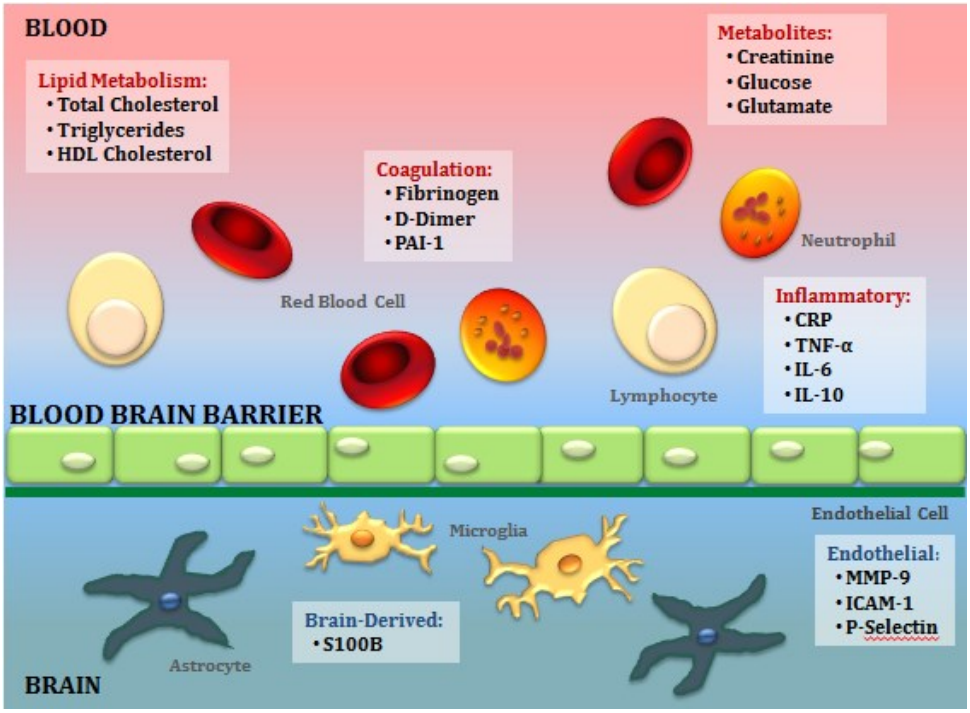


Figure 2.2 Candidate diagnostic and prognostic biomarkers of significant association with ischaemic stroke upon meta-analysis. Candidates were found to be either blood or tissue derived, and could be broadly separated into six main categories: lipid metabolism (total cholesterol, triglyceride, HDL cholesterol), coagulation (fibrinogen, D-dimer, PAI-1), inflammation (CRP, TNF-α, IL-6, IL-10), metabolites (creatinine, glucose, glutamate), brain-derived proteins involved in cell cycle progression (S100B) and proteins involved in endothelial cell adhesion and extracellular matrix integrity (MMP-9, ICAM-1, P-selectin). Proteins involved in cell adhesion and extracellular matrix integrity could derive from endothelial cells of the blood brain barrier, or the endothelium of unstable plaque regions causing the ischaemic stroke. **CRP**, C-reactive protein; **HDL**, high-density lipoprotein; **ICAM-1**, intercellular adhesion molecule-1; **IL**, interleukin; **MMP-9,** matrix metalloproteinase-9, **PAI-1,** plasminogen activator inhibitor-1; **TNF-α**, tumour necrosis factor-α.

2.3.1.1 Diagnostic Biomarkers of Ischaemic Stroke

Meta-analysis effect size was calculated as pooled standardised mean difference (SMD) between mean biomarker levels in study populations. Table 2.1 describes the SMDs associated with each candidate diagnostic biomarker for which meta-analysis could be performed.

Ten candidate biomarkers were found to be of statistical significance for differentiating between acute ischaemic stroke (AIS) and healthy controls after meta-analysis: triglyceride, HDL cholesterol, fibrinogen, CRP, IL-6, IL-10, TNF-α, glucose, ICAM-1 and P-Selectin. In particular proteins involved in lipid metabolism were associated with AIS at high significance upon meta-analysis, with triglycerides being higher and HDL cholesterol levels being lower in AIS patients as compared to healthy controls (triglyceride: SMD=0.80, 95% CI 0.41, 1.18, p=0.0001; HDL cholesterol: SMD=-0.65, 95% CI -0.89, -0.40, p=0.0001). However, the high level of heterogeneity ($p<0.0001$) between studies examining these candidate biomarkers lends uncertainty to the quality of the results. Inflammatory acute phase proteins TNF-α (SMD=0.39, 95% CI 0.18, 0.60, p=0.0002), CRP (SMD=0.78, 95% CI 0.11, 1.44, p=0.02) and fibrinogen (SMD=1.08, 95% CI 0.35, 1.82, p=0.004) were also found to be elevated in acute ischaemic stroke patients as compared to healthy controls. A forest plot of significant associations is provided in Figure 2.3.

Meta-analysis of two studies (Rainer et al. 2007; Glickman et al. 2011) consisting of 142 subjects, found that higher levels of S100B were present in AIS patients as compared to stroke mimics (SMD=0.60, 95% CI 0.26, 0.94, p=0.0005; Figure 2.4). Furthermore, meta-analysis of another two studies (Kavalci et al. 2011; Montaner et al. 2012) consisting of 266 patients found that lower MMP-9 levels could distinguish AIS from ICH (SMD=-0.93, 95% CI -0.30, -1.55, p=0.004; Figure 2.5).

Table 2.1: Candidate diagnostic biomarkers of ischaemic stroke identified through literature review. Results from RevMan meta-analysis for candidate biomarkers reported in two studies or more are detailed, and biomarkers of significance are highlighted in blue. **BNP**, brain natriuretic peptide; **CI**, confidence interval; **CRP**, c-reactive protein; **GFAP**, glial fibrillary acidic protein; **HDL**, high density lipoprotein; **H-FABP**, heart-type fatty acid binding protein; **ICAM**, intracellular adhesion molecule; **IL**, interleukin; **IMA**, ischaemia modified albumin; **LDL**, low density lipoprotein; **MMP**, matrix metalloproteinase; **NT-BNP**, N-terminal brain natriuretic peptide; **PAI,** plasminogen activator inhibitor; **SMD**, standardised mean differences; **TNF**, tumour necrosis factor. References in the table are numbered and listed separately at the end of the References section of the thesis.

	Biomarker	Number of Studies (References)	Number of Subjects	SMD	95% CI	p-value	Inter-study Heterogeneity (I^2; p-value)
AIS vs Healthy Control							
	ICAM-1	3(1–3)	420	1.72	0.12, 3.32	0.04	I^2 = 97%; p- value < 0.00001
	Glucose	2(4,5)	1395	0.71	0.35, 1.07	0.0001	I^2 = 84%; p- value = 0.01
	Total Cholesterol	9 (4–12)	2881	0.39	-0.03, 0.81	0.07	I^2 = 94%; p- value < 0.00001
	Triglyceride	6 (4–8,10)	1713	0.80	-0.41, 1.18	0.0001	I^2 = 88%; p- value < 0.00001
	HDL Cholesterol	6 (4–6,9–11)	2628	-0.65	-0.89, -0.40	0.00001	I^2 = 79%; p- value = 0.0003
	LDL Cholesterol	5 (4,6,7,10,11)	627	0.93	-0.05, 1.91	0.06	I^2 = 96%; p- value < 0.00001
	Creatinine	4 (4,8,11,13)	558	0.21	-0.09, 0.52	0.18	I^2 = 59%; p- value = 0.06
	TNF-a	2 (1,4)	396	0.39	0.18, 0.60	0.0002	I^2 = 0%; p- value = 0.66
	CRP	9 (4,5,10,14–19)	3765	0.78	0.11, 1.44	0.02	I^2 = 98%; p- value < 0.00001
	IL-6	4 (1,9,18,20)	1134	1.21	0.04, 2.38	0.04	I^2 = 96%; p- value < 0.00001
	Neopterin	2 (21,22)	168	0.92	-1.90, 3.75	0.52	I^2 = 98%; p- value < 0.00001
	IMA	2 (23,24)	132	4.10	-3.48, 11.68	0.29	I^2 = 99%; p- value < 0.00001
	Fibrinogen	4 (6,14–16)	1089	1.08	0.35, 1.82	0.004	I^2 = 97%; p- value < 0.00001
	P-Selectin	2 (7,25)	179	3.16	-1.04, 7.35	0.02	I^2 = 99%; p- value < 0.00001
	IL-10	2 (18,21)	64	-1.07	-1.67, -0.48	0.004	I^2 = 17%; p- value = 0.27
	H-FABP	2 (24,26)	74	0.65	-1.62, 2.91	0.58	I^2 = 94%; p- value < 0.0001
	CD40L	2 (8,27)	131	-3.98	-11.87, 3.91	0.32	I^2 = 99%; p- value < 0.00001
	Nitric Oxide	2 (5,28)	1186	-3.51	-14.75, 7.73	0.54	I^2 = 99%; p- value < 0.00001
	Homocysteine	3 (10,29,30)	759	4.05	0.05, 8.06	0.06	I^2 = 99%; p- value < 0.00001
AIS vs Neurological Disorders							
	CRP	2 (31,32)	389	10.87	-6.47, 28.49	0.23	I^2 =100%; p- value < 0.00001
AIS vs Stroke Mimic							
	S100B	2 (33,34)	142	0.60	0.26, 0.94	0.0005	I^2 = 0%; p- value = 0.63
AIS vs ICH							
	GFAP	3 ((35–37)	352	1.17	-0.09, 2.44	0.07	I^2 = 94%; p- value < 0.00001
	NT-BNP	2 (38,39)	133	-0.28	-0.66, 0.10	0.15	I^2 = 0%; p- value = 0.83
	BNP	2 (40,41)	266	-0.60	-1.77, 0.57	0.31	I^2 = 94%; p- value < 0.0001
	D-Dimer	2 (40,41)	266	-0.16	-1.46, 1.13	0.80	I^2 = 95%; p- value < 0.00001
	MMP-9	2 (40,41)	266	-0.93	-0.30, -1.55	0.004	I^2 = 77%; p- value = 0.04
	S100B	4 (33,37,40,41)	516	1.02	-0.24, 2.28	0.11	I^2 = 96%; p- value < 0.00001

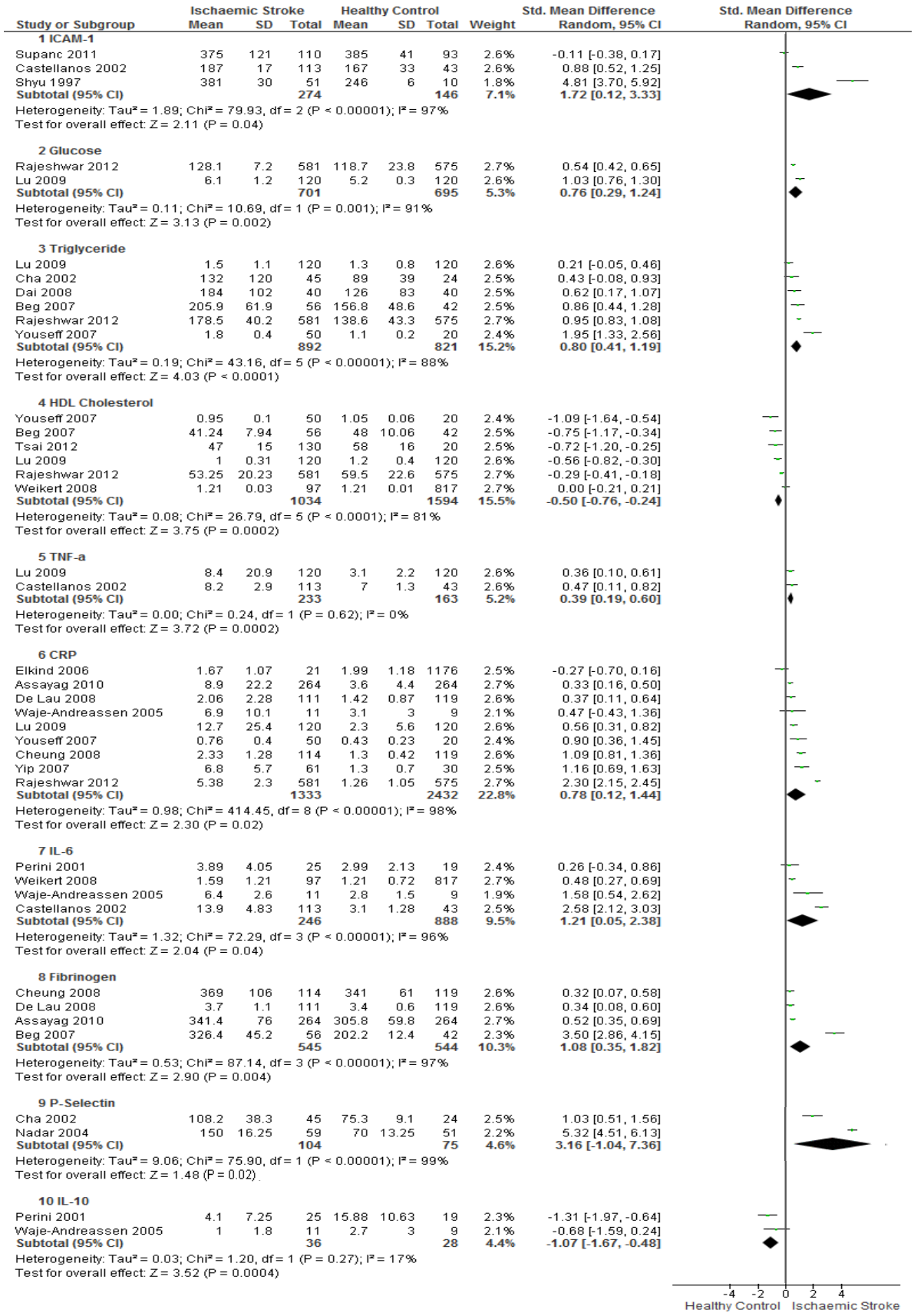


Figure 2.3: Forest plot of candidate diagnostic biomarkers able to significantly differentiate between acute ischaemic stroke patients and he althy controls. CI, confidence interval; **CRP,** C-reactive protein; **HDL**, high-density cholesterol; **ICAM-1,** intercellular adhesion molecule-1; **IL,** interleukin; **TNF-α**, tumour necrosis factor-α.

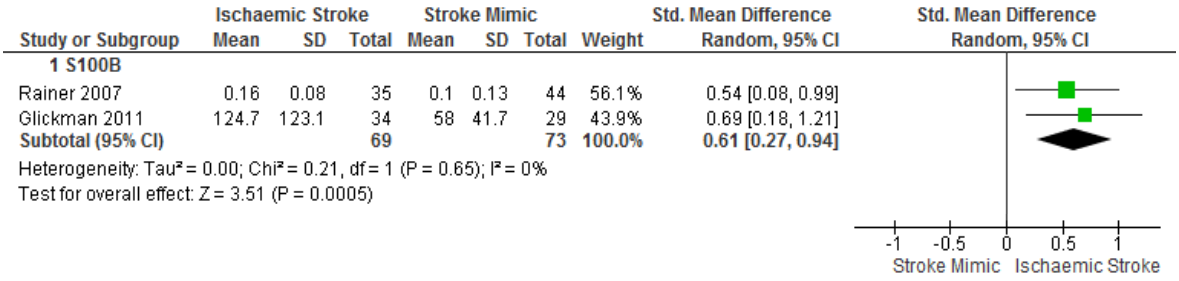


Figure 2.4: Forest plot comparing S100B levels in acute ischaemic stroke patients and stroke mimics. High S100B levels show a significant association with acute ischaemic stroke. **CI**, confidence interval.

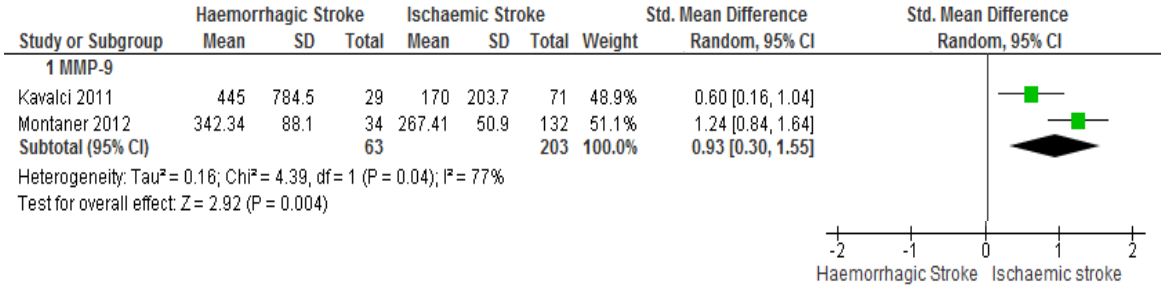


Figure 2.5: Forest plot comparing MMP-9 levels in acute ischaemic stroke and haemorrhagic stroke patients. Lower MMP-9 levels show a significant association with acute ischaemic stroke. **CI**, confidence interval; **MMP-9,** matrix metalloproteinase-9.

2.3.1.2 Prognostic Biomarkers of Ischaemic Stroke

Table 2.2 describes the SMDs associated with each candidate prognostic biomarker for which meta-analysis could be performed. Studies reported a number of post-AIS outcome measures, of which four were examined in more than one study cohort and amenable to meta-analysis: haemorrhagic transformation (HT) vs. no haemorrhagic transformation, in-hospital survival vs. death and good vs. bad outcome at 3 months (as measured modified Rankin Score or Barthel Score).

Prognostic biomarkers of significant association with ischaemic stroke were identified in two outcomes. Creatinine (SMD=-0.33, 95% CI -0.66,-0.01, p=0.04), D-dimer (SMD=-4.34, 95% CI -8,49, -0.18, p=0.04), fibrinogen (SMD=-5.51, 95% CI -9.06,-1.97, p=0.002) and glucose levels (SMD=-0.72, 95% CI -1.22,-0.22, p=0.005) were all higher in patients who died in hospital as compared to those who survived (Figure 2.6). High admission glucose levels were also found to be associated with poor functional outcome and death (mRS 3-6) at 3 months after stroke onset (SMD=-0.72, 95% CI -1.44, 0.01, p=0.04). Additionally, high fibrinogen levels were associated with poor outcome at 3 months (SMD=-0.59, 95% CI -1.10, -0.08, p=0.02) following iterative analysis and exclusion of one study (Welsh et al. 2009) that accounted for high heterogeneity.

Table 2.2: Candidate prognostic biomarkers of ischaemic stroke identified through literature review. Results from RevMan meta-analysis for candidate biomarkers reported in two studies or more are detailed, and biomarkers of significance are highlighted in blue. **CI**, confidence interval; **CRP**, c-reactive protein; **HbA1c**, haemoglobin A1C; **HT**, haemorrhagic transformation; **ICAM**, intracellular adhesion molecule; **IL**, interleukin; **mRS**, modified rankin scale; **SMD**, standardised mean differences; **TNF**, tumour necrosis factor. References in the table are numbered and listed separately at the end of the References section of the thesis.

	Biomarker	Number of Studies (References)	Number of Subjects	SMD	95% CI	p-value	Inter-study Heterogeneity (I^2; p-value)
HT vs no HT							
	Glucose	3 (42–44)	551	0.59	-0.06, 1.25	0.08	I^2 = 90%; p- value < 0.0001
	Fibrinogen	3 (42–44)	551	0.04	-0.43, 0.50	0.88	I^2 = 82% ; p- value = 0.004
Good Outcome (mRS 0-2) vs. Poor Outcome (mRS 3-6)							
	Glucose	5 (45–49)	2061	-0.72	-1.44, 0.01	0.04	I^2 = 98%; p- value < 0.00001
	Fibrinogen	5 (45,47–50)	2187	-0.14	-0.91, 0.62	0.71	I^2 = 98%; p- value < 0.00001
	Glutamate	3 (46,49,51)	616	-1.59	-3.50,0.32	0.10	I^2 = 99%; p- value < 0.00001
	IL-6	4 (48,50–52)	1146	-0.21	-2.96, 2.54	0.88	I^2 = 99%; p- value < 0.00001
	TNF-a	3 (50–52)	396	1.00	-0.06, 2.05	0.06	I^2 = 94%; p- value < 0.00001
	CRP	3 (46,48,50)	984	1.67	-1.46, 4.80	0.30	I^2 = 99%; p- value < 0.00001
	D-Dimer	2 (46,50)	234	0.37	-2.37, 3.11	0.79	I^2 = 99%; p- value < 0.00001
Good Outcome (Barthel >85) vs. Poor Outcome (Barthel <85)							
	ICAM-1	2 (1,3)	184	-0.29	-1.21, 0.62	0.53	I^2 = 87%; p- value = 0.005
Survival Inhospital vs. Death In-hospital							
	Glucose	3 (57–59)	1401	-0.72	-1.22, -0.22	0.005	I^2 = 75%; p- value = 0.02
	Total Cholesterol	2 (58,60)	566	0.33	-0.01, 0.67	0.06	I^2 = 0% ; p- value = 0.68
	CRP	2 (32,60)	431	-7.89	-16.90, 1.11	0.09	I^2 = 99% ; p- value < 0.00001
	Fibrinogen	2 (32,60)	431	-5.51	-9.06, -1.97	0.002	I^2 = 97% ; p- value < 0.00001
	D-Dimer	3 (32,58,60)	766	-4.34	-8.49, -0.18	0.04	I^2 = 99%; p- value < 0.00001
	HbA1c	2 (58,61)	352	0.74	-1.23, 2.70	0.46	I^2 = 87% ; p- value = 0.005
	Albumin	3 (32,58,61)	751	0.29	-1.00, 1.58	0.66	I^2 = 88% ; p- value = 0.0003
	Creatinine	2 (58,59)	441	-0.33	-0.66, -0.01	0.04	I^2 = 0% ; p- value = 0.78
	Sodium	2 (58,59)	441	-0.12	-0.69, 0.44	0.67	I^2 = 67% ; p- value = 0.08

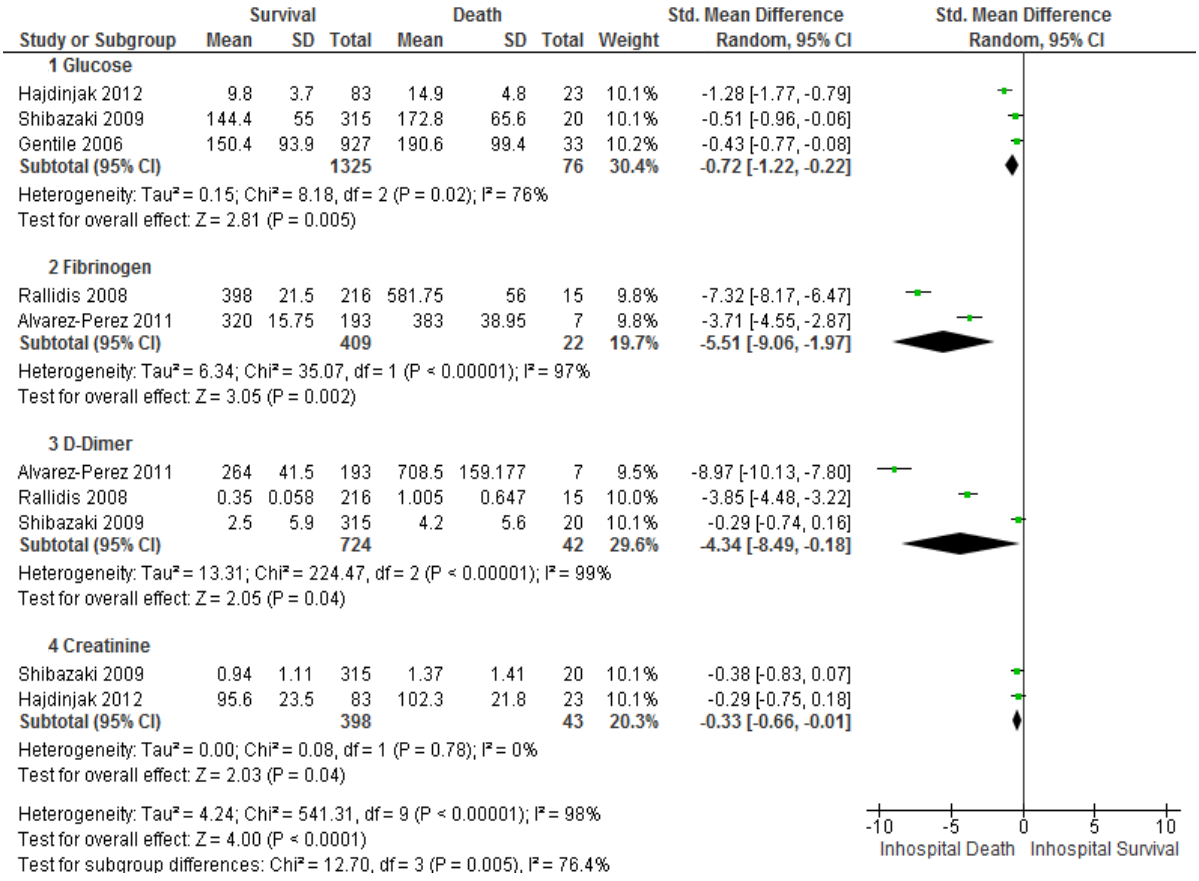


Figure 2.6: Forest plot of candidate biomarkers significantly associated with in-hospital death after ischaemic stroke. CI, confidence interval.

2.3.1.3 Prognostic Biomarkers for Thrombolysis

Thirty-nine papers reporting candidate prognostic biomarkers for intravenous rt-PA thrombolysis were identified for meta-analysis. These papers reported a number of prognostic outcomes, including survival vs. death, poor vs good outcome, efficiency of recanalisation and occurrence of haemorrhagic transformation and/or symptomatic intracerebral haemorrhage (SICH). Table 2.3 describes the SMDs associated with each candidate prognostic biomarker for which meta-analysis could be performed.

Of these, only one outcome, risk of SICH, was found to have candidate biomarkers of statistical significance upon meta-analysis. Meta-analysis of two studies found that lower levels of admission PAI-1 are associated with a higher risk of a patient developing SICH post-thrombolysis (SMD -0.54, 95% CI -1.09, 0.01, p=0.04). Admission glucose levels were also found to predict post-thrombolytic SICH. Results of this separate systematic review and meta-analysis are detailed in Section 2.3.2.

2.3.1.4 Clinical Relevance of Statistically Significant Biomarkers

Statistical association between a candidate biomarker and an outcome does not necessarily imply that the biomarker can discriminate between two population groups adequately (Pepe 2004), and moderately large effect sizes calculated by meta-analysis may still represent a substantial overlap in the distributions between study populations (Cook 2007). Similarly, in this current study, a difference in means between two populations does not eliminate the possibility of considerable overlap between the two distributions and high rates of false positives and false negatives at set thresholds. This concept was examined by plotting distribution curves from studies included in our meta-analysis of CRP levels in ischaemic stroke and healthy controls. Figure 2.7 demonstrates that CRP lacks the sensitivity and specificity required for effective discrimination between the two study populations, casting doubt upon its value as a clinically valuable biomarker.

Table 2.3: Candidate biomarkers predicting prognosis after thrombolysis for ischaemic stroke, as identified through literature review. Results from RevMan meta-analysis for candidate biomarkers reported in two studies or more are detailed, and biomarkers of significance are highlighted in blue. **APC**, activated protein C; **CI**, confidence interval; **CRP**, c-reactive protein; **mRS**, modified rankin scale; **PAI,** plasminogen activator inhibitor; **PH**, parenchymal haemorrhage; **SICH**, symptomatic intracerebral haemorrhage; **SMD**, standardised mean differences; **TAFI**, thrombin-activatable fibrinolysis inhibitor. References in the table are numbered and listed separately at the end of the References section of the thesis.

	Biomarker	Number of Studies (References)	Number of Subjects	SMD	95% CI	p-value	Inter-study Heterogeneity (I^2; p-value)
Good Outcome (mRS 0-2) vs. Poor Outcome (mRS 3-6)							
	Glucose	2 (62,63)	541	-0.98	-3.59, 1.62	0.46	I^2 = 99%; p- value < 0.00001
PH vs no PH							
	APC	2 (64,65)	234	0.56	-0.64, 1.76	0.36	I^2 = 88%; p- value = 0.004
SICH vs No SICH							
	TAFI	2 (66,67)	191	0.21	-1.22, 1.65	0.77	I^2 = 85%; p- value = 0.01
	PAI-1	2 (66,67)	191	-0.54	-1.09, 0.01	0.04	I^2 = 0%; p- value = 0.34
Survival vs Death							
	CRP	2 (68,69)	254	-0.64	-2.26, 0.97	0.44	I^2 = 94%; p- value < 0.0001

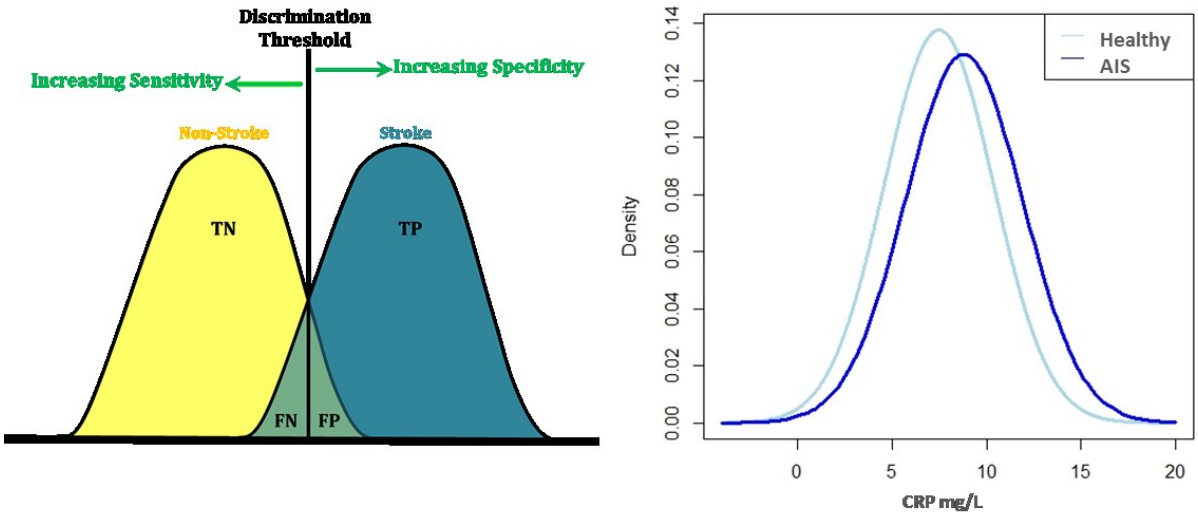


Figure 2.7: Assessing candidate biomarkers by examining distribution profiles. A) Biomarker discrimination limits are dependent upon the separation of distributions between two populations, affecting the numbers of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN). Moving an arbitrary discrimination threshold can alter the sensitivity and specificity of a biomarker test by changing the ratio of TPs and TNs. The discrimination threshold of a biomarker is usually reported to maximise the combination of sensitivity and specificity, as illustrated; however, the optimal discrimination threshold for a biomarker can vary depending on the relative importance of missing a TP versus that of misclassifying FPs. **B)** C-reactive protein (CRP) distribution profiles were created for ischaemic stroke and healthy control populations from data used in the meta-analysis. Despite the significant association between CRP and ischaemic stroke derived from meta-analysis, the distribution profiles indicate that an arbitrary discrimination threshold results in a high rate of FP and FNs, suggesting that CRP has questionable value as a diagnostic biomarker for acute ischaemic stroke.

2.3.2 Literature-Based Meta-Analysis of Admission Glucose Level as a Prognostic Biomarker for Post-Thrombolytic SICH

The findings of the systematic review demonstrated that admission glucose level could be a promising candidate biomarker for predicting prognosis following ischaemic stroke. It also has the advantage of being a simple, routinely measured blood-based clinical variable. Therefore the association between hyperglycaemia and post-thrombolytic SICH was investigated in more detail.

A separate literature search was conducted to obtain all studies reporting the association of admission glucose levels with SICH after intravenous rt-PA thrombolysis. Upon screening 204 studies, 66 studies reporting the presence of haemorrhagic transformation rather than SICH or examining prognosis after drugs other than intravenous rt-PA (Alteplase) were excluded. After reading the full texts of 138 papers, 116 were further excluded due to repeat data sets, or no quantitative assessment of the relationship between glucose levels and association with SICH. Thirteen observational studies and five clinical trials, encompassing 40,790 subjects, were identified for quantitative analysis (Figure 2.8).

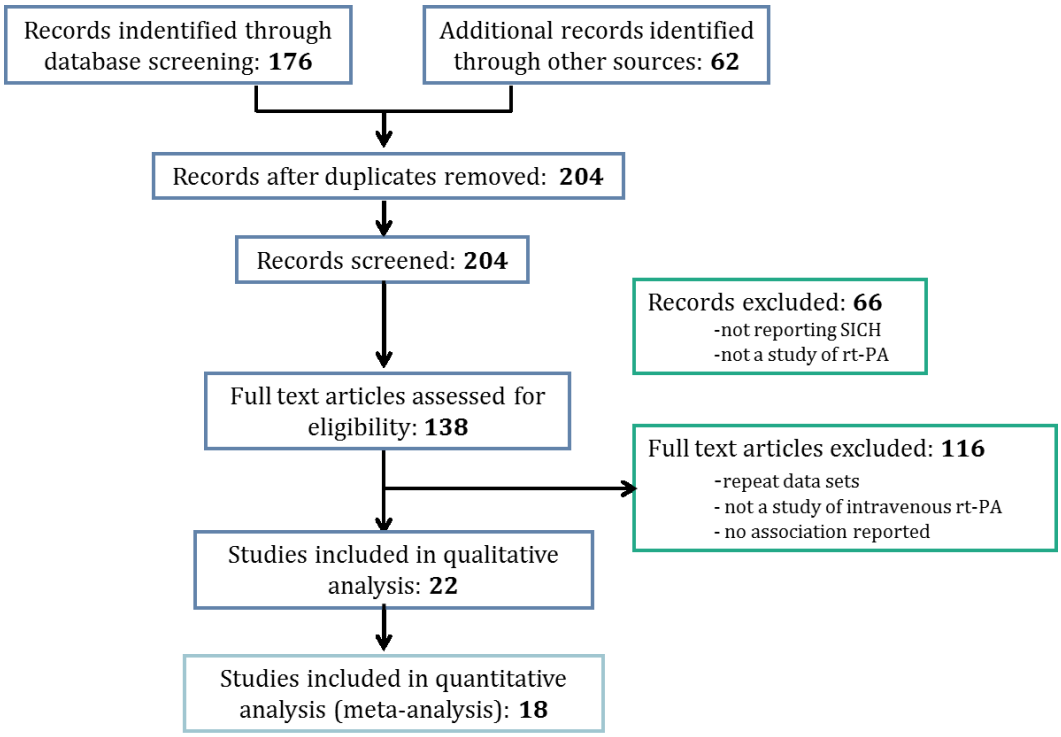


Figure 2.8 PRISMA flow diagram for systematic review and meta-analysis of glucose as a prognostic biomarker for post-thrombolytic SICH. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram illustrates the search strategy and number of studies included in the systematic review (qualitative analysis) and meta-analysis (quantitative analysis). **AIS**, acute ischaemic stroke, **SICH**, symptomatic intracerebral haemorrhage**, rt-PA**, recombinant tissue plasminogen activator.

Due to methodological differences between clinical trials and observational study protocols, it is not always appropriate to combine these two different study designs in a single meta-analysis (Professor Doug Altman, personal communication). For this reason, observational studies and clinical trials were analysed separately.

The mean admission glucose level in the 733 patients who developed SICH in the observational studies was calculated to be 8.7mmol/L where information was available, in contrast to a mean admission glucose level of 7.5mmol/L in the 30,611 patients who did not develop SICH. Eight observational studies reported continuous data, providing values of admission glucose levels (mean and standard deviation) for SICH and non-SICH populations (Table 2.4). Six observational studies reported the odds ratio (OR) of SICH per incremental increase in admission glucose level, with increments ranging from 1.0 – 5.6 mmol/L.

Table 2.4: Characteristics of studies included in the literature based meta-analysis of the association of glucose with post-thrombolytic symptomatic intracerebral haemorrhage. NR, not reported; **OR,** odds ratio, **SD**, standard deviation.

	Study	Data Reporting	SICH			non-SICH			% SICH
			Mean	SD	n	Mean	SD	n	
Observational Studies	Lansberg et al 2007	Continuous	9.8	5.4	7	7.3	2.8	67	9.5
	Katzan et al 2000	Continuous	11.0	4.7	11	8.4	3.4	59	15.7
	Tanne et al 2002	Continuous	8.8	3.5	72	8.3	2.9	86	6.0
	Alvarez-Sabin et al 2004	Continuous	9.5	1.7	8	8.6	3.8	130	5.8
	Cocho et al 2006	Continuous	9.6	4.8	8	7.6	3.7	106	7.0
	Strbian et al 2012	Continuous	7.2	0.8	68	6.6	0.4	906	7.0
	Mazya et al 2012	Continuous	7.1	0.5	546	6.6	0.4	29018	1.8
	Uyttenboogaart et al 2008	OR per increment and Continuous	6.9	2.3	13	6.4	1.9	239	5.2
	Meurer et al 2010	OR per increment	NR	NR	18	NR	NR	250	6.6
	Paciaroni et al 2009	OR per increment	NR	NR	2	NR	NR	63	3.2
	Naganuma et al 2011	OR per increment	NR	NR	25	NR	NR	553	4.3
	Hill et al 2005	OR per increment	NR	NR	52	NR	NR	1083	4.6
	Demchuk et al 1999	OR per increment	NR	NR	13	NR	NR	125	9.0
Clinical Trials	NINDS (Bruno et al 2002)	OR per increment	10.4	6.3	40	8.2	4.0	584	6.4
	ATLANTIS I & II (Bruno et al 2002)	OR per increment	NR	NR	NR	NR	NR	NR	5.2
	SAINT I & II (Cucchiara et al 2009)	OR per increment	7.5	2.9	54	7.3	2.7	911	5.6

2.3.2.1 Meta-Analysis of Observational Studies

Meta-Analysis of Observational Studies Reporting Continuous Data

Meta-analysis of the eight observational studies reporting means and standard deviations (SD) in SICH and non-SICH populations resulted in a pooled SMD of 0.87 (95% CI 2.61- 8.97; p<0.0001) using a random effects model. This indicates that a 2.3mmol/L (SMD x mean SD of 2.7) change in admission glucose levels can make the difference between a patient being safe or being at risk of developing SICH post-thrombolysis.

The OR of the association of glucose with SICH was calculated from the obtained SMDs for each of these studies. The resulting ORs and 95% CIs were then standardised to calculate the OR for SICH per 1mmol/L glucose (Table 2.5). Meta-analysis of these ORs using a random effects model resulted in a pooled OR of 2.41 (95% CI 1.23-3.59; p<0.0001) for the association of post-thrombolysis SICH with admission blood glucose level per 1mmol/L.

It was observed that the two largest datasets in the analysis (Mazya et al. 2012; Strbian et al. 2012) produced very large SMDs and consequently inflated ORs, most probably due to the small SDs commonly observed in large datasets. Removal of these two studies from the analysis resulted in a pooled OR of SICH per 1mmol/L glucose of OR 1.60 (95% CI 1.34-1.86; p<0.0001).

Meta-Analysis of Observational Studies Reporting Incremental Data

For the six observational studies reporting the OR of SICH for an incremental increase in admission glucose levels, data was standardised to calculate the OR for SICH per 1mmol/L of glucose (Table 2.5). Meta-analysis of these ORs resulted in a pooled OR 1.08 (95% CI 1.01-1.15; p<0.0001).

Pooled Meta-Analysis of Observational Studies

Studies reporting continuous data and studies reporting incremental data were combined in a single meta-analysis, using ORs calculated per mmol/L glucose from all studies, For one study (Uyttenboogaart et al. 2008a) reporting both continuous and incremental data, only the OR per1mmol/L calculated from incremental data was used in the meta-analysis. This resulted in a pooled OR 1.23 (95% CI 1.08-1.38; p<0.0001) (Table 2.5). However, Egger's regression test revealed a significant degree of publication bias (intercept value 2.65, standard error 0.64, two-tailed p-value 0.002). This was supported by outliers (Mazya et al. 2012; Strbian et al. 2012) found in the funnel plot of each study's odds ratio and standard error (Figure 2.9). Removal of these outliers still resulted in significant publication bias (intercept value 1.84, standard error 0.26, two-tailed p-value<0.0001).

2.3.2.2 Meta-Analysis of Clinical Trials

The three studies reporting data from five clinical trials (NINDS, ATLANTIS I and II, SAINT I and II) all reported the association of admission glucose levels with risk of SICH as OR per incremental increase in glucose levels. As with the observational studies, data was standardised to calculate the OR for SICH per 1mmol/L of glucose. Meta-analysis of these ORs resulted in a pooled OR of 1.01 (95% CI 0.87-1.14; p<0.0001) (Table 2.5).

2.3.2.3 Comment on the Evaluation of Biomarkers Using Literature Based Meta-Analysis

The benefits of analysing continuous rather than dichotomised/non-continuous data have been well described (Streiner 2002; Suissa & Blais 1995; Heritier & Ronchetti 2004; Royston et al. 2006). Continuous outcome measures usually provide greater analytical power that should be retained if possible (Chinn 2000; Streiner 2002). However, the use of SMDs to understand the relationship between glucose and risk of SICH is not ideal, as demonstrated by large SMDs calculated for the two largest studies (Mazya et al. 2012; Strbian et al. 2012). While meta-analysis of studies reporting ORs per increment of glucose may have led to a more accurate assessment of risk of SICH associated with admission glucose levels, the ORs reported by the respective studies are based upon either univariate analysis (Naganuma et al. 2011; Paciaroni et al. 2009; Cucchiara et al. 2009) or multivariate models incorporating different confounding variables (Uyttenboogaart et al. 2008a; Meurer et al. 2010; Hill et al. 2005; Demchuk et al. 1999). Therefore the ORs reported in these studies are not strictly comparable.

Meta-analysis of all published observational studies yielded OR 1.23 for the association of admission glucose levels with risk of post-thrombolytic SICH. However, the conversion of SMD to OR assumes that data in both the SICH and non-SICH study populations are normally distributed. In studies reporting glucose as a continuous variable, the standard deviations of glucose levels were higher in the SICH patients than non-SICH patients and large compared to the means, indicating that the data are not normally distributed (Altman & Bland 1996).

Additionally, analysis of data from a study by Uyttenboogaart et al. (2008) found that the OR of SICH per mmol/L glucose calculated from SMD was substantially larger than that reported in the published paper (OR 2.37 compared to OR 1.41). This discrepancy may be due to the fact that the relationship between glucose and OR may not be linear. This possibility, as well investigation of normal distribution and the effects of univariate or multivariate analysis cannot be examined using the published information available, but is essential for understanding the true prognostic value of glucose as a candidate biomarker for SICH following thrombolysis.

Table 2.5 Analysis of studies included in literature based meta-analysis of the association of glucose with post-thrombolytic symptomatic intracerebral haemorrhage. For studies reporting continuous data, standardised mean differences were calculated and then used to estimate odds ratios per 1mmol/L glucose. Studies reporting the association of symptomatic intracerebral haemorrhage (SICH) with incremental change sin admission glucose were standardised to an association (odds ratio) per 1mmol/L increase in admission glucose. Separate meta-analyses of observational studies and clinical trials were performed, using a random effects model to pool odds ratios per 1mmol/L glucose calculated for each study. **CI**, confidence interval.

		Standardised Mean Difference (SMD; 95% CI)	Odds Ratio per 1mmol/L Glucose (OR; 95% CI; p-value)
Observational Studies	Lansberg et al 2007	0.86	1.56; 95% CI: 1.14-2.14
	Katzan et al 2000	0.79	1.56; 95% CI: 1.00-2.44
	Tanne et al 2002	0.56	1.76; 95% CI: 1.04-2.97
	Alvarez-Sabin et al 2004	0.23	1.92; 95% CI: 0.26-14.07
	Cocho et al 2006	0.74	1.54; 95% CI: 1.05-2.26
	Strbian et al 2012	1.53	27.05; 95% CI: 15.62-46.86
	Mazya et al 2012	1.38	61.70; 95% CI: 47.77-79.70
	Uyttenboogaart et al 2008 (from continuous data)	0.26	2.37; 95% CI: 0.38-14.91
	Uyttenboogaart et al 2008 (from incremental data)	-	1.41; 95% CI: 0.96-1.92
	Meurer et al 2010	-	1.03; 95% CI: 0.86-1.23
	Paciaroni et al 2009	-	1.01; 95% CI: 1.00-1.01
	Naganuma et al 2012	-	1.11; 95% CI: 0.96-1.26
	Hill et al 2005	-	1.10; 95% CI: 1.04-1.18
	Demchuk et al 1999	-	1.16; 95% CI: 1.01-1.32
	Pooled Odds Ratio		**1.23; 95% CI: 1.08-1.38; p<0.0001**
Clinical Trials	NINDS (Bruno et al 2002)	-	1.11; 95% CI: 1.02-1.20
	ATLANTIS I & II (Bruno et al 2002)	-	0.84; 95% CI: 0.69-1.02
	SAINT I & II (Cucchiara et al 2009)	-	1.13; 95% CI: 1.05-1.20
	Pooled Odds Ratio		**1.01; 95% CI: 0.87-1.14; p<0.0001**

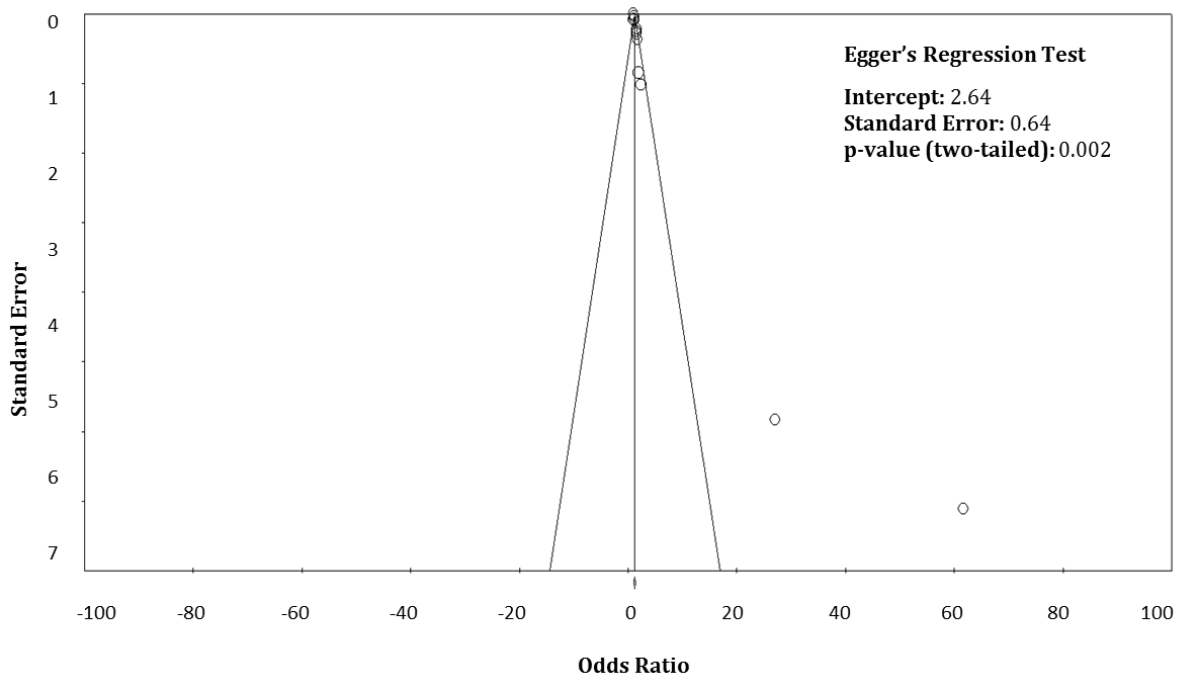


Figure 2.9 Examination of publication bias in all observational studies included in literature based meta-analysis of the association of glucose with post-thrombolytic symptomatic intracerebral haemorrhage. The funnel plot of odds ratio against standard error for each study revealed two outlying studies (Mazya et al. 2012; Strbian et al. 2012) and two-tailed Egger's Regression test found evidence of significant publication bias (p=0.002). Iterative analysis and removal of the two outlying studies still resulted in significant publication bias (p<0.0001).

2.3.3 Individual Patient Data Meta-Analysis of Admission Glucose Level as a Prognostic Biomarker for Post-Thrombolytic SICH

In light of the issues identified in performing a literature-based meta-analysis detailed in Section 2.3.2.3, it was concluded that an accurate investigation of the role of admission glucose levels in predicting post-thrombolytic SICH could only be performed by analysis of individual patient data (IPD).

Given the relatively low incidence of SICH within a thrombolysed patient population, reliable evaluation of the relationship between glucose levels and SICH requires accumulation of data from as many studies as possible. Therefore authors of 15 observational studies identified through systematic review and qualitative analysis were contacted for IPD, with the aim of quantifying the absolute risk of post-thrombolytic SICH associated with admission glucose levels.

Data was gained from 8 observational studies (53%), representing 83% of the potential patient data available (33,370 patients out of the possible 40,336 reported from literature). The characteristics of these 8 studies are detailed in Table 2.6. Mean admission glucose level was 6.59mmol/L (SD=0.34mmol/L), mean onset time to treatment 186 minutes (SD=73 minutes), and mean baseline NIHSS score 13 (SD=2.4) for studies from which information was available.

Table 2.6 Characteristics and patient demographics of studies contributing individual patient data for analysis of the association of admission glucose levels with risk of post-thrombolytic symptomatic intracerebral haemorrhage (SICH). Values are medians and interquartile ranges (IQR), unless specified otherwise. **ECASS**, European Cooperative Acute Stroke Study**; NINDS,** National Institute of Neurological Disorders and Stroke**; OTT,** onset to treatment time; **SITS-ISTR**, Safe Implementation of Thrombolysis in Stroke International Stroke Thrombolysis Register; **SITS-MOST,** Safe Implementation of Thrombolysis in Stroke-Monitoring Study.

	Alvarez-Sabin et al 2004	**Hill et al 2005**	**Lansberg et al 2007**	**Meurer et al 2010**	**Paciaroni et al 2009**	**Mazya et al 2012**	**Tanne et al 2002**	**Uyttenboogaart et al 2008**
Study Description	Spanish prospective cohort study	CASES Canadian prospective cohort study	DEFUSE multi-centre prospective open-label study	US multi-centre prospective cohort study	Italian multi-centre prospective cohort study	SITS-ISTR multi-national, prospective thrombolysis register	Multi-national prospective cohort study	Dutch prospective cohort study
Total Number of Subjects	406	1049	74	268	65	30088	784	636
Age	67.9 (10)	73 (63-80)	75 (64-82)	68 (60-76)	76 (62-88)	70 (60-77)	67 (58-72)	71 (58-79)
Women (%)	58 (34.3)	512 (45.1)	32 (43.2)	123 (45.9)	29 (44.6)	12647 (42.0)	624 (52.0)	290 (45.6)
OTT (mins)	141 (108-176)	155 (130-175)	315 (287-343)	-	-	145 (116-170)	-	174 (139–210)
NIHSS on admission	17 (9-25)	14(9-19)	12 (8-16)	-	-	12 (7-17)	-	11 (7-16)
Glucose, mmol/L	6.56 (5.39-8.00)	6.60 (5.70-8.10)	6.69 (5.65-7.96)	6.47 (5.51-7.99)	6.11 (5.28-7.25)	7.20 (5.70-7.90)	6.78 (5.72-8.22)	6.20 (5.40-7.40)
Diabetes Mellitus (%)	34 (20.2)	180 (15.9)	20 (27.0)	59 (22.0)	14 (20.7)	5172 (17.2)	240 (20.1)	29 (12.4)
SICH n (%)								
NINDS definition						2242 (8.2)	126 (16.1)	36 (5.7)
ECASS definition	20 (4.9)	49 (4.7)	5 (6.8)	23 (8.4)		1563 (5.6)		
SITS-MOST definition			5 (6.8)		2 (3.07)	543 (1.9)		

2.3.3.1 Absolute Risk of SICH Associated with Admission Glucose Level

Studies measured SICH by three different classifications: 1) SICH per the National Institute of Neurological Disorders and Stroke (NINDS) classification: any deterioration in NIHSS score or death within 7 days combined with intracerebral haemorrhage of any type, on any post-treatment imaging after the start of thrombolysis. 2) SICH per the European Cooperative Acute Stroke Study (ECASS) classification: any type of intracerebral haemorrhage on any post-treatment imaging after the start of thrombolysis and increase of ≥4 NIHSS points from baseline, or from the lowest value within 7 days, or leading to death. 3) SICH per the SITS-MOST classification: a local or remote type 2 parenchymal haemorrhage on imaging 22-36 hours after treatment combined with a neurological deterioration of ≥4 NIHSS points from baseline or from the lowest NIHSS score between baseline and 24 hours or leading to death within 24 hours. Meta-analyses were performed separately for each classification of SICH (NINDS, ECASS and SITS-MOST).

The incidence of SICH differed considerably depending on the classification of SICH, therefore each classification was considered separately in subsequent analyses. Three studies reported a NINDS classification of SICH (Mazya et al. 2012; Tanne et al. 2002; Uyttenboogaart et al. 2008) with a mean SICH rate of 10.0%, five reported an ECASS classification (Alvarez-Sabin et al. 2004; Hill et al. 2005; Lansberg et al. 2007; Meurer et al. 2010; Mazya et al. 2012) with a mean SICH rate of 6.1% and three reported a SITS-MOST classification (Lansberg et al. 2007; Paciaroni et al. 2009; Mazya et al. 2012) with a mean SICH rate of 3.9%.

Examination of glucose levels within each dataset found minimal evidence of a non-linear relationship between glucose levels and risk of SICH up to glucose levels of 30mmol/L. However, glucose levels in every dataset demonstrated a highly skewed distribution, therefore it was considered appropriate to log transform admission glucose values and use $\log_{10}$glucose in all subsequent regression analyses. Univariate logistic regression analysis performed independently on data from each study found a statistically significant association between admission glucose levels and risk of SICH in all but two studies (Paciaroni et al. 2009; Uyttenboogaart et al. 2008), with regression coefficients ranging from OR 2.03 (95%CI 1.43-2.63, p<0.0001) (Mazya et al. 2012; SICH per SITS-MOST classification) to OR 6.17 (95%CI 0.98-11.36, p=0.020) (Lansberg et al. 2007; SICH per SITS-MOST and ECASS classification) in those studies yielding a statistically significant association (Table 2.7). The small sample size and scarcity of SICH events in two studies (Paciaroni et al. 2009, n=65; Lansberg et al. 2007, n=74) was considered to result in a poor estimation of regression coefficients, therefore these studies were excluded from meta-analysis.

Table 2.7 Analysis of studies contributing individual patient data. Regression coefficients were calculated for each study in a univariate logistic model to assess the association between admission glucose levels and risk of symptomatic intracerebral haemorrhage (SICH). Glucose levels were log_{10} transformed before prior to regression analysis. **CI**, confidence interval; **ECASS**, European Cooperative Acute Stroke Study**; NINDS,** National Institute of Neurological Disorders and Stroke**; SITS-MOST,** Safe Implementation of Thrombolysis in Stroke-Monitoring Study.

	All Patients	Mean Admission Glucose mmol/L (SD)		Odds Ratio	95% CI	p-value
	SICH/ Total Population (%)	SICH	No SICH			
Alvarez-Sabin et al 2004						
ECASS	19 / 406 (4.68%)	7.96 (4.32)	7.00 (3.53)	3.01	0.19-5.83	0.037
Hill et al 2005						
ECASS	43 / 1098 (4.47%)	8.67 (4.15)	7.40 (2.86)	2.32	0.52-4.12	0.012
Lansberg et al 2007						
ECASS	5 / 74 (6.76%)	11.35 (5.70)	7.28 (2.77)	6.17	0.98-11.36	0.020
SITS-MOST	5 / 74 (6.76%)	11.35 (5.70)	7.28 (2.77)	6.17	0.98-11.36	0.020
Meurer et al 2010						
ECASS	23 / 268 (8.58%)	8.37 (3.34)	7.16 (2.78)	2.76	0.03-5.50	0.048
Paciaroni et al 2009						
SITS-MOST	2 / 65 (3.08%)	5.78 (1.10)	6.50 (2.32)	-0.99	-8.48-6.5	0.796
Mazya et al 2012						
SITS-MOST	543 / 29440 (1.84%)	7.94 (3.11)	7.17 (2.50)	2.03	1.43-2.63	<0.0001
ECASS	1563 / 29429 (7.60%)	8.08 (3.11)	7.14 (2.47)	2.56	2.21-2.92	<0.0001
NINDS	2242 / 29497 (16.07%)	8.03 (3.07)	7.12 (2.45)	2.55	2.25-2.85	<0.0001
Tanne et al 2002						
NINDS	126 / 658 (16.07%)	8.52 (3.13)	7.54 (3.23)	2.45	1.25-3.65	<0.0001
Uyttenboogaart et al 2008						
NINDS	36 / 636 (5.66%)	7.51 (2.73)	6.56 (1.90)	1.45	-0.16-3.06	0.077

Pooling of ORs obtained from individual regression analysis in meta-analyses gave the following results: OR 2.51 (95%CI 2.22-2.80, p<0.001) for SICH per NINDS classification, OR 2.56 (95%CI 2.22-2.90, p<0.001) for SICH per ECASS classification. Only one study (Mazya et al. 2012) reported SICH per SITS-MOST classification after exclusion of two studies (Paciaroni et al. 2009; Lansberg et al. 2007), giving OR 2.03 (95% CI 1.43-2.63, p<0.001) (Figure 2.10). Calculation of absolute risk of SICH associated with each mmol/L of glucose found that at the current thrombolysis glucose threshold of 22 mmol/L, the absolute risk of a patient developing SICH is 11.9% per NINDS classification, 8.2% per ECASS classification and 2.7% per SITS-MOST classification (Figure 2.11).

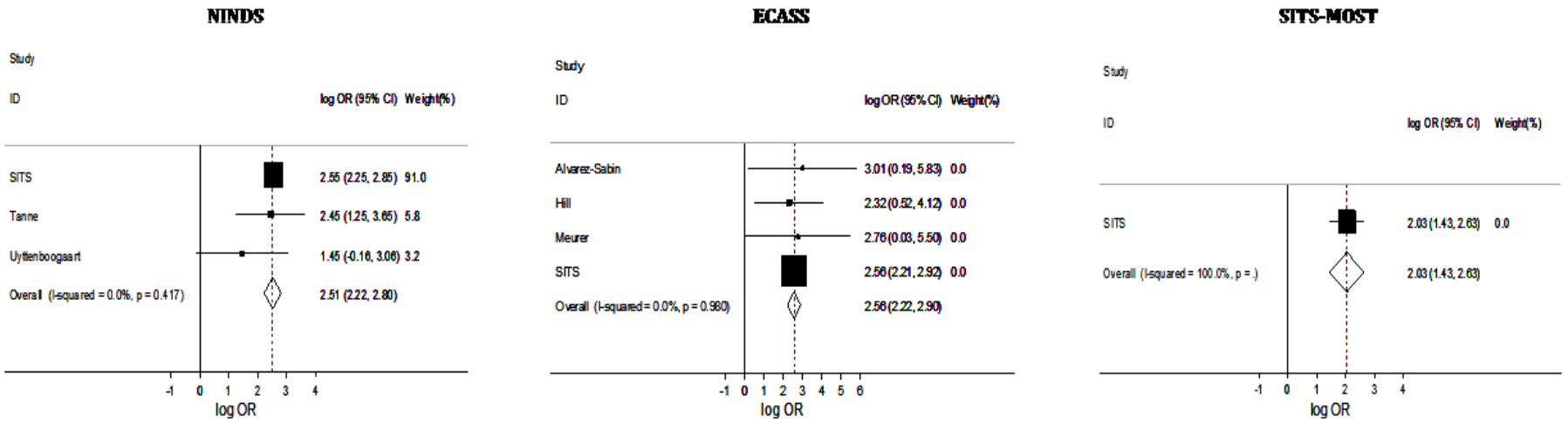


Figure 2.10: Meta-analysis of studies contributing individual patient data. A two-step approach meta-analysis was performed using weighted log odds ratios calculated from individual patient data of each study cohort. Meta-analyses were performed separately for each definition of symptomatic intracerebral haemorrhage (NINDS, ECASS, SITS-MOST) using a fixed effects model. **ECASS**, European Cooperative Acute Stroke Study**; I-squared (I^2),** measure of heterogeneity; **NINDS,** National Institute of Neurological Disorders and Stroke**; SITS-MOST,** Safe Implementation of Thrombolysis in Stroke-Monitoring Study.

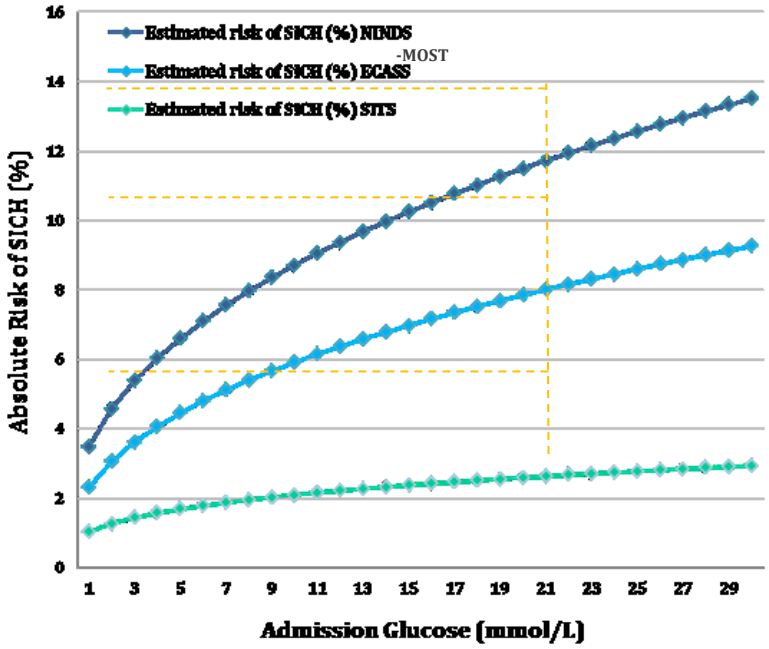


Figure 2.11 Estimated absolute risk of post-thrombolytic SICH associated with admission glucose levels. Using the odds ratios obtained from two-step meta-analysis, absolute risk of symptomatic intracerebral haemorrhage (SICH) associated with each 1mmol/L increment of admission glucose level was calculated for the three definitions of SICH. For example, at the current thrombolysis glucose threshold of 22mmol/L, an individual would be expected to have an 11.9% absolute risk of SICH per NINDS definition, 8.2% per ECASS definition and 2.7% per SITS-MOST definition. **ECASS**, European Cooperative Acute Stroke Study**; NINDS,** National Institute of Neurological Disorders and Stroke**; SITS-MOST,** Safe Implementation of Thrombolysis in Stroke-Monitoring Study.

Exploratory multivariable analysis on the largest study cohort (Mazya et al. 2012; >30,000 patients) found that risk of SICH was associated with four variables in addition to admission glucose level: age, baseline NIHSS score, hypertension and stroke onset to time of treatment (OTT). Multivariate logistic regression analysis still yielded a significant association between admission glucose level and SICH: OR 2.55 (95%CI 2.25-2.85, p<0.001) for SICH per NINDS classification, OR 2.20 (95%CI 1.81-2.58, p<0.001) for SICH per ECASS classification and OR 1.78 (95%CI 1.15-2.41, p<0.001) for SICH per SITS-MOST classification. Furthermore, the analysis did not substantially alter the absolute risk of SICH, with absolute risk of SICH associated at an admission glucose level of 22mmol/L calculated as 11.2% per NINDS classification, 7.8% per ECASS classification and 2.4% per SITS-MOST classification. Significant interaction effects (p<0.05) were not observed between any of the five variables.

2.3.3.2 Association of SICH with Poor Functional Outcome

While patients with SICH have been shown to have an increased risk of poor and fatal outcome post-thrombolysis (Strbian et al. 2011), not all SICH events may be detrimental. The relationship between SICH and poor functional outcome and survival was assessed in the largest study cohort (Mazya et al. 2012; >30,000 patients). Incidence of mortality and dependence (mRS 3-6) within 3 months follow-up was significantly greater in the SICH population than in the non-SICH population, regardless of the classification of SICH used (p<0.0001). SICH per SITS-MOST classification was the strongest predictor of mortality (OR 3.17, 95% CI 2.95-3.38, p<0.001, compared to NINDS classification: OR 2.46, 95% CI 2.37-2.56, p<0.001, ECASS classification: OR 3.09, 95% CI 2.96-3.12, p<0.001). SICH per ECASS classification was found to be the strongest predictor of dependence at 3 months (OR 3.70, 95% CI 3.41-3.99, p<0.001, compared to NINDS classification: OR 2.91, 95% CI 2.74-3.08, p<0.001, SITS-MOST classification: OR 3.33, 95%CI 2.90-3.75, p<0.001) (Table 2.8). Nonetheless, differences between SICH classifications were marginal, as indicated by overlapping confidence intervals.

Table 2.8 Investigation of functional outcome in patients with SICH. Three-month outcomes in patients with and without symptomatic intracerebral haemorrhage (SICH) were assessed in the largest study cohort (Mazya et al. 2012; >30,000 patients). Dependence was defined as modified rankin score (mRS) of 3-6 and independence was defined as mRS 0-2. Events and non-events (alive vs dead, independence vs dependence) were examined for statistical significance of using Pearson's Chi-square (χ^2) test. Odds ratios (ORs) of the association between each definition of SICH and functional outcome were calculated using a random effects model. **CI**, confidence interval.

	SICH NINDS				SICH ECASS				SICH SITS-MOST			
	No. of Patients / Total Population (%)				No. of Patients / Total Population (%)				No. of Patients / Total Population (%)			
	SICH	No SICH	p value	OR (95% CI); p value	SICH	No SICH	p value	OR (95% CI); p value	SICH	No SICH	p value	OR (95% CI); p value
Mortality within 3months	1179 / 2024 (58.3%)	2451 / 23076 (10.6%)	<0.0001	2.46 (2.37-2.56); <0.001	1059 / 1457 (72.7%)	2556 / 23593 (10.8%)	<0.0001	3.09 (2.96-3.21); <0.001	394 / 505 (78.0%)	3180 / 24474 (13.0%)	<0.0001	3.17 (2.95-3.38); <0.001
Dependence at 3 months	1864 / 2010 (92.7%)	9352 / 22739 (41.1%)	<0.0001	2.91 (2.74-3.08); <0.001	1400 / 1448 (96.7%)	9770 / 23252 (42.0%)	<0.0001	3.70 (3.41-3.99); <0.001	483 / 505 (95.4%)	10633 / 24124 (44.1%)	<0.0001	3.33 (2.90-3.75); <0.001

2.4 Discussion

While numerous molecules have been suggested as diagnostic and prognostic biomarkers of ischaemic stroke, their clinical relevance is unclear. To assess the utility of blood-based biomarkers in improving the diagnosis and prognosis of ischaemic stroke in the acute phase (within 24 hours of onset), a systematic literature review was conducted.

Systematic literature review identified 137 blood-based candidate biomarkers implicated for diagnosis or prognosis of ischaemic stroke. Subsequent meta-analysis found that few withstood statistical scrutiny, with only 12 diagnostic biomarkers (ICAM-1, glucose, HDL cholesterol, triglyceride, TNF-α, CRP, IL-6, fibrinogen, P-selectin, IL-10, S100B and MMP-9) and 5 prognostic biomarkers (glucose, fibrinogen, D-Dimer, creatinine and PAI-1) being significantly associated with acute ischaemic stroke. These findings demonstrate the value of performing meta-analysis in order to discount potential candidate biomarkers that have inflated associations in single, under-powered studies. It also highlights the importance of designing well-powered studies for biomarker discovery, which contain multiple independent levels of replication and validation, in order to reduce false positive associations. In line with this, Ioannidis & Panagiotou (2011) recently commented upon the risk of individual studies exaggerating biomarker associations and the caution needed in interpreting the reported results.

Diagnostic and Prognostic Biomarkers of Ischaemic Stroke

Ten candidate biomarkers were identified that could significantly differentiate between ischaemic stroke and healthy controls. The majority of these candidates fell into three main categories: endothelial cell adhesion (ICAM-1, P-selectin), lipid metabolism (triglyceride, HDL cholesterol) and immune response (TNF-α, CRP, IL-6, IL-10). Upregulation of ICAM-1 and P-selectin has been widely reported in experimental models of cerebral ischaemia (Nishiwaki et al. 2003; Love & Barber 2001), being responsible for neutrophil adhesion and an indication of compromised vessel integrity (Zhang et al. 1999). Dyslipidaemia is a well-known risk factor for CVD and high triglyceride and low HDL cholesterol levels have been reported during acute ischaemic stroke, although this association likely represents an intricate biological interaction that may not be most suitably represented by the results of previous multivariate analyses (Miller et al. 2011; Kisialiou et al. 2012).

A number of studies have also established a key role for inflammatory responses in atherosclerosis, cardiovascular disease and in the acute phase pathophysiology of ischaemic stroke (Ross 1999; Pearson et al. 2003; Tamam et al. 2005). Initial inflammation during acute

ischaemic stroke stimulates the production of acute phase proteins involved in the innate immune response, including TNF-α, fibrinogen and CRP. Consistent with this knowledge, acute phase proteins were found to be significantly higher in patients with ischaemic stroke as compared to healthy controls in this meta-analysis. In particular, CRP has gained popularity in recent years and has been proposed to be useful as a risk stratifying, diagnostic and prognostic biomarker for ischaemic stroke (Di Napoli 2001). However, in the current study, deeper examination of CRP levels in ischaemic stroke and healthy control populations demonstrated considerable overlap in distribution profiles, indicating that despite CRP being a recognised pathophysiological player in ischaemic stroke, a biomarker test based upon CRP is unlikely to provide the sensitivity or specificity required for a clinically valuable diagnostic tool (Greenland & O'Malley 2005).

The candidate biomarkers above can differentiate between ischaemic stroke patients and healthy controls. However, the clinical relevance of such biomarkers is questionable. An ideal biomarker test should be able to distinguish between ischaemic strokes patients and individuals with suspected stroke and stroke-like symptoms (stroke mimics), to replicate a realistic clinical environment. One candidate biomarker, MMP-9, was found to be able to significantly differentiate between ischaemic and haemorrhagic stroke. Differentiating between ischaemic and haemorrhagic stroke is a critical step in planning immediate stroke management. Both types require specific protocols of action, for example, rt-PA treatment is beneficial to ischaemic stroke patients but can be potentially lethal to haemorrhagic stroke patients. While CT imaging can distinguish between ischaemic and haemorrhagic stroke types with high sensitivity, a blood-based biomarker test for stroke could allow patients to be triaged even before entering the hospital e.g. inside an ambulance. MMP-9 plays an important role in cell adhesion and extra-cellular matrix breakdown, thus mediating the integrity of microvasculature and the blood-brain barrier, explaining why MMP-9 levels were found to be much higher in haemorrhagic stroke as compared to ischaemic stroke patients (Kavalci et al. 2011).

A diagnostic biomarker test with the ability to distinguish between ischaemic stroke and stroke mimics would also be clinically useful for effective stroke management. Systematic review of the literature revealed few studies examining biomarker differences between ischaemic stroke patients and stroke mimics (Glickman et al. 2011; Ahn et al. 2011; Foerch et al. 2012; Doehner et al. 2012; Airas et al. 2008; Rainer et al. 2007). Only one biomarker, S100B, was reported in more than one study and amenable to meta-analysis. S100B is an astrocyte-derived protein known to be elevated during ischaemic stroke and cerebral damage. Disruption of the blood-brain barrier during stroke causes the release of brain-derived molecules into the bloodstream (Aoki et al. 2002; Kastrup et al. 2008; Jin et al. 2010). Thus blood-based, brain-derived

biomarkers have the potential to be accurate measures of cerebral damage following stroke. However, S100B is a common brain-derived protein which is up-regulated in a number of traumatic situations, therefore its specificity and sensitivity in a clinical setting is debatable.

Candidate biomarkers identified as being significantly associated with poor prognosis following ischaemic stroke could be classified as metabolites (glutamate, glucose, creatinine) and proteins involved in coagulation and fibrinolysis (fibrinogen, D-dimer, PAI-1). The amino acid glutamate is the most abundant neurotransmitter in the brain, and it is likely that high levels are an indication of the extent of neuronal damage and therefore prognosis following stroke (Jin et al. 2010). Elevated creatinine levels have also been associated with ischaemic stroke, implicated as an indication of subtle renal damage consequent to raised blood pressure and thus a risk factor for cerebrovascular disease (Wannamethee et al. 1997).

The benefits of thrombolytic therapy for AIS have been well documented, significantly reducing the proportion of patients dead or dependent six months after stroke (Wardlaw 2009). Nevertheless haemorrhagic transformation is a common complication, and while SICH occurs in an estimated 6% of treated stroke patients, it has a fatality rate of 50-80% (Hacke et al. 2004). Studies have shown that poor outcome and risk of intracranial haemorrhage following tPA administration increases outside the optimum 3-hour treatment window, determined in part by disruption of the blood-brain barrier (Aoki et al. 2002; Kastrup et al. 2008; Jin et al. 2010). However, clinicians may overestimate the risk of ICH when considering rt-PA treatment and an excessive fear of haemorrhagic transformation complications remains a key reason for clinicians to withhold thrombolytic therapy (Derex & Nighoghossian 2008). Therefore a prognostic biomarker would be helpful for predicting poor outcome following thrombolysis and aiding patient selection and stratification, allowing safe administration thrombolytic therapy as quickly as possible.

PAI-1 was found to be significantly associated with risk of SICH post-thromobolysis. PAI-1 is a protein involved in the coagulation/fibrinolysis cascade, and is a major inhibitor of tissue plasminogen activator, the molecule found in licensed thrombolytic therapy. Thus the decrease in PAI-1 in patients developing SICH can be explained by the fact that lower PAI-1 levels enhance a fibrinolytic state, predisposing to intracranial bleeds (Ribo et al. 2004).

In addition to PAI-1, elevated admission glucose levels were shown to be associated with post-thrombolytic SICH as well as poor outcome (mRS 3-6) and in-hospital death following ischaemic stroke. It is well established that conditions such as stroke and myocardial infarction can lead to an increase in glucose levels, with hyperglycaemia occurring in up to 60% of stroke patients

without known diabetes (Kiers et al. 1992; Van Kooten et al. 1993). There is also ample evidence to suggest that high glucose levels are detrimental in cerebral ischaemia, as summarised in a meta-analysis by Capes et al. (2001). Hyperglycaemia has been shown to aggravate microvascular lesions, leading to a cascade of events including cellular acidosis, increased free radical production and the initiation of cell death. With the up-regulation of matrix metalloproteinases (MMPs) during thrombolytic therapy, ischaemic damage to the vessel basal lamina may increase, damaging the integrity of vessel walls and causing haemorrhage (Montaner et al. 2003).

Admission Glucose Level as a Prognostic Biomarker of Post-Thrombolytic SICH

The possible use of glucose in a blood-based biomarker to predict risk of SICH has the advantage of glucose being a simple, routinely measured clinical variable, thus removing the need for a test based upon biomarkers that could potentially require longer and more complex laboratory analysis.

High blood glucose levels can be particularly detrimental after thrombolysis, with hyperglycaemic patients being less likely to recanalise (Ribo et al. 2005) and more likely to have SICH and poor outcome, even after recanalisation (Demchuk et al. 1999; Alvarez-Sabin et al. 2003; Dharmasaroja et al. 2012). To explore the prognostic value of admission glucose values in predicting post-thrombolytic SICH, separate literature-based systematic review and meta-analyses were performed. Meta-analyses of observational studies and clinical trials suggested a small yet significant association of glucose levels with SICH. However, these findings were subject to a number of limitations, as discussed in Section 2.3.2.3, and could not provide accurate quantification of the relationship between admission glucose levels and absolute risk of SICH. Therefore an individual patient data (IPD) analysis, consisting of 32,000 subjects and the largest of its kind to date, was conducted to address these issues.

Current international glucose guidelines for thrombolysis, set at 22mmol/L, are based upon exclusion criteria from the NINDS trial (n=624) in 1995. However, these guidelines have no scientific foundation, and post-hoc analysis of the NINDS trial found that admission blood glucose level is associated with poor outcome and SICH post thrombolysis (Ingall et al. 2004). A number of recent studies have also supported this finding: Putaala et al. (2011) recently reported that hyperglycaemia (≥8.0 mmol/l) up to 48 hours after intravenous rt-PA is strongly associated with poor outcome, SICH and death; Demchuk et al. (1999) reported that glucose levels >11mmol/L are associated with a 25% increased risk of SICH and Alvarez-Sabin et al. (2004) that a value of >7.7mmol/L may counterbalance the beneficial effect of thrombolysis.

A survey by the American College of Emergency Physicians (Brown et al. 2005) found that its members would only tolerate a risk of 3.4% SICH before thrombolysing (as compared to the 6.4% risk reported in the NINDS trial). The results of this current IPD study found that at the current thrombolysis guideline admission glucose threshold of 22mmol/L, the absolute risk of a patient developing SICH was 11.9% using the NINDS classification, 8.2% with the ECASS classification and 2.7% with the SITS-MOST classification. Furthermore, using the SITS-MOST classification of SICH, the role of admission glucose in predicting SICH is slight (2.7% risk of SICH at 22mmol/L, 3.0% risk at 30mmol/L). Thus, contrary to previous studies, the findings lend evidence to suggest that the current glucose guideline threshold could be extended beyond 22mmol/L.

Many patients who experience SICH may already be destined for a poor outcome and may not have had their outcome altered as a result of the haemorrhage (Saver 2007). The Helsinki Stroke Thrombolysis Registry Group recently reported that despite SICH having a small contribution to stroke prognosis, patients with SICH had increased risk of death and poor outcome, with the ECASS (rather than NINDS or SITS-MOST) classification of SICH providing the largest contribution to worst outcome (Strbian et al. 2011). Similarly, the current study found a larger incidence of poor prognosis at 3 months in SICH patients than non-SICH patients, with SICH per SITS-MOST classification being the strongest predictor of mortality and SICH per ECASS classification being the strongest predictor of dependence (mRS 3-6).

When interpreting these results, it is also important to remember that hyperglycaemia is not the only factor contributing to poor prognosis following thrombolysis. A number of multivariate scores have been developed for predicting SICH post-thrombolysis (HAT score, 3 variables (Lou et al. 2008); SEDAN score, 5 variables (Strbian et al. 2012); SITS score, 9 variables (Mazya et al. 2012); Multicenter Recombinant Tissue-Type Plasminogen Activator Stroke Survey Group score, 4 variables (Cucchiara et al. 2008)). However, these scores are often limited by the consideration of variables in a dichotomised fashion. The results from this current IPD meta-analysis benefit from analysing glucose as a continuous variable, having accounted for skewness in the distribution via $\log_{10}$ transformation, thus providing greater analytical power (Chinn 2000; Streiner 2002). Investigation of confounding variables in the largest IPD study (Mazya et al. 2012) identified four factors (age, baseline NIHSS score, hypertension and stroke onset to time of treatment) aside from admission glucose levels that were associated with SICH. All of these factors have previously been identified in previous risk scores; nonetheless, adjustment of the model to include these factors did not substantially alter the absolute risk of SICH associated with an admission glucose level. The predictive capacity of the multivariate scores above has been moderate at best (Mazya et al. 2013), and this current study supports previous findings

that a mere association between risk factors and SICH does not assure their ability to predict risk of SICH in individual patients (Whiteley et al. 2012). Thus despite being associated with SICH outcome, it seems admission glucose level is of modest value for predicting the absolute risk of SICH after thrombolysis.

Limitations of the Study

As with any meta-analysis, there are a number of limitations worth consideration. It is difficult to avoid publication bias when conducting a meta-analysis. This was accounted for in part by manually searching abstracts and letters and performing Egger regression asymmetry tests. However, it has been noted that when considering fewer than ten studies, test power is usually too low to distinguish chance from real asymmetry (Sterne et al. 2011), therefore the lack of statistical significance for publication bias found in this study must be interpreted with caution.

The scarcity of statistically significant biomarkers may in part be due to the fact that data could not be extracted from a number of studies, and many candidate biomarkers were only reported by a single study, therefore a meta-analysis could not be performed. However, biomarkers reported by more than one paper were either not significant upon analysis, or the studies were too statistically heterogeneous to be able to draw any firm conclusions. This underscores the need for studies examining candidate biomarkers to adhere to strict protocols and data reporting, for example by following the REMARK guidelines (McShane et al. 2005).

Heterogeneity in study design and data presentation inhibited the inclusion of many studies into quantitative analysis. Papers that reported ORs without detailing the mean/ median and variability were excluded, as were papers reporting only ORs associated with a threshold value for a candidate biomarker. Meta-analyses were therefore conducted collectively to maintain high statistical power, rather than splitting studies based upon their sample of measurement (plasma/serum/whole blood); iterative analysis found that inclusion/exclusion of certain sample types did not make significant differences to the results achieved. It is likely that the timing of a blood sample affects the concentrations of the biomarkers measured but this could not be explored owing to the paucity of data. Despite the presence of heterogeneity, a number of measures have been taken to limit its impact e.g. the use of the random-effects model and subgroup analysis.

Examination of the relationship between admission glucose levels and SICH using IPD has overcome many of the challenges of heterogeneity associated with literature based meta-analyses. The relative advantages of analysing IPD over published data are summarised in Table 2.9. Importantly, the differing results obtained with literature based meta-analysis and IPD

meta-analysis demonstrate that the deductions of any meta-analysis must be considered within the context of its inherent limitations.

Table 2.9: The relative merits of individual patient data meta-analysis over published literature-based meta-analysis.

Published Data	Individual Patient Data (IPD)
Data quality unclear and limiting	Ensure data quality and standardise between studies
Sub-group analysis not always possible	Accurate sub-group analysis can be performed
Multivariate analysis not always possible	Option of performing multivariate analysis
Outcome data not always available	Flexible analysis of outcomes
Analysis options limited	Ensure appropriateness of analysis

Conclusions from the IPD analysis are based upon data obtained from observational studies and not RCTs, which could be considered a limitation. However, several sources have queried the value of RCTs over other study designs (Sanson-Fisher et al. 2007; Rothwell 2005) or commented upon the similarities between data obtained from RCTs and observational studies (Golder et al. 2011). The sizeably larger subject population in the collection of observational studies as compared to clinical trials investigating stroke thrombolysis adds statistical power to the analysis conducted in this IPD meta-analysis. Thus the use of data from observational studies in our meta-analysis has allowed robust examination of the relationship between admission glucose levels and the risk of SICH, to reach conclusions that are likely to be more applicable to an uncontrolled clinical environment.

Concluding Remarks

This study has found that despite a plethora of candidate biomarkers being investigated for the diagnosis or prognosis of ischaemic stroke, most of these markers are evaluated in too few studies to infer any true clinical relevance. While meta-analysis found candidate biomarkers of statistical significance likely to be indicative of the pathophysiology of stroke, these may have only marginal translational value (as demonstrated upon examination of CRP and glucose) and prospective studies will be required to further evaluate whether they stand up to scrutiny. Thus it can be concluded that at present, no diagnostic or prognostic biomarkers for acute ischaemic stroke can be recommended for routine clinical practice. This supports the pursuit of novel biomarkers for ischaemic stroke using non-bias, non-targeted approaches that are not constrained by the current understanding of stroke pathophysiology.

Systematic review of the literature also highlighted many limitations of study design that are likely to explain why many reported impressive biomarker associations could not be replicated upon meta-analysis. Effective assessment of a candidate biomarker requires a well-powered study with multiple levels of replication and validation. The study must also be performed in clinically relevant disease and control populations. These findings helped to inform and develop the design of biomarker discovery studies conducted in Chapter 3 and Chapter 4 of this thesis.

CHAPTER THREE:

A PROTEOMICS APPROACH TO DIAGNOSTIC BIOMARKER DISCOVERY FOR ACUTE ISCHAEMIC STROKE

3.1 Introduction

In Chapter 2, systematic literature review revealed a number of studies advocating biomarkers for ischaemic stroke, the vast majority of which were examined through targeted, candidate-based investigations. The data obtained from these reports provides evidence for the possibility that molecules released into the bloodstream within the acute phase of stroke could form the basis for sensitive diagnostic tests. Nonetheless, the findings also highlighted the current absence of clinically valuable biomarkers.

The challenges of AIS diagnosis have been previously discussed in Chapter 1, with CT imaging being insensitive to acute lesions and MRI imaging being expensive and unavailable in many hospitals. Thus the search for novel diagnostic biomarkers for acute ischaemic stroke has received growing attention (Whiteley et al. 2008).

With blood being readily accessible for clinical evaluation, analysis of human plasma is a promising approach to biomarker discovery. Studying the proteome through high-throughput plasma proteomic pattern analysis can directly inform on energy changes and molecular mechanisms underpinning acute insults. Therefore proteins are useful not only for understanding the pathogenesis of complex diseases such as AIS, but also have the potential to be used as accurate diagnostic biomarkers.

Early investigations have established SELDI-TOF MS protein profiling as a valuable platform for biomarker discovery in ovarian and breast cancer, with studies reporting biomarkers with over 90% sensitivity and specificity (Petricoin et al. 2002; Li et al. 2002). In vascular disease, SELDI-TOF MS has been used to study acute coronary syndrome (Florian-Kujawski et al. 2004) and carotid atherosclerosis (Degraba et al. 2011). Studies have also employed SELDI-TOF MS for the detection of diagnostic biomarkers for ischaemic and haemorrhagic stroke (Allard et al. 2004; Zhang et al. 2008; Staals et al. 2008), providing further evidence for the utility of this experimental approach (Table 1.3, Chapter 1).

Aims of the Study

The aim of the following study was to make progress towards developing a blood-based diagnostic test for AIS by:

- Comparing plasma protein profiles from patients with AIS against non-AIS individuals.
- Assessing the merits of both single marker and multi-marker models.
- Identifying candidate markers with discriminating potential.

3.2 Methods

3.2.1 Materials

Proteinchip CM10 arrays, all-in-one peptide standard and all-in-one protein-standard were obtained from Bio-Rad Laboratories Inc. (Hercules, CA, USA). Sinapinic acid (SPA) and alpha-cyano hydroxy cinnamic acid (CHCA) (>99% purity for MALDI-MS) were obtained from Fluka/Sigma-Aldrich Company Ltd. (Dorset, UK). High-performance liquid chromatography (HPLC) grade acetonitrile (ACN) was obtained from VWR International Ltd. (Lutterworth, UK). Trifluoroacetic acid (TFA) was from Rathburn Chemicals Ltd. (Walkersburn, Scotland).

3.2.2 Patient Selection and Sample Acquisition

3.2.2.1 Pilot Study

For the pilot study, blood samples were collected after written consent from recruited patients, under guidelines approved by the Hammersmith and Queen Charlotte's and Chelsea Hospitals Research Ethics Committee. Thirteen patients diagnosed with AIS were recruited for the study, having had their lesions confirmed by MRI on recruitment and 10 days later. Blood was collected from these patients within 24 hours of symptom onset and then again in a follow-up session 2–4 months after the initial stroke. These follow-up samples served as controls for the study, with the assumption that the patients were healthy and no longer had any acute stroke symptoms. Furthermore, two patients diagnosed with TIA and two stroke mimics were included as comparative subjects in the study.

Blood samples were collected in ethylene di-amino tetra-acetic acid (EDTA) tubes and centrifuged at 1300g for 10 minutes, within 30 minutes of collection. The plasma samples obtained from centrifugation were aliquoted into 200µl volumes and immediately frozen at -80^{o}C until required. All samples were collected and processed under a stringent protocol, developed through consideration of Human Proteome Organisation (HUPO) Plasma Proteome Project recommendations (Rai et al. 2005). This included drawing blood from the median cubital vein only, and leaving sufficient time between sterilisation and blood withdrawal to allow complete evaporation of the sterilising alcohol. Blood was drawn until one whole EDTA tube was filled, and then the tube was inverted ten times to ensure effective mixing of the blood with the anticoagulant EDTA.

3.2.2.2 Replication: Discovery Phase Study

For the discovery phase study, samples collected under similar conditions as detailed above were obtained from the University of Edinburgh under a Materials Transfer Agreement. This cohort consisted of acute plasma samples from 404 subjects: 235 patients with CT or MRI-confirmed AIS, 15 with haemorrhagic stroke (13 ICH, 2 Subarachnoid Haemorrhage), 72 with TIA and 82 stroke mimics.

The cohort was divided into three sets to allow independent replication of experimental conditions and candidate markers:
Training Set: 20 AIS + 20 Non AIS (10 TIA, 10 Stroke Mimics)
Test Set: 32 AIS + 32 Non AIS (16 TIA, 16 Stroke Mimics)
Validation Set: 300 subjects (183 AIS, 46 TIA, 13 ICH, 2 SAH, 56 Stroke Mimics)

3.2.3 SELDI-TOF MS Proteomic Profiling

3.2.3.1 Principles of SELDI-TOF MS

SELDI-TOF MS is a variation of MALDI-TOF MS, coupling selective protein binding using ProteinChip technology with quantitative mass detection. Two major types of ProteinChip surfaces exist: chemical (e.g. anionic, cationic, hydrophilic, hydrophobic, immobilised metal affinity metal) and biological (e.g. antibody, receptor, DNA, enzyme). This spectrum of different arrays allows a broad range of the proteome to be analysed: the variety of ProteinChip surface chemistries selectively capture specific proteins from complex biological samples while simultaneously removing detergents and salts. Hence information about a protein's chemical properties can be determined, as proteins bind to the array surface through specific chemical interactions.

The MS element of the SELDI-TOF MS technique is based upon the principles of matrix-assisted laser desorption/ionisation (MALDI)-TOF technology. In SELDI-TOF MS, some peptides and proteins are selectively adsorbed onto chemically modified surfaces while others are washed away. An energy-absorbing matrix solution crystallises with the sample, responsible for converting laser energy to thermal energy to facilitate the desorption/ionisation process.

The mass spectrum is recorded by the ProteinChip Reader, a laser desorption ionisation time-of-flight MS instrument. When the matrix on the surface of the ProteinChip absorbs energy at the wavelength of the pulsed ultraviolet nitrogen laser of the ProteinChip Reader, proteins within the sample/matrix mixture desorb and are ionised in the gas phase. These ions are

accelerated down a flight tube through an electric potential and a detector at the end of the tube measures the time of flight. The mass-to-charge ratio of each ion can be determined from the length of the tube, the kinetic energy given to ions by the electric field, and the time taken to travel the length of the tube (Figure 3.1).

The analyser used for SELDI is specially designed to achieve high-sensitivity quantification. The ion source and detector are adapted for efficient ion transmission and detection over a wide mass range, with the system being generally more favourable for proteins and peptides with molecular weight lower than 20kDa (Wiesner 2004).

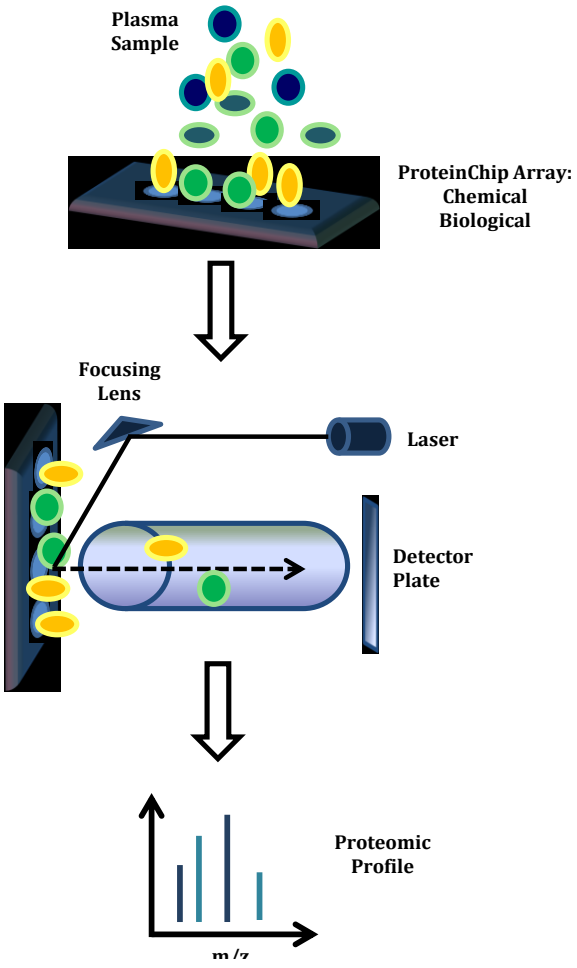


Figure 3.1: Illustration of SELDI-TOF MS technology. Upon application of a plasma sample onto a ProteinChip array, a subset of plasma proteins is retained on the chip surface, dependent on the interacting physico-chemical properties. These proteins are co-crystallised with a matrix solution. Inside the mass spectrometer, a laser induces desorption & ionisation of protein ions, which are accelerated down a flight tube through an electric potential based upon their mass to charge (m/z) ratio (normally z=1). The detector converts the arrival time of the protein ions into a spectral profile depicting the quantity of detected protein ions versus its m/z ratio.

3.2.3.2 Sample Pre-Processing

Experiment design and protocol were created according to recommendations from Bio-Rad Laboratories Inc. (2008). Samples were applied to a weak cation exchange (CM10) chip. Thawed plasma samples were denatured by the addition of a solution of 9M urea, 2% CHAPS, 5mM DTT, and were incubated for 1 hour on a vortex mixer. This was followed by a 1:10 or 1:20 plasma dilution in binding buffers of pH3, pH4, pH6 (0.1M sodium acetate plus 1M acetic acid to adjust to final pH) or pH8 (0.1M trizma base, plus 1M acetic acid to adjust to final pH). The interactive relationship between the CM10 chip, binding buffer and plasma proteins are illustrated in Figure 3.2.

3.2.3.3 Target Chip Preparation and Sample Incubation

CM10 chips were assembled in a bioprocessor consisting of a plastic cassette and aluminium frame, which together allow isolation of each individual spot on the chip for application of different samples and buffers. The interactive spots of the chips were equilibrated twice with 200µl of binding buffer within the bioprocessor assembly (Ciphergen Biosystems Inc., Freemont, USA) for 5 minutes, followed by incubation of 150µl of the diluted plasma samples for one hour with orbital shaking at 1050rpm (Heidolph Titramax 100 shaker). Solutions in the wells were discarded and the chips were washed three times with binding buffer for 5 minutes per wash, followed by one brief wash with deionised water to remove excess salts. Each washing step was performed with orbital shaking. The chips were then removed from the bioprocessor and allowed to air dry. Saturated solutions of SPA and CHCA energy absorbing matrix were prepared according to the recommendations of the manufacturer (in solution of 50% ACN and 0.5% TFA). The matrix was added twice to each spot (either SPA: 0.5µl x 2 or CHCA: 0.8µl x 2) and the spots were allowed to air dry.

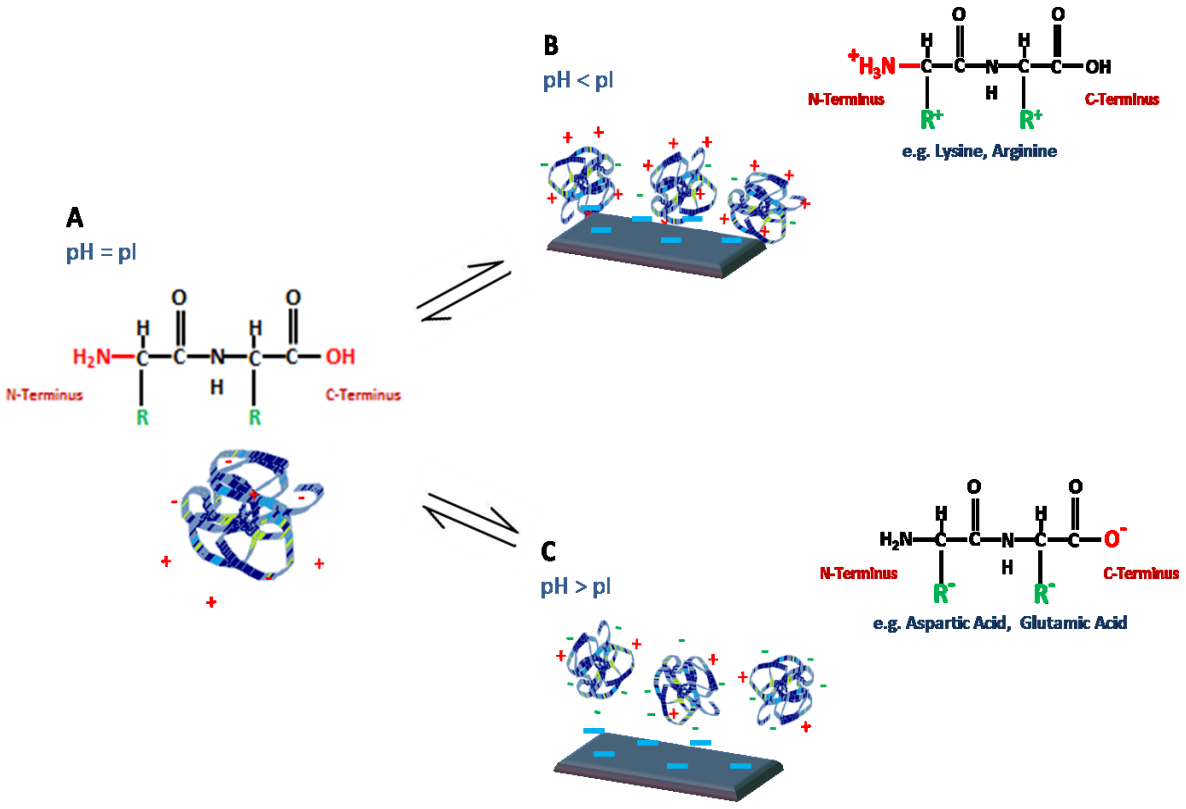


Figure 3.2: **The dynamic interaction between the CM10 ProteinChip array surface, experimental buffers and plasma proteins.** The CM10 ProteinChip array incorporates carboxylate chemistry (negative charge), thus operating as a weak cationic exchanger whose selectivity can be varied depending on buffer pH. The net charge of a protein can be manipulated by buffer pH and depends upon the side chain ('R'), which is unique to each amino acid, as well as the amino (H_2N) and carboxyl (COOH) groups of each amino acid. The positively charged amino groups of lysine and arginine amino acids and the negatively charged carboxyl groups of aspartic acid and glutamic acid amino acids primarily determine the electrochemical character of a protein. **A)** When the buffer pH matches a protein's isolectric point (pI), the protein's amino acids have an overall neutral charge. Most plasma proteins have a pI in the range of pH5-7. **B)** Decreasing the buffer pH (more acidic) neutralises acidic amino acids and imparts an overall net positive charge on plasma proteins due to amino groups and basic amino acids, such as lysine and arginine carrying a positive charge. Therefore a greater number of proteins bind to the CM10 surface. **C)** Conversely, increasing pH (more basic) neutralises basic amino acids and imparts an overall net negative charge on amino groups and acidic amino acids such as aspartic acid and glutamic acid within the proteins, resulting in less binding. Protein binding specificity can also be modified by varying the ionic strength of the buffer; increasing ionic strength increases competition between protein ions and buffer ions, causing weakly bound proteins to be eluted from the CM10 surface. Low stringency buffers (e.g. sodium acetate, trizma base) were used throughout to maximise protein yield.

3.2.3.4 Data Acquisition

The CM10 chips were analysed using a Protein Biology System IIc Reader (Ciphergen Biosystems Inc., Freemont, USA) which was calibrated for mass accuracy using 'All in One' peptide and protein standards. During the discovery phase study, a pooled 'quality control' sample from the pilot study was included on eight spots during each experiment to ensure reproducibility.

The TOF-MS spectra were generated by averaging 117 pulsed UV laser shots with an intensity of 200-210 units when SPA was used and 160-180 units when CHCA was used. Two warming shots were fired at a laser intensity of 5 units greater than acquisition and the data from these shots was discarded. Peak detection was performed using Biomarker Wizard software 3.1 software (Ciphergen Biosystems Inc., Freemont, USA). All peaks from m/z 2000 to 30,000Da with a signal to noise ratio above 5 at first pass and 2 at second pass were auto-detected and the peaks were aligned within 0.3% of mass error. Molecular masses below 2000Da were not analysed due to masking of peptide signals by peaks from the matrix. All peaks detected were required to be present in at least 5% of the spectra, in an attempt to avoid inclusion of peaks with poor discriminatory power. Automatic baseline correction was used and data filtering was set as default at 0.2 times expected peak width.

3.2.3.5 SELDI-TOF MS Protocol Optimisation

The SELDI-TOF MS protocol was optimised in a number of ways to ensure high quality data acquisition and processing. Plasma samples were diluted 1:100 and 1:200 in buffer to ascertain which dilution gave a better protein profile, without saturating the detector. Matrix application was standardised by using different volumes (but same concentration) of matrix solution to ensure the solution covered the entirety of each spot, and drying times between matrix applications were also controlled. The data acquisition protocol was optimised on the Ciphergen software by adjusting the following parameters: detector sensitivity, mass range, focus mass, matrix attenuation, warming shot energy and data shot energy. The number of peaks detected, dependent on settings used, ranged from 37 to 106.

In order to investigate data quality, the total ion current (TIC) from m/z 2000 to 20,000 was measured and compared between spectra of different samples. This normalisation process calculates an average TIC from all spectra over a mass range appropriate to the matrix used (2–10kDa for CHCA and 10– 50kDa for SPA) and adjusts the intensity scales for all the spectra to this average.

3.2.4 Data Interpretation and Analysis

Protein ions detected by the Biomarker Wizard software (and distinguished as 'peaks') were exported into Microsoft Excel as intensity values. Means and the intra-assay coefficient of variation (CV) values for each peak were calculated from replicates, and differences between mean peak intensities were compared between sample groups. Mann-Whitney test (unpaired) or Wilcoxon Signed Rank test (paired), were used to examine statistical significance. Proteins showing significant differences between analysis groups ($p<0.05$) were selected for further analysis.

Data were subjected to unsupervised analysis by Principal Component Analysis (PCA), and visualised with scatter plots and volcano plots using GraphPad Prism v4 and Statistica v6.1 (Statsoft Ltd) software.

The relative merits of a single marker and multi-marker diagnostic test for cerebral ischaemia were assessed. Discriminant function analysis, utilising a forward step-wise regression model, was performed to determine variables that were able to discriminate between the clinical groups. With this type of model, at each step all variables are assessed to evaluate which one contributed most to the discrimination between groups; this variable is incorporated into the model and the process starts again, in a forward step-wise manner until a robust model of the most discriminating factors is built. The process is guided by *F* to enter and *F* to remove values, with the *F* value of a variable (the ratio of the intergroup variance over the average intragroup variance) determining whether or not that variable has the ability to discriminate between the groups. All protein ions as well as continuous phenotypic variables (age, admission NIHSS, time of blood sampling) were used as variables, with an *F* to enter value set at 4 and *F* to remove set at 3.9.

Protein ions identified through discriminant function analysis (DFA) were used to build a multiple regression model. Iterative reverse step-wise analysis was performed to generate classification matrices. Receiver Operating Characteristic (ROC) curves were generated from the regression model predicted values to calculate sensitivity and specificity and further assess the value of these ions independently or combined in a multi-marker model. The DeLong, Delong and Clarke-Pearson method (DeLong et al. 1988) was used for calculating the statistical difference between ROC curve areas. Standardised Mean Difference (SMD) and odds ratio (OR) were calculated using Comprehensive Meta-Analysis v2.0 software, as detailed in Chapter 2.

3.2.5 Identification and Quantification of Candidate Biomarker

3.2.5.1 Immunodepletion of m/z6640

Immunodepletion of the candidate biomarker was performed using two anti-human apolipoproteinC-1 (ApoC-1) antibodies, one mouse monoclonal (Abnova, Taipei City, Taiwan) and one rabbit polyclonal (Abcam, Cambridge, UK). A pooled plasma sample (previously used as a quality control sample) was incubated with anti-ApoC-1 antibody in 1:200 dilution in pH4 buffer, with or without the addition of 9M urea, 2% CHAPS, on rotation for one hour. An antibody concentration curve was performed, from 0.18µg/ml to 18µg/ml. Control incubations using anti-ApoC-1 antibody only, plasma only, and affinity-purified rabbit IgG control (courtesy of Dr Robert Edwards) were used to confirm specificity. All samples were applied to CM10 ProteinChips then processed by SELDI-TOF MS and analysed as detailed above.

3.2.5.2 ApolipoproteinC-1 ELISA

A total of 40 plasma samples were chosen from both Training and Test Sets to cover the range of intensities of m/z6640 detected by SELDI-TOF MS. Duplicate samples were measured by enzyme-linked immunosorbent assay (ELISA) kit (AssayPro, St Charles, MO, USA) as per manufacturer's instructions. The assay employs a quantitative sandwich enzyme immunoassay technique, with immobilised polyclonal anti-ApoC-1 antibody on a 96-well microplate. Plasma samples were diluted 1:100 in the provided diluent, then 50µl of samples and standards were added to each well of the microplate and incubated at room temperature for 2 hours. Following five washes with 200µl of wash buffer, 50µl of biotinylated polyclonal ApoC-1 antibody was added to each well and incubated for 2 hours. The plate was washed again, as above, then 50µl of streptavidin-peroxidase conjugate was added to each well and incubated for 30 minutes. Following another washing step, 50µl of chromogen substrate was added to each well and incubated for 30 minutes, after which 50µl of stop solution was added. Colour intensity was measured at 450nm on a microplate reader within 10 minutes.

3.3 Results

3.3.1 Pilot Study

SELDI-TOF MS employs a variety of ProteinChip arrays and conditions to allow biomarker discovery. A pilot study was performed to optimise a number of variables for biomarker elucidation. The study consisted of 13 AIS patients (11 of which had paired acute and follow-up blood samples) as well as acute blood samples taken from 2 TIA patients and 2 stroke mimics (Table 3.1).

3.3.1.1 Evaluation of Data Quality

Establishing good reproducibility is critical before determining whether blood protein profiles differ in various clinical states. It is expected that the median intra-assay CV should be between 10-20% (Bio-Rad Laboratories Inc. 2008). Preliminary experiments demonstrated a very high CV (>50%), therefore adjustments were made to the protocol to improve data quality. Extrinsic variables including instrument performance, buffer composition, sample preparation and acquisition settings were altered. However, the greatest improvement in CV was made by increasing plasma sample dilution from 1:100 to 1:200 to minimise detector saturation, and refining matrix preparation. A sample dilution of 1:200 combined with freshly made, saturated matrix, applied with careful pipetting to each interactive spot, decreased CV to a mean of 18%, similar to previously reported values (Schaub et al. 2004). Intra-assay CV values, pre- and post-optimisation of the experimental protocol, are shown for 6 protein ions of interest in Table 3.2.

3.3.1.2 Investigating Matrix and pH Binding Conditions for Biomarker Discovery

Three ProteinChip arrays are recommended for analysis of plasma proteins (Bio-Rad Laboratories Inc. 2008): the CM10 cation exchange, H50 reverse-phase and the IMAC30 immobilised metal affinity chips. Previous work within the lab with plasma samples has found that CM10 chips produce the greatest number of detectable protein ions and also yield the highest number of significant differences between study groups (Dr Waheedah Abdul-Salam, personal communication). Therefore the CM10 chip was chosen for investigation.

The CM10 ProteinChip array provides flexibility for biomarker discovery by allowing different subsets of proteins to bind to its surface by using a variety of pH buffers. In a preliminary study of 11 paired AIS samples (acute and follow-up) a range of pHs were used, from less stringent (pH3, pH4) to more stringent (pH6, pH8). Two different matrices were also used: CHCA (better for measuring proteins <20kDa) and SPA (for proteins >20kDa). Figure 3.3 depicts

representative spectra obtained from AIS plasma samples, using CM10 chips and SPA matrix at various pH binding conditions.

The majority of proteins detected lay in the 2-20kDa molecular weight range. Nonetheless, it was found that SPA matrix yielded more detectable protein ions than CHCA, as well as more significantly different ($p<0.05$) ions (Figure 3.4A). The proportion of significantly differentially expressed ions was greater using SPA rather than CHCA matrix, implying that more of these ions detected with SPA were true positive associations rather than being false positives (type 1 error rate of 5%, at a threshold of $p<0.05$). Therefore this matrix was chosen for subsequent investigation.

Table 3.1: Pilot study patient characteristics. Demographic and clinical information relating to the 13 AIS patients (from which paired acute and follow-up plasma samples were taken from 11 subjects), as well as 2 patients with TIA and 2 stroke mimics. Values are median and interquartile ranges (IQR), or range when n=2.

	Total Ischaemic Population (n=15)	**Large AIS (>10 cm^3 lesion volume) (n=4)**	**Moderate AIS (>3 cm^3 lesion volume) (n=3)**	**Mild AIS (<3 cm^3 lesion volume) (n=6)**	**TIA (no lesion on MRI) (n=2)**	**Stroke Mimics (no lesion on MRI) (n=2)**
Age	70.00 (62.00-81.00)	63.00 (54.50-78.25)	82.00 (70.00-91.50)	63.00 (47.50-74.50)	73.00 (65-81)	74.50 (70.00-79.00)
Gender	5F, 10M	1F, 3M	2F, 1M	2F, 4M	0F, 2M	0F, 2M
Admission NIHSS	4.00 (2.50-6.00)	4.00 (2.50-10.00)	8.00 (4.00-12.00)	3.50 (1.75-4.00)	2.00 (1-3)	3.50 (2-5)
Time of Acute Blood Sampling (hours)	14.00 (7.50-22.00)	13.00 (4.38-20.50)	17.00 (13.00-21.00)	7.75 (4.38-21.00)	6.50 (2.00-11.00)	20.50 (18.00-23.00)
Lesion Volume (cm^3)	3.30 (0.75-13.40)	15.50 (12.65-16.70)	3.80 (3.30-5.60)	0.75 (0.43-1.15)	n/a	n/a

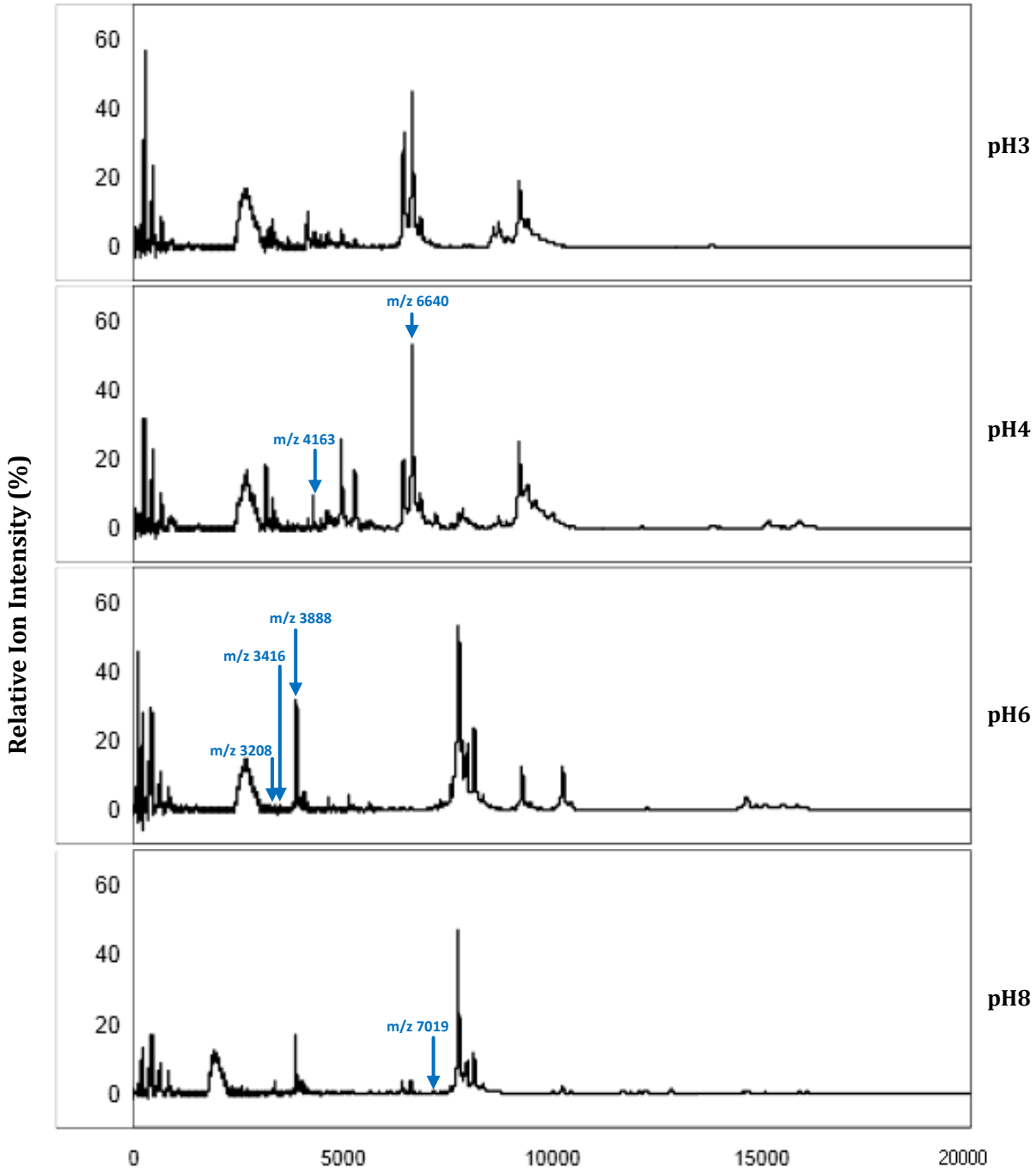
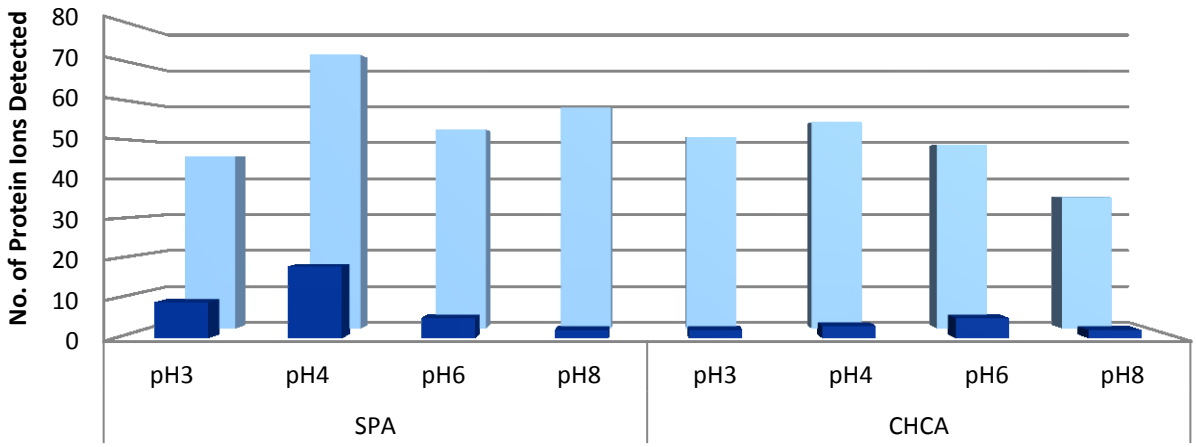


Figure 3.3: Representative SELDI-TOF MS spectra from the pilot study using CM10 ProteinChip arrays with SPA matrix. The pilot study consisted of 13 AIS subjects (11 from which paired acute and follow-up plasma samples were taken, the other two from which only acute plasma samples were available), 2 TIA subjects and 2 stroke mimics. The spectra show profiles of one acute AIS patient plasma sample, analysed on CM10 ProteinChip arrays with 0.1 M sodium acetate and 0.1% triton X-100 as the binding buffer at pH's 3, 4, 6 and 8. Some protein ions appeared under many conditions, whereas others were unique to a single pH. Protein ions considered of interest (see Table 3.2) are labelled.

A



B

	pH3 SPA	pH4 SPA	pH6 SPA	pH8 SPA
Total Number of Peaks Detected:				
	46	**73**	**53**	**59**
Number of Peaks Significantly Different (% of Total Number of Peaks Detected):				
Acute AIS (n=11) vs Follow-Up AIS (n=11)	**9 (19.6%)** m/z 3901, 4163, 4316, 4643, 6498, 6850, 7765, 7927, 9243	**18 (24.7%)** m/z 3093, 3224, 3331, 3901, 3925, 4142, 4163, 4196, 4286, 4803, 5397, 6440, 6498, 6640, 6694, 6765, 6850, 8706	**5 (9.4%)** m/z 3136, 3579, 3888, 7967, 8143	**2 (3.4%)** m/z 3369, 22665
AIS+TIA (n=15) vs Stroke Mimics (n=2)	**7 (15.2%)** m/z 5199, 5237, 6640, 9243, 9510, 15873, 18305	**8 (11.0%)** m/z 2719, 3013, 4142, 15135, 15910, 16104	**11 (20.8%)** m/z 3208, 3312, 3416, 3579, 3888, 6833, 11662, 12552, 12836, 15872, 16047	**4 (6.8%)** m/z 7019, 7765, 10467, 14852

Figure 3.4: The effect of matrix choice and pH on plasma protein profiles in the pilot study. A) Total number of protein ions detected using SPA and CHCA matrix at various pHs (light blue bars) and number of significantly different ions (Wilcoxon Signed Rank test, $p<0.05$) between paired acute and follow-up plasma samples from 11 AIS patients (dark blue bars). **B)** A number of protein ions were significantly differentially expressed between acute and follow-up paired AIS samples and between AIS and TIA (n=15) and stroke mimics (n=2) using SPA matrix and different pH binding conditions, as calculated by Mann-Whitney or Wilcoxon Signed Rank test . In all but pH8, the proportion (%) of significantly different protein ions found, from the total number of peaks detected, is greater than 5%, indicating that the pH3, pH4 and pH6 binding conditions yield more significantly different protein ions than would be expected by chance alone

3.3.1.3 Preliminary Diagnostic Biomarkers

Using SPA matrix, each pH binding condition produced different sets of protein ions that varied significantly in ion intensity between the various study groups (Mann-Whitney or Wilcoxon Signed Rank test, $p<0.05$), as shown in Figure 3.4B.

Of the 63 significantly differentially expressed protein ions detected with SPA matrix at various pHs, six (9.5%) were considered as potential putative biomarkers for AIS (Table 3.2), based upon degree of statistical significance, fold difference and visual inspection of the ability of the protein ions to distinguish between study groups. Box and whisker plots for these nine protein ions are presented in Appendix 2. Two protein ions (m/z 4163, 3888) could differentiate between acute and follow-up AIS samples and four protein ions (m/z 6640, 3208, 3416, 7019) between AIS and TIA against stroke mimics. Thus the pilot study confirmed the hypothesis that SELDI-TOF MS is able to detect a protein profile that is significantly affected during acute ischaemic cerebral event.

Table 3.2: Preliminary diagnostic biomarkers identified from the pilot study. Six protein ions from the pilot study were selected as 'putative markers' from the pilot study. Protein ions were captured using binding conditions of pH3, pH4, pH6 and pH8, on CM10 ProteinArray chips with SPA matrix. The putative markers were selected based upon visual inspection of their ability to distinguish between study groups, as well as degree of significance (Wilcoxon Signed Rank test p-value) and fold difference between study groups. Intra-assay coefficient of variation (CV) values for each peak are shown pre- (in brackets) and post- optimisation of the experimental protocol. Associated box and whisker plots for the nine protein ions are presented in Appendix 2.

	CM10 Binding Condition	Protein Ion m/z	Relative Intensity of Peaks (mean ± sd)	Fold Difference	p-value	Intra-Assay CV %
Acute (n=11) vs Follow-Up (n=11) AIS	pH4	4163	Acute = 4.9 ± 1.8 Follow-Up = 2.2 ± 1.7	2.23	0.001	15.8 (63.2)
	pH6	3888	Acute = 42.6 ± 20.4 Follow-Up = 11.2 ± 8.4	3.80	0.027	9.6 (62.80)
AIS + TIA (n=15) vs Stroke Mimics (n=2)	pH4	6640	AIS+TIA = 45.9 ± 10.3 Stroke Mimics = 32.8 ± 10.6	1.40	n/a	12.6 (47.9)
	pH6	3208	AIS+TIA = 21.9 ± 6.0 Stroke Mimics = 11.6 ± 2.9	1.89	n/a	14.1 (39.1)
	pH6	3416	AIS+TIA = 8.1 ± 2.4 Stroke Mimics = 3.4 ± 0.1	2.38	n/a	14.8 (68.4)
	pH8	7019	AIS+TIA = 0.8 ± 0.5 Stroke Mimics = 2.3 ± 0.3	0.35	n/a	17.3 (56.3)

3.3.2 Discovery Phase Study

Preliminary results from the pilot study provided justification for the feasibility of SELDI-TOF MS in elucidating diagnostic biomarkers for AIS. Therefore a second patient cohort was obtained for a well-powered discovery study utilising the same experimental approach, following a study design that has proven successful in the past (Abdul-Salam et al. 2010).

Samples were obtained from the University of Edinburgh (with consent from Professor Peter Sandercock and Dr. William Whiteley) and consisted of 404 well-characterised patients: 235 patients with CT or MRI-confirmed AIS, 15 with haemorrhagic stroke (13 ICH, 2 Subarachnoid Haemorrhage), 72 with TIA and 82 stroke mimics.

Power calculations were performed based upon the intensities of the nine protein ions from the pilot study deemed to be putative biomarkers. An average sample size of 10 from each arm (AIS and TIA vs non-stroke) delivered a 95% power at $p<0.01$, and a sample size of 7 from each arm delivered a 95% power at $p<0.05$ to detect a 2-fold difference in means.

With consideration of the power calculations above, the discovery cohort was randomised by a third party into three independent 'Training', 'Test' and 'Replication' sets, as suggested by Burke (2005), to ascertain the validity of any potential biomarkers. This approach serves to minimise variations arising from sample processing and possible changes in instrument performance, thus ensuring the robust confirmation of findings.

The Training Set consisted of 40 patients (20 AIS, 20 Non-AIS [10 TIA, 10 Stroke Mimic]), the Test Set of 64 patients (32 AIS, 32 Non-AIS [16 TIA, 16 Stroke Mimic]) and the Validation Set of the remaining 300 patient samples. A 'quality control' sample obtained from the pilot study was run alongside Training and Test samples to ascertain suitable replication of data. Patient characteristics for the Training Set cohort are given in Table 3.3.

Table 3.3: Discovery Study Training Set patient characteristics. Demographic and clinical information relating to 20 AIS and 20 Non-AIS (10 TIA, 10 Stroke Mimics) patient plasma samples comprising the Training Set. Values are median and IQR.

	Total Population (n=40)	AIS (n=20)	TIA Positive (NIHSS<5) (n=10)	AIS and TIA Positive combined (n=30)	Stroke Mimics (NIHSS<5) (n=10)	AIS&TIA vs Stroke Mimics (p-value)
Age	75.00 (59.50-84.75)	81.50 (72.75-85.00)	71.00 (60.50-76.50)	78.00 (69.00-85.00)	57.00 (46.75-76.75)	0.021
Gender	24F, 16M	14F, 6M	7F, 3M	21F, 9M	3F, 7M	0.025
Confirmation of Stroke	20 CT, 17 MRI	12 CT, 8 MRI	8 CT, 2 MRI	20 CT, 10 MRI	3 CT, 7 MRI	0.087
OCSP Classification	4 LACS 9 PACS 3 POCS 10 TACS, 4 Uncertain 10 Non-Stroke	2 LACS 7 PACS 10 TACS 1 Uncertain	1 PACS 2 POCS 7 Uncertain	2 LACS 8 PACS 2 POCS 10 TACS 8 Uncertain	10 Non-Stroke	0.065
Time of Blood Sampling (hours)	6.62 (3.08-17.17)	3.79 (2.50-16.38)	6.38 (5.13-8.25)	5.63 (2.92-14.19)	14.67 (6.17-25.50)	0.039
Previous MI/Angina	6Y, 34N	5Y, 15N	0Y, 10N	5Y, 25N	1Y, 9N	0.609
Previous Stroke/TIA	10Y, 30N	5Y, 15N	4Y, 6N	9Y, 21N	1Y, 9N	0.206
Admission NIHSS	5.50 (1.00-16.75)	16.50 (13.25-21.00)	0.50 (0.00-1.00)	13.50 (1.00-18.25)	1.00 (0.5-3.00)	0.018

3.3.2.1 Training Set: Establishing Optimum Experimental Conditions

Experimental conditions were re-established and optimised in the Training sample set. Following the findings from the pilot study, CM10 Protein Chips were used along with SPA matrix and binding buffers of pH3, pH4, pH6 and pH8.

Preliminary Data Analysis

As in the pilot study, the number of protein ions detected and number of significantly different ions ($p<0.05$) between various study groups was examined, alongside visual inspection of the distribution of protein ion intensities. Of all the experimental conditions, pH4 provided the greatest number of significantly differentially expressed protein ions between ischaemic (AIS and TIA) and non-ischaemic (stroke mimic) subtypes (Table 3.4).

Visualisation of Data

In order to gain a better understanding of the value of SELDI-TOF MS for biomarker discovery for AIS, complete data sets obtained from each pH condition were examined by unsupervised analysis using Principal Components Analysis (PCA).

Dimensionality reduction decreases the number of redundant peaks or noise that might otherwise confuse the final analysis. PCA has become a standard tool in large dataset analysis to study causes of variation, and can be used to infer continuous axes of variation (eigenvectors) that reduce the data to a small number of dimensions, whilst describing as much of the variability between individuals as possible. PCA analysis of protein ions obtained from all pH conditions showed that patient diagnosis (AIS/TIA/Stroke Mimic) was a key factor for classification, and in particular demonstrated clustering of ischaemic events (AIS and TIA) versus non-ischaemic events (stroke mimics). Other demographic variables (e.g. age, gender, evidence of infection) did not exhibit this separation. This result indicated that differential ion expression can be used for patient diagnostic classification, unconfounded by demographic variables (Figure 3.5, Appendix 3).

Table 3.4: Differentially expressed protein ions in the Discovery Study Training Set. Peaks found to be differentially expressed between study groups at various pH binding conditions with SPA matrix (Mann-Whitney test, $p<0.05$).

	pH3	pH4	pH6	pH8
Total Number of Peaks Detected:				
	55	**66**	**61**	**68**
Number of Peaks Significantly Different:				
AIS(n=20) vs TIA(n=10) and Stroke Mimics (n=10)	**29 (52.7%)** m/z 2648, 2670, 2698, 2971, 3228, 3328, 3432, 3700, 4301, 5866, 6448, 6646, 6853, 7044, 7083, 7777, 7856, 7984, 8158, 8479, 8584, 9210, 9306, 9419, 9598, 11566, 11710, 15950, 16113	**37 (56.1%)** m/z 3228, 3333, 3382, 3432, 3452, 3550, 4366, 4482, 4639, 4822, 5866, 6451, 6652, 6855, 6963, 7043, 7082, 7228, 7581, 7784, 7853, 7970, 8156, 8589, 8720, 8905, 9601, 9748, 10023, 11557, 11702, 11962, 12143, 13795, 15157, 15341, 15916	**20 (32.8%)** m/z 3697, 3898, 3917, 4030, 5052, 5116, 5933, 6442, 6486, 6643, 6684, 6846, 7039, 7580, 7679, 10844, 13778, 15145, 15331, 23482	**38 (55.9%)** m/z 2892, 3080, 3458, 3522, 3547, 3663, 3860, 3880, 3931, 3991, 4173, 4271, 4303, 4600, 4700, 4900, 4934, 5066, 5578, 5791, 6285, 6438, 6634, 7034, 7546, 7837, 8466, 9921, 11618, 12105, 13161, 13545, 14304, 15138, 15354, 16976, 27084, 28119
AIS (n=20) +TIA (n=10) vs Stroke Mimics (n=10)	**31 (56.4%)** m/z 2648, 2670, 2698, 2971, 3164, 3228, 3328, 3381, 3432, 3453, 3700, 4610, 4973, 5281, 6448, 6646, 6853, 7044, 7083, 7777, 7856, 7984, 8158, 9139, 9210, 9306, 9419, 9598, 11566, 11710, 15950	**52 (78.8%)** m/z 2642, 2687, 2710, 2857, 2971, 3165, 3228, 3333, 3382, 3432, 3452, 3550, 4366, 4609, 4639, 4797, 4977, 5177, 5291, 5866, 6451, 6652, 6855, 6963, 7043, 7082, 7228, 7581, 7784, 7853, 7970, 8156, 8589, 8720, 8905, 9210, 9314, 9394, 9601, 9748, 10023, 11557, 11702, 11962, 12143, 13795, 15157, 15341, 15916, 18592, 18774, 28170	**24 (39.3%)** m/z 3697, 3898, 3917, 4030, 4903, 5052, 5933, 6075, 6486, 6643, 6684, 6846, 7039, 7580, 7679, 7847, 10036, 10844, 11649, 12879, 13778, 15145, 23482, 28157	**46 (67.6%)** m/z 2729, 2822, 2861, 2892, 3080, 3131, 3152, 3458, 3522, 3663, 3698, 3860, 3880, 3931, 3991, 4027, 4099, 4173, 4271, 4303, 4372, 4424, 4900, 4934, 5066, 5145, 5791, 6040, 6096, 6285, 6438, 6634, 7034, 7837, 7882, 8167, 9448, 9921, 12105, 12901, 14304, 15138, 15354, 16976, 27084, 28119
TIA (n=10) vs Stroke Mimics (n=10)	**13 (23.6%)** m/z 2971, 3164, 3228, 3328, 3381, 3432, 3453, 3700, 6448, 6646, 6853, 7044, 7083	**9 (13.6%)** m/z 2642, 2687, 2710, 2857, 3228, 3333, 3432, 3701, 6652	**5 (8.2%)** m/z 3697, 6684, 7039, 12879, 28157	**11 (16.2%)** m/z 2822, 2861, 2892, 3860, 3931, 3991, 4099, 4173, 4424, 8167, 12901

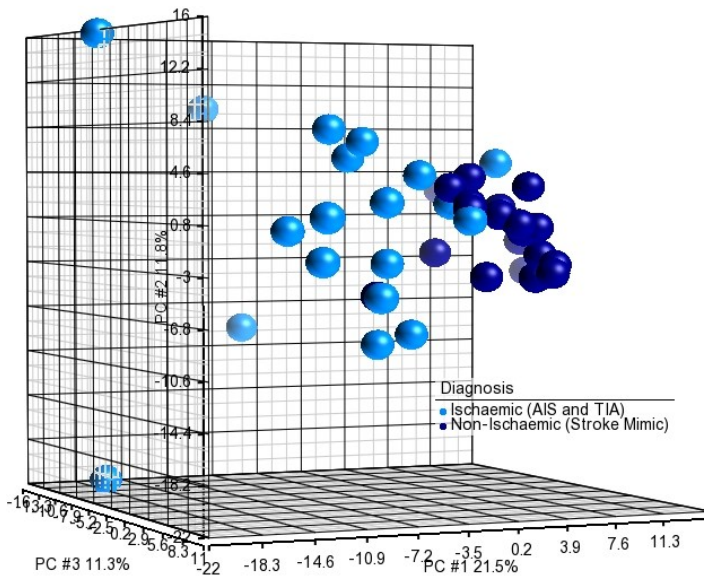
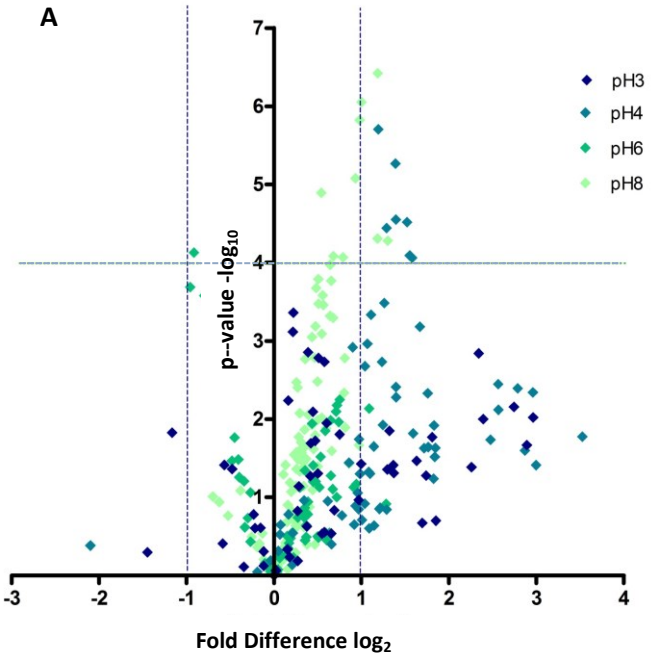


Figure 3.5: Unsupervised analysis of Training Set data by principal component analysis (PCA). Unsupervised analysis of all protein ions obtained at pH4 binding conditions from the Training Set data (comprising 20 AIS patients, 10 TIA patients and 10 stroke mimics) demonstrated a natural ability of the collective protein ions to separate patients by their diagnostic classification. PCA analysis of protein ions at pH3, pH6 and pH8 binding conditions found similar grouping of AIS, TIA and stroke mimic populations. The same analysis performed on a number of demographic variables (admission NIHSS, time of blood sampling,, history of MI, stroke or TIA and gender) did not demonstrate this ability for classification (Appendix 3), indicating that differential protein ion expression can be useful for patient diagnostic classification, independent of other confounding variables.

For all pH conditions, a volcano plot was generated to visualise protein ions along two key parameters: fold difference of ion intensity between ischaemic and non-ischaemic samples, and the corresponding Mann-Whitney p-value for each ion. A volcano plot generated from all Training Set results (Figure 3.6A) demonstrated that pH4 and pH8 were the only two binding buffer conditions that yielded protein ions exhibiting a fold difference >2 between ischaemic (AIS and TIA) versus non-ischaemic (stroke mimic) groups, and a Mann-Whitney p-value<0.0001. The substantial skew observed in the volcano plot also indicated that the majority of protein ions detected by SELDI-TOF MS are up-regulated during acute cerebral ischaemia.

To explore the relationships between the eleven protein ions found to have high fold difference and low p-value by the volcano plot (pH4: m/z 3333, 3432, 6640, 6855, 7043, 7228, 8720; pH8: m/z 3860, 4173, 6440, 6640) the protein ion intensities were examined for co-linearity through the creation of correlation matrices (Figure 3.6B). The seven protein ions captured at pH4 showed substantial correlation with each other (r^2=0.46-0.92). Conversely, of the four protein ions captured at pH8, only m/z 6440 and m/z 6640 exhibited a lower degree a correlation (r^2=0.34-0.79). This implies that the ions captured at pH8 represent three distinct proteins whose expression alters independently in response to acute ischaemic insult. This lack of relationship between proteins is essential when building an effective multi-marker model: each protein ion behaves differently and therefore cumulatively adds to the sensitivity and specificity of the model. Nonetheless, the high degree of correlation between many of the ions captured at pH4, and the high intensity of m/z 6640 (also captured at pH8), indicated the potential to find one very valuable biomarker for the differentiation between ischaemic (AIS and TIA) and non-ischaemic (stroke mimic) events. Therefore, the pH4 binding buffer was chosen for further investigation.

A



		Mean Relative Ion Intensity ±SD	
		AIS + TIA	Stroke Mimic
pH4	m/z 3331	9.07 ± 5.27	2.62±1.75
	m/z 3432	2.44 ± 1.61	0.46 ± 0.59
	m/z 6640	47.09 ± 20.91	16.64 ± 12.06
	m/z 6855	12.40 ± 7.98	3.58 ± 2.70
	m/z 7043	3.65 ± 2.02	1.25 ± 0.81
	m/z 7228	1.55 ± 1.10	0.46 ± 0.28
	m/z 8720	3.14 ± 1.81	1.12 ± 0.73
pH8	m/z 3699	2.29 ± 1.03	3.10 ± 1.21
	m/z 3860	13.47 ± 5.98	5.79 ± 1.98
	m/z 4173	13.60 ± 5.26	6.65 ± 2.89
	m/z 6440	5.20 ± 2.94	2.46 ± 1.40
	m/z 6640	7.46 ± 4.57	3.15 ± 1.42

B

pH4	m/z 3331	m/z 3432	m/z 6640	m/z 6855	m/z 7043	m/z 7228	m/z 8720
m/z 3331	1.00	0.88	0.89	0.92	0.65	0.46	0.72
m/z 3432		1.00	0.75	0.76	0.49	0.38	0.69
m/z 6640			1.00	0.92	0.74	0.62	0.76
m/z 6855				1.00	0.74	0.59	0.74
m/z 7043					1.00	0.67	0.59
m/z 7228						1.00	0.47
m/z 8720							1.00

pH8	m/z 3860	m/z 4173	m/z 6440	m/z 6640
m/z 3860	1.00	0.61	0.34	0.47
m/z 4173		1.00	0.35	0.40
m/z 6440			1.00	0.79
m/z 6640				1.00

Figure 3.6: Evaluating pH binding conditions in the Discovery Study Training Set to identify potential diagnostic biomarkers. A) A volcano plot of fold difference (log_2) against Student's t-test p-value ($-log_{10}$) was generated from all protein ions detected in the Training Set. This displayed eleven protein ions with high fold difference (>2) and significance (p-value<0.0001) between ischaemic (AIS and TIA) and non-ischaemic (stroke mimic) patient groups. **B)** Correlation matrices revealed high correlation between ions detected at pH4, in contrast to those detected at pH8. This indicates that the proteins detected at pH8 have an independent relationship to the presence of ischaemic and could constitute a multi-marker diagnostic test by cumulatively increasing the sensitivity and specificity of the test.

As no single protein ion was able to completely differentiate between ischaemic and non-ischaemic patients based upon ion intensity, a multi-marker approach was considered to select a minimum number of protein ions that allowed maximum discrimination between the two groups. Forward stepwise discriminant function analysis was performed on all 66 protein ions identified from the pH4 binding condition in the Training Set, resulting in the selection of 2 protein ions (m/z 3699 and 6640) that in combination gave high measures of sensitivity and specificity. Multiple regression analysis on these 2 protein ions again confirmed their high discriminatory power.

A receiver operating characteristic (ROC) curve combining the 2 protein ions selected by discriminant function analysis gave a very high area under the curve (AUC) based upon multiple regression predictive values (AUC=0.97), and an overall test accuracy (proportion of individuals correctly classified) of 90.5% (Figure 3.7). Furthermore, scatterplots of the 2 individual proteins demonstrate relatively good discrimination between ischaemic versus non-ischaemic groups, with m/z 6440 being upregulated in the presence of ischaemia and m/z 3699 being downregulated (Figure 3.8A). This result demonstrates the ability of discriminant function analysis and multiple regression analysis to select protein ions that are poorly correlated with one another and that could collectively contribute to a multi-marker test.

3.3.2.2 Test Set: Replication of Training Set Findings

With consideration of the findings from the Training Set, SELDI-TOF MS was performed on the 32 AIS, 16 TIA and 16 stroke mimic samples comprising the Test Set (Table 3.5), using CM10 Protein Chips with SPA matrix and pH4 binding buffer. A total of 101 protein ions were detected, of which 30 showed significant ($p<0.05$) differential expression between ischaemic (AIS and TIA) versus non-ischaemic (stroke mimic) populations.

The performance of putative biomarkers identified from forward stepwise discriminant function analysis of the Training Set was examined in the Test Set (Figure 3.8B). Logistic regression analysis found that previous history of stroke/TIA and admission NIHSS were associated with m/z 6640 relative ion intensity (Table 3.6). However, these associations were not as strong as that demonstrated by diagnostic classification (previous history of stroke/TIA: OR 1.10, 95% CI 0.24-4.18, p=0.033; admission NIHSS: OR 0.89, 95% CI -0.35-3.18, p=0.007; Diagnosis: OR 12.91, 95% CI 4.22-39.49 $p<0.0001$).

The sensitivity of m/z6640 was lower in the Test Set than in the Training Set, with a value of 83% (as compared to 94%), and specificity increased to 94%. Addition of m/z3699 resulted in a sensitivity of 85% and specificity of 95%, with an AUC of 0.93 (95% CI 0.87-0.98; $p<0.0001$).

A

	m/z 3699	m/z 6640
m/z 3699	1.00	0.03
m/z 6640	0.03	1.00

B

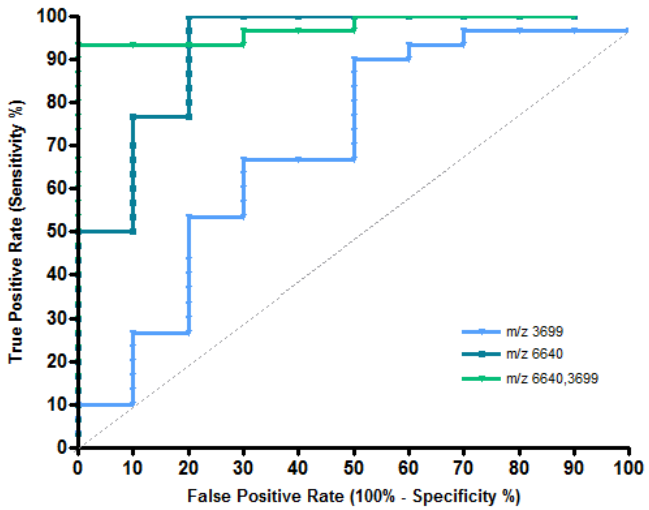
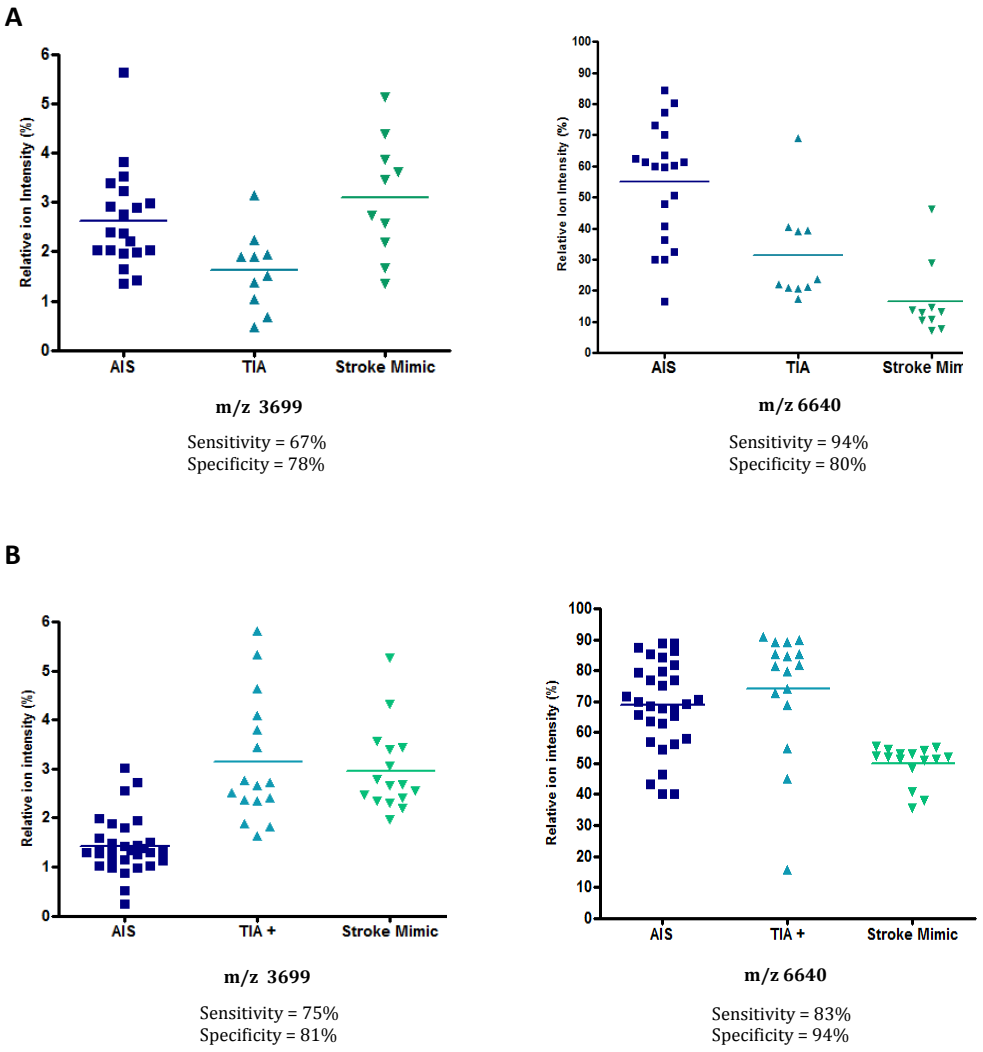


Figure 3.7: Examination of putative biomarkers identified through discriminant function analysis (DFA) of Training Set data. DFA performed on all 66 protein ions detected from the pH4 binding condition with Training Set plasma samples yielded 2 protein ions (m/z 3699 and m/z 6640) that could differentiate between ischaemic (AIS and TIA; n=30) and non-ischaemic (stroke mimics; n=10) acute cerebral events. These two protein ions demonstrated little inter-correlation, implicating their utility in a multi-marker diagnostic test **(A)**. Predicted values from multiple regression analysis were used to generate Receiver Operator Characteristic (ROC) curves. Use of only m/z 6440 in the model (sensitivity=94%; specificity=80%) gave an area under the curve (AUC) of 0.93 (95% confidence interval 0.82-1.03; p<0.0001), while use of only m/z 3699 in the model (sensitivity=67%; specificity=78%) gave an area under the curve (AUC) of 0.70 (95% confidence interval 0.50-0.90; p=0.065). Incorporation of both ions (sensitivity=93%; specificity=90%) into the model resulted in an AUC value of 0.97 (95% confidence interval 0.93-1.02; p<0.0001). Comparison of AUCs derived from m/z 6640 and m/z 3699 alone or in combination demonstrated a statistically significant improvement in prediction using a dual-marker model (m/z 6640 alone vs m/z 3699 alone: p=0.012; m/z 6640 + m/z 3699 vs m/z 3699 alone: p=0.007; m/z 6640 + m/z 3699 vs m/z 6640 alone: p=0.023) **(B)**.

Table 3.5: Discovery Study Test Set patient characteristics. Demographic and clinical information relating to 32 AIS and 32 Non-AIS (16 TIA, 16 stroke mimics) patient plasma samples comprising the Test Set. Values are median and IQR.

	Total Population (n=64)	AIS and TIA Positive combined (n=48)	AIS (n=32)	TIA Positive (NIHSS<5) (n=11)	TIA Positive (NIHSS≥5) (n=5)	TIA Positive Total (n=16)	Stroke Mimics (NIHSS<5) (n=8)	Stroke Mimics (NIHSS≥5) (n=8)	Stroke Mimics Total (n=16)	AIS&TIA vs Stroke Mimics (p-value)
Age	69.50 (56.25-79.00)	69.50 (57.00-79.00)	75.50 (66.25-81.00)	55.00 (43.00-71.00)	57.00 (35.50-74.50)	56.00 (41.50-70.75)	73.00 (46.25-79.75)	72.00 (59.00-82.75)	72.00 (51.50-79.75)	0.774
Gender	33F, 31M	26F, 22M	16F, 16M	6F, 5M	4F, 1M	10F, 6M		4F, 4M	7F, 9M	0.47
Confirmation of Stroke	34 CT, 26 MRI	25 CT, 19 MRI	19 CT, 13MRI	2 CT, 5 MRI	4 CT, 1 MRI	6 CT, 6 MRI	3 CT, 5 MRI	6 CT, 2 MRI	9 CT, 7 MRI	0.491
OCSP Classification	10 LACS, 20 PACS, 4 POCS, 12 TACS, 2 Uncertain, 16 Non-Stroke	5 LACS, 19 PACS, 2 POCS, 12 TACS, 10 Uncertain	3 LACS, 16 PACS, 1 POCS, 12 TACS	2 LACS, 2 PACS, 1 POCS, 6 Uncertain	1 PACS, 4 Uncertain	2 LACS, 3 PACS, 1 POCS, 10 Uncertain	3 LACS, 1 PACS, 1 POCS, 3 Non-Stroke	2 LACS, 1 POCS, 1 Uncertain, 4 Non-Stroke	5 LACS, 1 PACS, 2 POCS, 1 Uncertain, 7 Non-Stroke	0.003
Time of Blood Sampling (hours)	4.42 (2.44-8.06)	3.84 (2.25-6.94)	3.34 (2.25-7.19)	4.00 (2.00-7.50)	5.00 (2.25-6.63)	4.25 (2.13-7.06)	4.46 (2.38-12.88)	11.63 (5.00-20.63)	7.75 (3.11-15.63)	0.053
Previous MI/Angina	23Y, 41N	16Y, 32N	12Y, 20N	3Y, 8N	1Y, 4N	4Y, 12N	4Y, 4N	3Y, 5N	7Y, 9N	0.452
Previous Stroke/TIA	18Y, 46N	13Y, 35N	7Y, 25N	2Y, 9N	4Y, 1N	6Y, 10N	3Y, 5N	2Y, 6N	5Y, 11N	0.748
Admission NIHSS	6.00 (3.00-10.00)	7.50 (5.00-12.50)	10.00 (7.00-15.50)	1.00 (0.00-2.00)	6.00 (5.00-8.50)	2.00 (1.00-5.00)	1.00 (1.00-2.75)	5.50 (5.00-7.00)	4.50 (1.00-5.75)	0.002



C

	Training Set		Test Set	
	m/z 3699	**m/z 6640**	**m/z 3699**	**m/z 6640**
Ischaemic Events Mean Relative Ion Intensity % (sd)	2.29 (1.03)	47.09 (20.91)	2.01 (1.19)	70.62 (16.52)
Non-Ischaemic Events Mean Relative Ion Intensity % (sd)	3.10 (1.21)	16.64 (12.06)	2.97 (0.87)	49.94 (6.16)
Fold Difference (Ischaemic/Non-Ischaemic)	0.86	1.73	0.68	1.41
p-value	0.006	<0.0001	0.0006	<0.0001

Figure 3.8: Individual scatter plots of m/z 3699 and m/z 6640. The 2 protein ions selected by forward stepwise discriminant function analysis of the Training Set data at pH4 (m/z 3699, m/z 6640) demonstrated good discrimination between ischaemic (AIS and TIA) and non-ischaemic (stroke mimic) groups. Scatter plots of ion intensity from both Training Set **(A)** and Test Set **(B)** samples illustrate their ability to discriminate between the patient groups with high sensitivity and specificity. Mean intensities of each classification group are presented, with data points representing the average intensity of duplicates from each patient sample. In both Training and Test sets, m/z3699 is significantly downregulated during ischaemia, while m/z 6640 is significantly upregulated (Mann-Whitney test p-value) **(C)**.

Table 3.6: Logistic regression analysis of Test Set demographic variables. A number of demographic variables available for the Test Set patients (32 AIS and 32 non-AIS) were tested for their association with relative ion intensity of m/z 6640. Diagnostic classification (ischaemic vs non-ischaemic) proved to have the strongest association with m/z 6640 levels.

	Odds Ratio	95% CI	p-value
Diagnosis (Ischaemic vs Non-Ischaemic)	12.91	4.22-39.49	<0.0001
Gender	0.43	0.12-1.57	0.201
Time of Blood Sampling	-3.33	-8.87-2.40	0.252
Previous History of MI	0.83	-2.58	0.078
Previous History of Stroke/TIA	1.10	0.24-4.18	0.033
Admission NIHSS	0.89	-0.35-3.18	0.007

3.3.2.3 Selection of a Candidate Biomarker

Both m/z 3699 and m/z 6640 were identified as being significantly differentially expressed in both Training and Test sets (Figures 3.6, 3.7). In particular, m/z 6640 had been detected in both the pilot study (pH3, 4 and 8) and Training set (pH4 and 8) and exhibited potential to discriminate between TIA and stroke mimic groups. Furthermore, m/z 6640 was detected at a higher intensity (AIS and TIA mean = 70.6, sd = 16.5, stroke mimic mean = 49.9, sd = 6.1) than m/z 3369 (AIS and TIA mean = 2.0, sd = 1.2, stroke mimic mean = 3.0, sd = 0.8), indicating a more robust fold difference and allowing greater ease in purification of the protein. For these reasons, m/z 6640 was chosen as the best candidate to take forward for identification.

As mentioned in Chapter 2, a significant difference in mean biomarker levels between study groups must be considered within the context of the degree of overlap between the two populations. The normal distributions for ischaemic and non-ischaemic populations based upon relative ion intensities of m/z 6640 from the Test Set demonstrated reasonably good separation (Figure 3.9). Furthermore, calculation of Standardised Mean Difference (SMD) between the two study groups resulted in SMD 1.41 (95% CI 0.79 – 2.03; $p<0.0001$), relating to an estimated odds ratio (OR) of 12.91 (95% CI 4.22 – 39.49; $p<0.0001$).

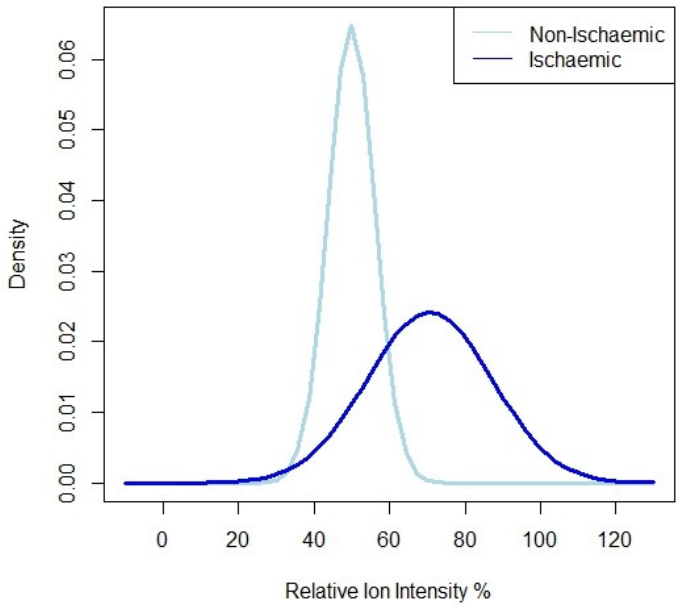


Figure 3.9: Distribution profile of m/z 6640 from Test Set ischaemic and non-ischaemic populations. Normal distribution profiles of m/z 6640 relative ion intensities taken from Test Set ischaemic and non-ischaemic populations demonstrate reasonable separation, implying good discriminatory power with few false positive and false negative results.

3.3.3 Identification of Candidate Marker m/z 6640

A number of studies have reported the identity of m/z 6640 as apolipoproteinC-1 (ApoC-1), using both MALDI-TOF (Chang et al. 2013) and SELDI-TOF (Allard et al. 2004; Takano et al. 2008; Xue et al. 2010) techniques. Hence it was considered that this was the most likely identity of the candidate marker.

3.3.3.1 Immunodepletion of m/z 6640

In order to test the above hypothesis, SELDI-TOF MS immunodepletion was attempted with one monoclonal and one polyclonal anti-human ApoC-1 antibody. A subgroup of 24 samples (8 AIS, 8 TIA, 8 stroke mimics) were randomly selected to comprise a pooled plasma sample for this immunodepletion pull-down experiment. SELDI-TOF MS processing was performed according to the protocol performed in the Test Set discovery study, using CM10 ProteinChip arrays and pH4 binding buffer.

Incubation of the pooled plasma samples directly with a rabbit polyclonal anti-ApoC-1 antibody demonstrated a depletion of m/z 6640 (full-length, single charged ApoC-1 ion), as well as m/z 6440, the single-charged ApoC-1 fragment ion without the N-terminal threonine-proline residues (Chang et al. 2013). However this effect was not specific to these two ions. Incubation of the plasma sample with IgG control antibody showed that this lack of specificity was not due to antibody isotype (Figure 3.10). Therefore this antibody was not deemed to be appropriate for confirmation of m/z 6640 identity. Subsequent incubation of the pooled plasma sample with a mouse monoclonal anti-ApoC-1 antibody failed to demonstrate any pull-down effect, even at high antibody concentrations (18μg/ml). With both antibodies, the findings above were unchanged by the removal of Urea/CHAPS from the binding buffer, which was theorised to be interfering with the binding of antibody to the ApoC-1 protein (Figure 3.10).

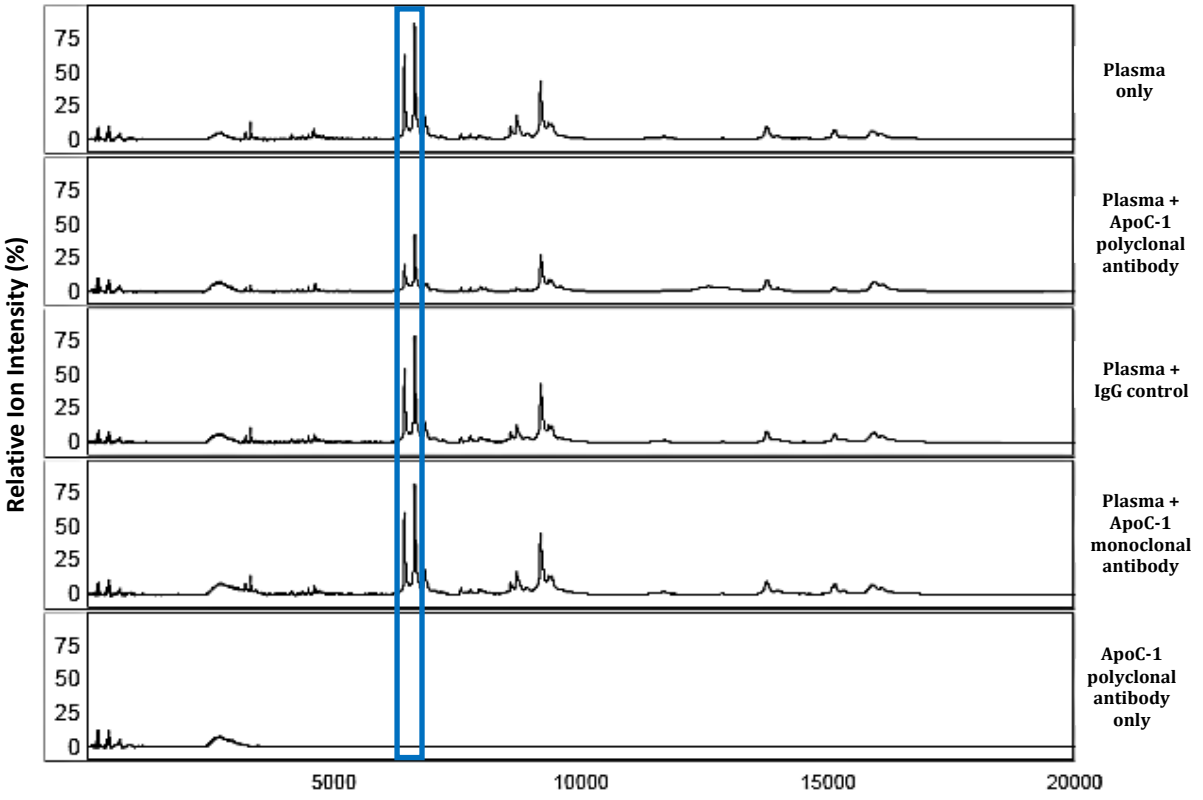


Figure 3.10: Immunodepletion of m/z 6640, theorised to be apolipoproteinC-1 (ApoC-1). A pooled plasma sample was diluted in pH4 buffer and incubated with either a rabbit polyclonal or mouse monoclonal anti-human ApoC-1 antibody, applied to CM10 ProteinChips and then processed by SELDI-TOF MS. The full ApoC-1 m/z 6640 ion, as well as its fragment without the N-terminal Threonine-Proline (m/z 6440) are highlighted in blue. Control incubations were performed using anti-ApoC-1 antibody only, plasma only, and affinity-purified IgG control antibody. Rabbit polyclonal ApoC-1 antibody depletion (6µg/ml) was not specific to m/z6640, while the monoclonal antibody (18µg/ml) failed to demonstrate any depletion of the m/z 6640 peak. Similar results were observed with or without the addition of 9M urea, 2% CHAPS to the plasma sample. No interaction was observed between plasma and IgG control antibody, and no peaks were observed when only ApoC-1 antibody was added to CM10 ProteinChips. All spectra represent samples profiled at pH4 with CM10 chips, according to the original protocol.

3.3.3.2 Measurement and Quantification of m/z 6640 by ELISA

Despite immunodepletion of m/z 6640 failing to confirm the protein ion's identity as ApoC-1, an ApoC-1 ELISA was performed to further test this hypothesis.

Plasma samples from 40 patients from the Training and Test Sets were chosen at random, covering the range of m/z6640 ion intensities observed via SELDI-TOF MS (30-90%). An ELISA performed on these samples allowed quantification of ApoC-1 protein levels in μg/ml. One plasma sample was also spiked with a range of concentrations (0 - 16μg/ml) of authentic ApoC-1 protein (obtained from the ELISA kit) and then analysed on a CM10 ProteinArray chip at pH4 with SELDI-TOF MS. The intensity of m/z6640 increased linearly with increasing concentrations of ApoC-1 and the association between relative ion intensity and concentration of ApoC-1 demonstrated good correlation (r^2=0.91). This relationship was used to estimate ApoC-1 levels in a subgroup of 24 samples which were run on ELISA and also on SELDI-TOF MS. ApoC-1 levels determined by ELISA corresponded well with m/z 6640 relative ion intensities obtained from SELDI-TOF MS (r^2=0.80; Figure 3.11A).

The high correlation between SELDI-TOF MS relative ion intensities and ApoC-1 levels measured by ELISA provided strong evidence to suggest that the identity of m/z 6640 was indeed ApoC-1. To further confirm this identification, SELDI-TOF MS was performed on a preparation of authentic ApoC-1 protein, obtained from the ELISA kit. This yielded two peaks at m/z 6440 and m/z 6640 with similar shapes to the ions observed in plasma, as expected. Additionally, spiking of a plasma sample containing detectable levels of m/z 6440 with authentic ApoC-1 protein increased the intensity of these two ions without altering the intensity of other protein ions in the profile (Figure 3.11B).

ApoC-1 levels, determined by ELISA in the total 40 plasma samples (AIS=10; TIA=10; stroke mimic=15) were significantly higher in the ischaemic group than the non-ischaemic group (ischaemic=1.42μg/ml, non-ischaemic=0.90μg/ml; p=0.001; Student's t-test on log-transformed data). An ROC curve constructed from these levels gave an AUC of 0.84 (95% CI 0.71-0.97), with sensitivity of 83% and specificity of 70% at a cut-off level of 1.85μg/ml. This again confirmed the potential of ApoC-1 to be a valuable diagnostic biomarker for distinguishing ischaemic cerebral events from stroke mimics (Figure 3.12).

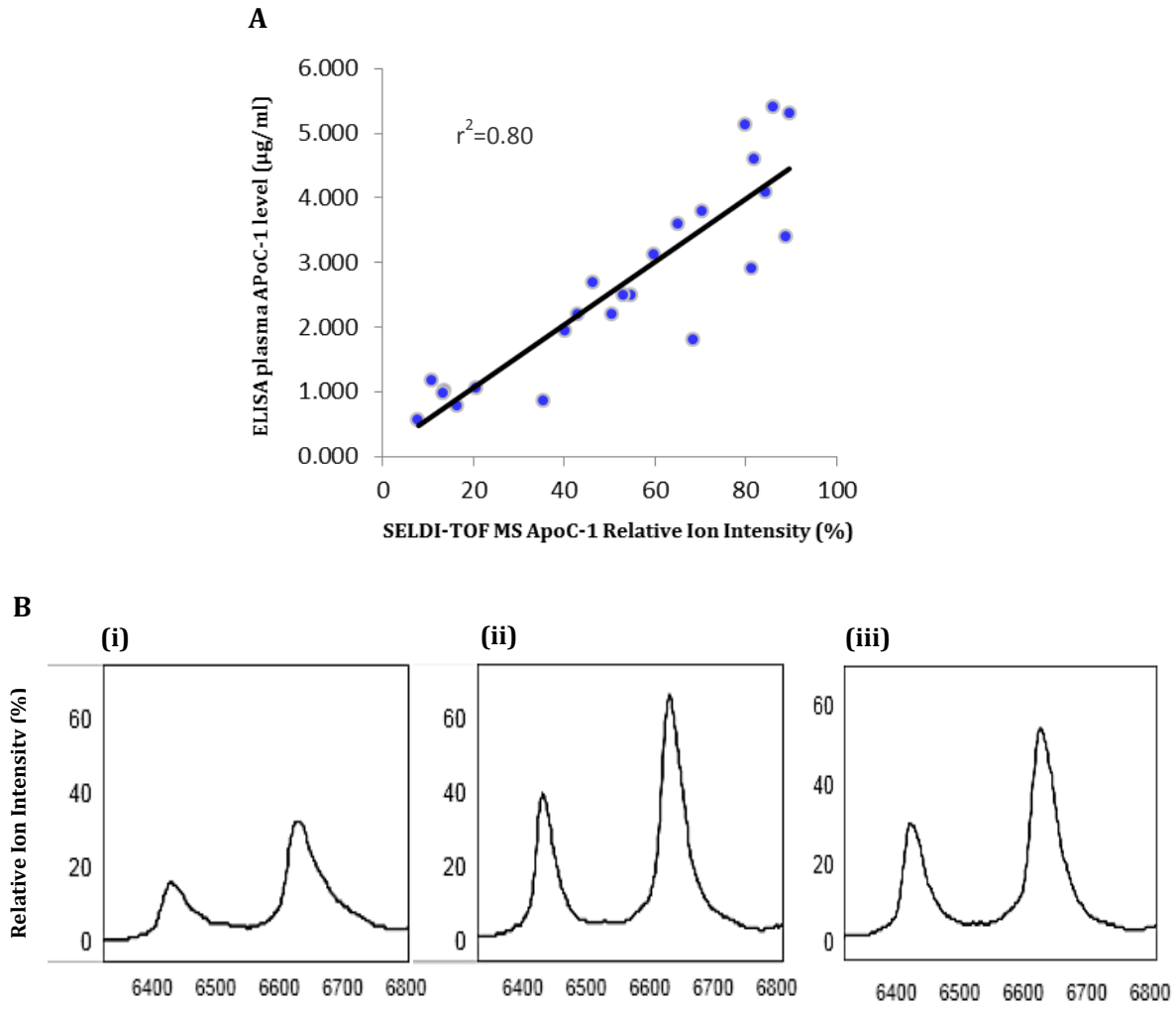


Figure 3.11: Measurement and quantification of m/z 6640 by ELISA. (A) Authentic human ApoC-1 was added to a plasma sample at various concentrations and the intensity of the m/z 6640 ion determined by SELDI-TOF MS. This relationship was used to correlate the ApoC-1 concentration in 24 plasma samples run on ELISA as well SELDI-TOF MS. ApoC-1 relative ion intensities determined by SELDI-TOF MS corresponded well to ApoC-1 levels determined by ELISA (r^2=0.80), regardless of the sample classification grouping. **(B)** Confirmation of the identity of m/z 6640. The intensity of m/z 6640 is shown in samples containing (i) plasma only (15µl) (ii) a mixture of ApoC-1 (4µg/ml) and plasma (15µl) and (iii) authentic human ApoC-1 only. All samples were applied to CM10 ProteinChip arrays in pH4 binding buffer and analysed under the conditions described earlier. The intensity of m/z 6440 and m/z 6640 increased when plasma was spiked with authentic human ApoC-1, without altering the intensity of other protein ions in the spectrum. Analysis of authentic human ApoC-1 only with SELDI-TOF MS produced major peaks at m/z 6440 and m/z 6640 that matched the size and shape of these ions in plasma.

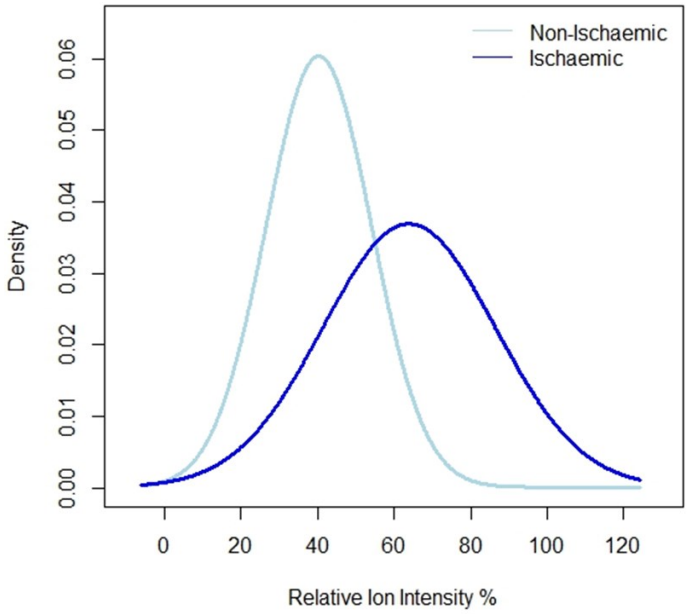


Figure 3.12: Distribution profile of ApoC-1 in ischaemic and non-ischaemic populations. Normal distribution profiles of ApoC-1 relative ion intensities taken from ischaemic and non-ischaemic patient samples (AIS=10; TIA=10; stroke mimic=15) on which ApoC-1 ELISA was performed. The distribution profiles demonstrate reasonable separation, implying good discriminatory power, relating to an AUC of 0.84 (95% CI 0.71-0.97), with sensitivity of 83% and specificity of 70% at a cut-off level of 1.85µg/ml.

3.4 Discussion

The rapid development of ischaemic stroke pathophysiology means that prompt diagnosis is critical for preventing large-scale brain damage and early intervention is crucial, with approximately 2 million neurons lost per minute after stroke onset (Saver 2006). A blood-based diagnostic test that is able to detect acute cerebral ischaemia could substantially improve patient management. The aim of this study was to utilise a whole proteome approach to identify plasma proteins with diagnostic value for acute ischaemic stroke. Using SELDI-TOF MS, two candidate biomarkers were obtained for the differentiation of acute cerebral ischaemia from stroke mimics. One of these plasmatic markers was identified as apolipoproteinC-1 (ApoC-1) and was found to be up-regulated during acute cerebral ischaemia. The other marker, m/z 3699, exhibited down-regulation during acute cerebral ischaemia.

The difficulties in diagnosing TIA and the importance of accurate diagnosis in order to provide adequate patient management and prevent subsequent ischaemic events, has been previously described in Chapter 1. The differentiation between TIAs and stroke mimics poses a particular challenge as both appear negative upon CT imaging. While recruiting patients from the Charing Cross Hospital Hyper Acute Stroke Unit it was noted that almost a third of patients admitted to the ward transpired to be stroke mimics, consistent with findings from an observational study that reported a diagnosis of stroke mimics in 31% of patients, of whom 40% were labelled as labelled "possible stroke or TIA" (Hand et al. 2006). Therefore, even though a recent cohort study has demonstrated the relative safety of (incorrectly) treating these patients with thrombolysis (Zinkstok et al. 2013), the ability to triage stroke mimics effectively would greatly improve time and resource management for clinicians.

With this in mind, studies have attempted to investigate biomarkers differentiating between ischaemic events and stroke mimics using a candidate protein approach (detailed in Chapter 2), and have postulated a role for 11 candidate biomarkers including NMDA receptor peptide NR2 (Dambinova et al. 2012), NSE, S100B (Gonzalez-Garcia et al. 2012), plasma DNA (Rainer et al. 2007), GFAP (Foerch et al. 2012), CRP, D-Dimer (Glickman et al. 2011) IMA (Ahn et al. 2011), PTA (Doehner et al. 2012) and VAP-1 (Airas et al. 2008). However, thus far none of these markers seem to have shown sufficient value within a clinical setting.

ApoC-1 as Diagnostic Biomarker for AIS

The combination of selective protein binding with quantitative mass detection makes SELDI-TOF MS a simple and attractive method for comparative protein analysis. Nonetheless, proteins

are not subjected to digestion or tandem mass spectrometry; therefore candidate protein ions discovered through SELDI-TOS MS require downstream processing for identification.

It has been argued that proteomic patterns should be based upon proteins of known identity (Diamandis 2002). However this current study, like others (Petricoin et al. 2002), maintains that the diagnostic utility of markers identified through proteomics can be independent of protein identity.

Identification of m/z 6640 was unsuccessful by antibody immunodepletion. Direct immunodepletion requires precise titration of plasma in order to allow a network of interactions to form between the antibody and antigen; the inability to find the exact antibody:antigen ratio required may explain the failure of this method. Immobilisation of the anti-ApoC-1 antibody, through protein-A agarose beads (Takano et al. 2008) or RS100 ProteinChip array (Xue et al. 2010) may have been a more reliable technique for immunoprecipitation. ELISA provided convincing evidence of the identity of m/z 6640 being ApoC-1. Nonetheless, conclusive verification of this would be provided by tryptic digestion of plasma run on 2D SDS PAGE, with subsequent processing by liquid chromatography and tandem MS (LC-MS/MS).

The finding that ApoC-1 levels increase during acute ischaemic events is supported by its key role in lipid metabolism. ApoC-1 is a secreted plasma protein produced by the liver and comprises ~10% of circulating very low density lipoprotein (VLDL) and ~2% of high density lipoprotein (HDL), being up-regulated when monocytes differentiate into macrophages. In humans the protein has a high basic pI due to its high content of lysine, and the acidic form, present in bonobos, chimpanzees and orang-utans, has become a pseudogene (Jong et al. 1999).

ApoC-1 is synthesised with a 26-residue signal peptide that is cleaved in the rough endoplasmic reticulum. Subsequently, residues 7 to 24 and 35 to 53 of the protein are important for lipoprotein binding (Rozek et al. 1995). MALDI-TOF MS has ascertained that two isoforms of the protein can be detected in plasma (Jin & Manabe 2005): full ApoC-1 (m/z 6640) and truncated ApoC-1 lacking N-terminal threonine and proline (m/z 6440); both these isoforms were detected in this current study (Figures 3.4 and 3.6).

Upregulation of ApoC-1 displaces ApoE from the VLDL complex, decreasing their clearance from plasma (Kolmakova et al. 2004). ApoC-1 inhibits phospholipase A2 activity and cholesterylester transfer protein, a plasma protein that facilitates the homoexchange of cholesteryl esters from HDL to triglycerides on VLDL and LDL. ApoC-I also regulates plasma triglycerides levels through

inhibition of triglyceride-rich lipoprotein uptake via hepatic receptors, particularly the LDL receptor-related protein (Cohn et al. 2003). Thus ApoC-1 presence on VLDL and HDL may prolong triglyceride residence time within the circulation (Jong et al. 1999), enabling the conversion of triglyceride-rich lipoproteins to LDL and exacerbating a hypertriglyceridemic state that could have a pro-atherogenic effect (Madden 2012; Cohn et al. 2003).

The findings of this current study provide validation of previous studies that have reported an upregulation of ApoC-1 in AIS, with ApoC-1 being implicated as a diagnostic biomarker to distinguish between AIS and ICH using proteomics techniques such as SELDI-TOF MS (Allard et al. 2004) and Selective Reaction Monitoring on a triple quadrupole mass spectrometer (Lopez et al. 2012). ApoC-1 has been shown to be uniquely elevated in post-prandial VLDL and chylomicron particles in men with early asymptomatic atherosclerosis (Bjorkegren et al. 2002), and ApoC-1 enriched HDL activation of the neutral sphingomyelinase-ceramide signalling pathway leads to apoptosis in aortic smooth muscle cells, possibly promoting plaque rupture (Kolmakova et al. 2004). Global gene expression analysis also uncovered an ApoC-1-centric network associated with atherosclerotic carotid artery disease (Diez et al. 2010), reinforcing findings that ApoC-1 content of VLDL particles is associated with plaque size in patients with carotid atherosclerosis (Noto et al. 2008). The location of ApoC-1 is of significance, with systemic ApoC-1 increasing development of atherosclerosis, but local macrophage production of ApoC-1 in the arterial wall having little effect on plasma lipid levels or atherosclerotic lesion area in mice (Westerterp et al. 2007). Based upon its role in promoting carotid artery disease and plaque rupture (a process that is not as likely to be present in stroke mimics), these studies provide strong biological evidence that ApoC-1 is a plausible diagnostic biomarker for AIS and TIA.

A Multi-Marker Diagnostic Test for Acute Cerebral Ischaemia

Given the complexity of AIS pathophysiology, it is very likely that a multi-marker test would provide superior sensitivity and specificity to a single marker. Proteomic approaches can interrogate hundreds of proteins simultaneously and are therefore ideal for assessing diseases such as stroke, in which complex protein cascades are induced in response to injury. A number of studies have attempted this multi-marker approach (Dambinova et al. 2002; Reynolds et al. 2003; Lynch et al. 2004; Laskowitz et al. 2005; Montaner et al. 2005; Tuttolomondo et al. 2009), with a panel of four markers (S100B, von Willebrand Factor, MMP9, VCAM-1) demonstrating 90% sensitivity and specificity to distinguish AIS from controls (Lynch et al. 2004), and a similar panel of five markers (S100B, von Willebrand Factor, MMP9, MCP-1, BNGF) demonstrating 92% sensitivity and 93% specificity to distinguish AIS from healthy controls (Reynolds et al. 2003).

However, the relevance of these markers is doubtful as the clinical problem is not the difficulty in differentiating AIS from healthy controls, but rather to be able to diagnose ischaemic events from non-ischaemic events. The "Triage® Stroke Panel", comprising many of the markers above, failed to demonstrate any significant difference between AIS and stroke mimic populations in a recent prospective study (Knauer et al. 2012). Although the identification and validation of m/z 3699 was not possible in the current study, its ability to improve the specificity of a test based upon m/z 6640 alone warrants further investigation into its incorporation into a multi-marker test to distinguish between ischaemic and non-ischaemic cerebral insults.

Clinical Utility of Candidate Biomarkers

In order to examine their relative contribution to ischaemic stroke diagnosis, the utility of the candidate biomarkers must be assessed within the context that they will be used. Existing clinical variables have shown to be unhelpful for the classification of stroke mimics (Tobin et al. 2009). In Chapter 2, meta-analysis of published data identified only one protein, S100B, as being significantly able to differentiate between AIS and stroke mimics. The current study found that ApoC-1 has 83% sensitivity and 70% specificity at a cut-off of 1.85μg/ml, indicating that ApoC-1 may be a more valuable biomarker than S100B, which was recently reported as having 54% sensitivity and 84% specificity for discriminating AIS from stroke mimics (Park et al. 2013).

A recent meta-analysis reported sensitivity and specificity values of 39% and 100% respectively for CT imaging, and 99% and 92% respectively for MRI imaging (Brazzelli et al. 2009). Being a retrospective study, the value of using the proposed single/multi-marker biomarker test in conjunction to CT or MRI could not be investigated; prospective analysis of this will shed light on the actual value of the test, over and above standard clinical practice. The finding that ApoC-1 alone has 83% sensitivity and 70% specificity to detect acute cerebral ischaemia implies that a diagnostic test based upon this protein has the potential to identify double the number of true positives than currently identified by CT, and could be a valuable diagnostic tool when used in conjunction with CT imaging. The value of being able to distinguish between TIAs and stroke mimics in particular has been discussed in Chapter 1, and the substantial improvement of sensitivity provided by an ApoC-1 test over CT imaging hold promise for its value as a diagnostic biomarker. Both this single marker test and a test based upon ApoC-1 in combination with m/z 3699 do not provide the diagnostic utility of MRI imaging. However, given the limited availability of MRI within an acute clinical setting (particularly in developing countries), the tests could have substantial value either pre- hospital admission (e.g. while a patient is in an ambulance) or for triaging before CT imaging.

Choice of Proteomic Platform

The presence of a large number of proteins in blood plasma and the ease of sample collection within a clinical setting makes human plasma an excellent material for discovering diagnostic biomarkers. However, the complexity and large concentration ranges of proteins within plasma samples present major challenges for their analysis. A handful of highly abundant proteins, such as albumin, represent over 90% of the plasma proteome and can mask the detection of low-abundance proteins that are likely to be more valuable as clinical biomarkers (Anderson & Anderson 2002). Indeed, a mere 21 proteins represent 99% of the plasma proteome, hence the challenge is to probe the remaining 1% for disease-specific markers (Righetti et al. 2005). Furthermore, the complete measured dynamic range of proteins spans more than 10 logs of molar abundance (Hortin & Sviridov 2010). Nonetheless, the HUPO PPP Specimens Committee recommends that plasma is preferable to serum because it contains less peptides of degradation and therefore presents less variability (Omenn et al. 2005; De Bock et al. 2010).

There are a number of advantages of using SELDI as a proteomic technique (reviewed in Bons et al. 2006). The unique chromatographic properties of the ProteinChip arrays preclude the necessity for pre-fractionation steps, which are usually performed on plasma samples in order to simplify the proteome and maximise protein separation and resolution. Another advantage of the surface enhanced process is that components such as detergents, lipids and salts that commonly cause problems with other analytical tools are washed away in the SELDI process (Wiesner 2004).

Importantly, unlike the relatively time-consuming process of purification, processing and analysis required for LC/MS, SELDI-TOF MS allows hundreds of samples to be screened in a short time period, providing an opportunity to compare patient-patient variability and a rapid screening for potential biomarkers. As a soft ionisation technique SELDI-TOF MS reduces the frequency of protein fractionation, which aids in distinguishing between mature proteins and fragments produced from larger precursor proteins. This is more difficult to ascertain with LC/MS-MS, which is susceptible to higher levels of fractionation.

The analyser used for SELDI-TOF MS is designed to achieve high-sensitivity quantification, and the ion source and detector are adapted for efficient ion transmission and detection over a wide mass range. Although the system is generally more favourable for proteins and peptides with molecular weight lower than 20kDa, it is able to analyse proteins over a wide molecular weight range (0-200kDa) and the entire isoelectric point (pI) range, giving it an advantage over other proteomic techniques such as 2D electrophoresis (limited to protein separation >10kDa, pI 3-10; Wiesner 2004).

Choice of Study Design

A commonly recommended study design for large-scale proteomics and SELDI-TOF MS experiments is to collect a large number of samples and randomly assign them to Training and Test sets (Burke 2005; Petricoin et al. 2002). This design is especially important for avoiding overfitting if considering a multi- rather than single biomarker model (Baker et al. 2002). Furthermore, it is suggested that candidate markers be further replicated in an independent Validation set (Baker et al. 2002). The scarcity of published literature using this study design suggests that previously reported biomarkers for AIS identified via high-throughput approaches may not have been reproduced and confirmed in subsequent studies.

Given the challenges of collecting blood samples within the acute phase (<24hours) from stroke patients who are able to consent, in this study an initial 13 AIS subjects were recruited for investigation in a pilot study. This allowed optimisation of conditions favouring biomarker discovery, and more importantly, justified the use of SELDI-TOF MS for detecting a protein profile that is deranged during acute ischaemic insult. Despite the small number of patients, the effect of inter-subject variability was minimised by comparison of 11 paired plasma samples, taken from the same individual at acute and follow-up time intervals. Thus there is greater likelihood that differences observed between these groups represent genuine biological changes in peptide expression than if independent healthy subjects had been chosen as control samples.

The findings of the pilot study allowed attainment of over 400 blood samples from collaborators at Edinburgh University. These samples were divided into well-powered Training (40 subjects), Test (64 subjects) and Validation (300 subjects) sets. The Training set was used to optimise experimental conditions and found two ions (m/z 3699 and m/z 6640) that could be incorporated into a multi-marker diagnostic test for distinguishing acute ischaemic events from stroke mimics. Results from the Test set further confirmed the performance of these two ions, although m/z 6640 was deemed to be the best candidate for further investigation.

Limitations

SELDI-TOF MS investigations have been criticised due to a general lack of standardised methodology, rendering results difficult to reproduce (Alaoui-Jamali & Xu 2006). Reproducibility and CVs are affected by factors including accuracy of matrix application, plasma concentration and settings used on the SELDI-TOF MS instrument for data collection. These issues were optimised in the pilot study, and variability due to blood collection and plasma processing were minimised by following HUPO PPP Specimen Collection and Handling Committee recommendations (Rai et al. 2005).

Appropriate characterisation of patients, especially individuals with TIA, was difficult during the acute phase of ischaemia in which blood samples were collected. Similarly, diagnosis of stroke mimics was also challenging. Ironically, this again highlights the potential utility of a test that could correctly differentiate between TIA and stroke mimics.

A number of other variables may impact the findings of this study; statistically significant differences between age, gender, OCSP classification, admission NIHSS and time of blood sample collection between ischaemic and non-ischaemic populations are likely to confound the results. In particular, the time course of ApoC-1 and m/z 3699 over the acute phase of cerebral ischaemia needs investigation in order to optimise the threshold values of the biomarkers for diagnosis and determine how soon after AIS/TIA onset the test could be used.

Concluding Remarks

This study has demonstrated the ability of SELDI-TOF MS to detect differential protein profiles between ischaemic (AIS, TIA) and non-ischaemic (stroke mimic) cerebral events and its value in preliminary diagnostic biomarker discovery. One protein in particular, ApoC-1, has been shown to have promising potential for being a novel diagnostic biomarker for acute cerebral ischaemia.

CHAPTER FOUR:

AN INTEGRATIVE GENOMICS APPROACH TO RISK BIOMARKER DISCOVERY FOR ISCHAEMIC STROKE

4.1 Introduction

It is predicted that by 2030, the global number of people suffering from first-ever stroke will be 23 million in the absence of population-wide intervention (Strong et al. 2007). In the UK, stroke incidence and mortality have decreased over the past decade. However, there is still a pressing need for better risk stratification and improvements in risk factor reduction in high-risk individuals (Lee et al. 2011).

Carotid atherosclerosis is associated with ~20% of ischaemic strokes and extracranial carotid atherosclerosis is the most common cause of large artery atherosclerotic (LAA) strokes in Caucasian populations (Mohammad et al. 2010). While carotid plaques causing substantial stenosis may lead to reduced blood flow and ischaemic stroke, the Asymptomatic Carotid Atherosclerotic Study (ACAS) and the North American Symptomatic Carotid Endarterectomy Trial (NASCET) demonstrated that the size of a plaque alone does not determine stroke symptomology, with stroke rates differing significantly between asymptomatic and symptomatic individuals with the same level of stenosis (Gasecki et al. 1995; Baker et al. 2000).

The composition of atherosclerotic plaques is considered to be an important determinant of arterial thrombus development and embolic complications. Biological attributes such as inflammation, matrix degradation, intra-plaque haemorrhage and intra-plaque angiogenesis have been acknowledged as key determinants of plaque stability and future stroke. However, while it is known that atherosclerosis is a complex disease involving multiple genes, pathways and cell types (Lusis 2000), the specific genetic mechanisms underlying plaque development and progression remain unclear. Thus, there exists a translational need for molecular and cellular characterisation of plaque stability for efficient stroke risk stratification. Such elucidation is essential for understanding the pathways responsible for disease progression and in offering biomarkers for identification of atherosclerotic burden and instability, in addition to putative therapeutic drug targets. Identification of high-risk individuals could improve their clinical management and select candidates who would benefit most from stroke-prevention therapies. Furthermore, an integrated model combining traditional risk factors with common genetic variants could more effectively discriminate patients likely to suffer from recurrent ischaemic events.

The proportion of ischemic stroke in a population that can be attributed to a particular risk factor is called the population attributable risk and has been estimated as 60% in total, indicating a 40% gap that may in part be accounted for by genetic variants (Whisnant 1997). A

number of studies have demonstrated that ischaemic stroke and carotid disease share similar environmental and genetic risk factors (Humphries & Morgan 2004). The role of the *9p21* locus has previously been implicated in carotid disease (Congrains et al. 2012) and atherosclerotic stroke (Gschwendtner et al. 2009). This finding has been replicated in stroke GWAS, which have also identified two other loci (*HDAC9*, *6p21.1*) associated with the LAA ischaemic stroke subtype (Holliday et al. 2012; ISGC & WTCC2 2012). Meta-analysis of GWAS from the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium recently identified multiple SNPs associated with carotid intima-media thickening and plaque, mapping to genes related to cellular signalling, lipid metabolism, blood pressure and homeostasis (Bis et al. 2011).

These results demonstrate the ability of GWAS to identify genetic variants that could provide new insights into the mechanisms underlying carotid atherosclerosis and LAA ischaemic stroke. However, while GWAS offers unbiased ways to examine the association of more than a million SNPs associated with disease, the list of significant SNPs obtained often fails to provide mechanistic insights into the underlying biology of the disease being studied, and the challenge remains in extracting functionally and biologically relevant information from GWAS results.

Functional enrichment analysis approaches have been used to overcome this challenge. Hypergeometric testing and GSEA have been successfully employed to investigate molecular pathways underlying cardiovascular and metabolic diseases using data derived from gene expression studies (Sluimer et al. 2007; Diez et al. 2010; Cagnin et al. 2009) and GWAS (Morris et al. 2012; Voight et al. 2010; de las Fuentes et al. 2012). Both approaches have the advantage of reducing the complexity of data obtained from GWAS and gene expression studies from thousands of genes to a handful of pathways. Furthermore, the characterisation of functional molecular pathways that differ between two study populations often holds more information than simple gene lists alone (Khatri et al. 2012). However, these analyses may be susceptible to false positive results and thus require replication and scrutiny in independent cohorts.

One method of replication is to use an integrative genomics approach to correlate the findings from GWAS with gene expression data to elucidate novel molecular mechanisms underlying complex diseases such as carotid atherosclerosis and stroke. Subsequent identification of robust disease-causing pathways can provide new biomarkers for targeted diagnostic and therapeutic intervention. This approach also allows investigation of eQTLs and QTTs or differential gene expression as detailed in Chapter 1, underpinned by the hypothesis that mutations or polymorphisms (non-disease causing genetic variation) near or in genes can have an effect on the expression of a disease-causative gene and are also associated with the disease itself (Ware et al. 2013). Given that 15% of disease-causing point mutations regulate splicing in humans

(Pagani & Baralle 2004), integrative genomics can also be used to investigate the genetic mechanisms underlying alternative splicing, a key process in disease risk and progression but one that is often overlooked. For example, *9p21*, implicated in ischaemic stroke and carotid disease as mentioned above, has been found to have a complex pattern of variable splicing and gene expression change in a study investigating the effects of the haplotype block in expression microarray datasets (Folkersen et al. 2009).

Aims of the Study

Understanding global expression, splicing and genetic regulation of mRNA is of key importance in the field of human carotid atherosclerosis, however to date, no study of this kind has been performed. The aim of this study was to identify novel biomarkers that could aid in the risk-stratification of ischaemic stroke by:

- Using the Affymetrix Human Exon 1.0 ST Array to characterise gene expression and alternative splicing profiles underlying carotid stenosis and plaque instability.
- Utilising functional enrichment analysis approaches to identify gene sets associated with symptomatic carotid stenosis in both the gene expression study and a GWAS meta-analysis of carotid disease in ischaemic stroke patients.
- Employing an integrative genomics approach to correlate results from the whole transcriptome study with the GWAS meta-analysis to elucidate genetic mechanisms underlying symptomatic carotid stenosis.

4.2 Methods

An overview of the study workflow and methodologies used in this chapter is given in Figure 4.1.

4.2.1 Gene Expression Profiling Study

4.2.1.1 Sample Selection

Specimens from the Charing Cross Hospital carotid endarterectomy tissue bank (maintained by Professor Alun Davies, Imperial College London) were chosen from eighteen highly characterised individuals with asymptomatic or symptomatic carotid disease. Patients characterised as 'symptomatic' exhibited focal neurological symptoms pertaining to the ipsilateral cerebral hemisphere or amaurosis fugax, as confirmed by CT or MRI. This diagnosis was further supported by individual discussion of patients in a clinical MDT meeting of vascular surgeons and neurologists, considering contributing factors such as previous history of vascular events or vascular risk factors.

Each flash-frozen carotid endarterectomy specimen was dissected on dry ice into two distinct areas relating to stenosing plaque (>20% stenosis) or non-stenosing tissue (<20% stenosis) by an experienced vascular surgeon (Mr Joseph Shalhoub, Imperial College NHS Trust). Multiple sections were taken where possible. Percentage stenosis, defined by North American Symptomatic Carotid Endarterectomy Trial (NASCET) criteria, was calculated from the internal carotid artery peak systolic velocity/common carotid artery end diastolic velocity ratio, as measured by Doppler ultrasound prior to surgery.

Plaque segments were classified as follows: stenosing areas from symptomatic plaques were considered to be unstable (ruptured), whilst stenosing areas from asymptomatic plaques were considered stable (Figure 4.1). This classification was also based upon plaque histology by macroscopic examination, with stable segments covered by a smooth luminal surface indicative of an intact fibrous cap, while unstable plaques commonly displayed an ulcerated surface alongside evidence of thrombosis or haemorrhage. Patient characteristics and tissue section classifications are detailed in Table 4.1 and Figure 4.2.

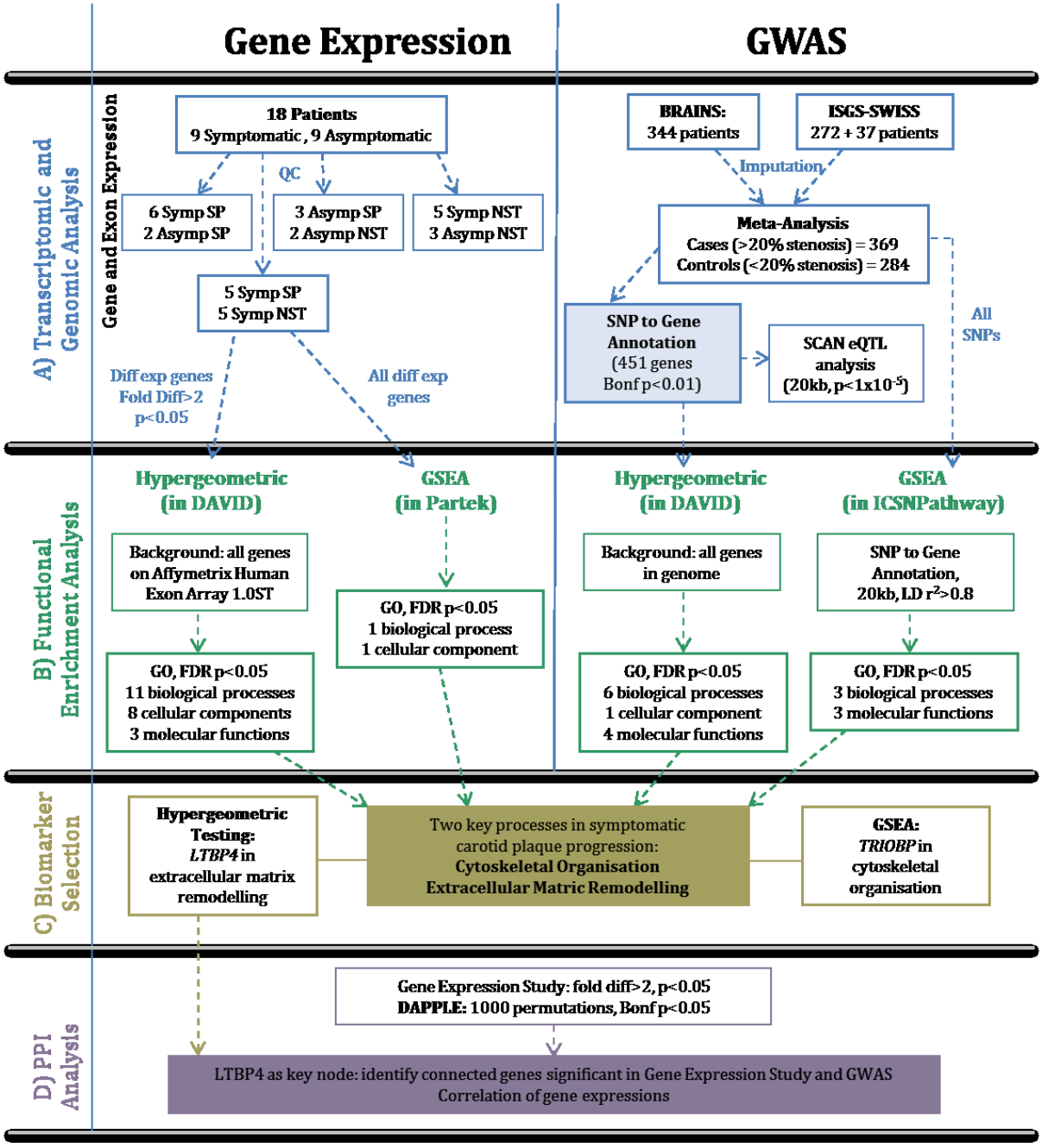


Figure 4.1: Study design and methodology flowchart. A) The gene expression study and GWAS were conducted in different cohorts and analysed separately. **B)** Functional enrichment analysis was conducted using two methods, hypergeometric testing and gene set enrichement analysis (GSEA). **C)** Assessment of gene ontology (GO) terms yielded two molecular processes that were found to be significantly enriched in both the gene expression study and GWAS. Two genes (*LTBP4* and *TRIOBP*) within these GO terms, common to both the gene expression study and GWAS, were selected as candidate biomarkers and investigated in more detail to understand whether their expression and splicing profiles were under the control of SNPs. D) Protein-protein interaction (PPI) analysis was performed using significantly differentially expressed genes from the gene expression study, to further examine the role of *LTBP4* as a key node in the PPI disease network and a biomarker able to predict symptomatic carotid plaque progression.

Table 4.1: Gene expression study total patient population characteristics. Demographic, clinical and histological information relating to 10 patients from whom SP and NST carotid endarterectomy tissue samples were taken for gene expression analysis. **AIS,** acute ischaemic stroke; **IHD**, ischaemic heart disease; **NST**, non-stenosing tissue; **PAD**, peripheral artery disease; **SP**, stenosing plaque, **TIA,** transient ischaemic attack.

Patient ID	Tissue Type	Symptoms	Carotid Stenosis (%)	Gender	Age	IHD	PAD	Hypertension	Diabetes Mellitus	Smoking History	Statin Use	Antiplatelet Use
V0012	2 SP + 1 NST	TIA	90-95	F	76	x	x	✓	x	✓	✓	✓
V0162	3 SP + 2 NST	-	90-95	M	75	✓	✓	✓	✓	x	✓	✓
V1139	2 SP + 1 NST	TIA	45-65	M	72	✓	✓	✓	x	x	✓	✓
V1236	1 SP + 1 NST	-	85	M	85	x	✓	✓	x	✓	✓	✓
V1241	2 SP	-	85-95	M	50	x	x	x	x	✓	✓	✓
V1299	2 SP + 1 NST	TIA	90-95	M	79	x	x	✓	x	x	✓	x
V1355	2 SP	TIA	95	M	63	x	x	✓	✓	✓	✓	✓
V1376	1 SP + 1 NST	AIS + AmFu	90	M	60	✓	x	x	✓	✓	✓	✓
V1377	2 NST	AIS	30	M	73	x	x	✓	✓	✓	✓	✓
V1388	1 NST	-	80	F	66	x	x	✓	x	x	✓	x
V1420	1 SP + 1 NST	AIS	55-65	M	51	x	x	✓	x	✓	x	x
V1465	1 SP + 1 NST	-	75-80	M	67	x	x	✓	✓	x	x	✓
V1468	1 SP + 1 NST	AIS	30	M	55	x	x	x	x	x	x	✓
V1470	2 SP + 1 NST	-	95	M	78	x	x	x	x	x	x	x
V1471	1 SP + 1 NST	-	80	M	56	✓	x	✓	x	x	✓	✓
V1477	2 SP + 1 NST	AmFu	60-75	M	54	x	x	✓	x	x	✓	✓
V1478	1 SP + 1 NST	-	90	M	73	x	x	✓	x	x	✓	✓
V1486	3 SP	-	80	M	67	x	✓	x	x	✓	✓	✓

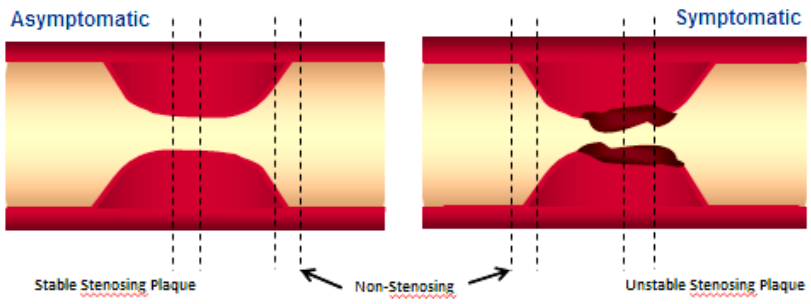


Figure 4.2: Carotid endarterectomy specimen dissection into stenosising plaque or non-stenosing tissue areas. Where possible more than one section was taken from each area, resulting in a total of 44 tissue samples (27 stenosing plaque, 17 non-stenosing tissue) taken from 18 patients. Of these 18 patients, 9 were asymptomatic (yielding 14 stenosing plaque and 8 non-stenosing plaque sections) and 9 were symptomatic (yielding 13 stenosing plaque and 9 non-stenosing plaque sections). Patient characteristics are further detaileid in Table 4.1.

4.2.1.2 RNA Isolation

Total RNA was isolated from carotid endarterectomy tissue samples (50-100mg) using the miRNeasy 96-well kit (Qiagen, UK), and a minimum of one extraction was performed from each tissue sample. Samples were homogenised using the TissueLyser system with 1ml of QIAzol lysis reagent for 4-5 minutes. Following incubation at room temperature for 10 minutes, 200µl of chloroform was added to each sample and the tubes were mixed vigorously for 15 seconds. Tubes were again incubated at room temperature for 10 minutes and then centrifuged at 12,000 *x g* at 4°C for 15 minutes. The aqueous phase (upper layer containing RNA) was transferred to new collection tubes and RNA was precipitated by adding 800µl of 100% ethanol (1.5 volume of the aqueous phase). The samples were transferred into RNeasy Mini spin columns and these were centrifuged at 8000 *x g* for 1 minute to allow the RNA in the solution to bind to the spin column membrane. The supernatant was discarded and samples were washed to remove salt traces and impurities by adding 700µl of washing buffer and centrifuging for 15 seconds at 8000 *x g*. This was repeated 3 times, with the final centrifugation lasting for 2 minutes. Total RNA was eluted from the spin column by the addition of 60µl of RNase-free water, and each sample was divided into 20 µl aliquots.

4.2.1.3 Expression profiling using Affymetrix GeneChip® Human Exon 1.0 ST Arrays

Expression profiling on the Affymetrix GeneChip® Human Exon 1.0 ST Arrays (Affymetrix, High Wycombe, UK) was outsourced to AROS Applied Biotechnology AS company laboratories (http:// www.arosab.com). Total RNA (200ng) was used for generation of cRNA and subsequent cDNA preparation with the Ambion WT Expression Kit, then cDNA was biotin labelled with the Affymetrix Gene-Chip Whole Transcript Sense Target Labelling Assay, as per

manufacturer's instructions. Following this *in vitro* transcription reaction, unincorporated nucleotides were removed using RNeasy columns (Qiagen). Fragmented cDNA was heated to 95^{O}C for 5 minutes and then 45^{O}C for 5 minutes, then loaded onto the Affymetrix Human Exon 1.0 ST array cartridge (all samples on one cartridge, to eliminate batch effect errors). The cartridge was then incubated at 45^{O}C for 16 hours, with constant rotation (60rpm), then washed 10 times with 6x SSPE-T (sodium chloride + sodium phosphate + EDTA + 0.01%Tween-20) buffer at 25^{O}C, followed by 4 washes of 0.5x SSPE-T at 50^{O}C.

For detection, the biotinylated cDNA was stained with a streptavidin-phycoerythrin conjugate (final concentration 2mg/ml; Affymetrix) in 6x SSPE-T for 30 minutes at 25^{O}C then 10 washes in 6x SSPE-T at 25^{O}C. An antibody amplification was performed with biotinylated anti-streptavidin goat antibody (final concentration 3mg/ml; Affymetrix), using goat IgG as a blocking reagent (final concentration 0.1mg/ml; Affymetrix). This was followed by another staining step with the streptavidin-phycoerthrin conjugate (final concentration 2mg/ml; Affymetrix) in 6x SSPE-T for 30 minutes at 25^{O}C then 10 washes in 6x SSPE-T at 25^{O}C. The hybridised Affymetrix Human Exon 1.0 ST arrays were scanned at 560nm using a confocal laser scanning microscope (GeneChip Scanner 3000 7G) and visually inspected for hybridisation artefacts.

4.2.1.4 Quality Control

Following RNA extraction, the NanoDrop ND-1000 Spectrophotometer V3.3.0 was used to assess the concentration and purity of each RNA sample. The concentration of each sample was noted (ng/µl), and samples with 260 nm/280 nm and 260 nm/230 nm ratios of absorbance <1.8 were removed. RNA samples were also applied to an RNA 6000 Nano-LabChip and analysed using the Agilent 2100 Bioanalyzer (Agilent Technologies UK Ltd). RNA Integrity Number (RIN) values were calculated by the 2100 Expert Software (Agilent Technologies UK Ltd) and each RNA electropherogram was inspected to further asses RNA quality. Samples with RIN <3.0 were removed.

Following exon array processing, quality control (QC) parameters were examined using the Affymetrix Expression Console software version 1.2 (www.affymetrix.com), as recommended by Affymetrix guidelines (Affymetrix White Paper: Quality Assessment of Exon and Gene Arrays). Labelling, hybridisation, scanning and background signals were evaluated by probe set summarisation using Robust Multichip Analysis (RMA) (Irizarry, Bolstad, et al. 2003; Irizarry, Hobbs, et al. 2003; Bolstad et al. 2003).

4.2.1.5 Exon Array Data Analysis

Normalisation, filtering and statistical analyses were performed using Partek Genomics Suite v6.6 (Partek Inc., St. Louis, USA). Arrays were pre-processed using RMA quantile normalisation with GC background correction and probesets summarised with median polish were $\log_2$ transformed. Probesets containing SNPs that could affect array hybridisation were identified using the SNPinProbe1.0 database (Duan et al 2008) and PiP Finder Tool (Ramasamy et al 2013) and were excluded from downstream analysis. Exons were summarised to genes using the 90% winsorised mean of the corresponding probe sets, based upon the meta probeset provided by Affymetrix. Filtering was performed so that only the 'core' meta probeset was considered (21,980 gene and 232,448 exon probesets, consisting of RefSeq and full-length GenBank mRNAs) and Affymetrix probe sets were re-mapped onto the human RefSeq build 19 (GRCh37), as detailed in the NetAffx annotation file (HuEx-1_0-st-v2 Probeset Annotations, Release 31).

Differences in gene-level expression and splicing were investigated with ANOVA models to examine significant associations between tissue type (stenosing plaque vs. non-stenosing tissue) or patient diagnosis (asymptomatic vs. symptomatic) with age and gender included as covariates. Paired analysis was performed using repeated measures ANOVA, with covariates as above. Significance p-values for the ANOVA models were calculated using $\log_2$-transformed data, and expression fold differences calculated by least squares (LS) means.

A range of fold change (1.2-2.5) and significance thresholds ($p<0.01$-0.1), with and without FDR correction, were tested on the gene expression and exon level data. Optimal thresholds were deemed to be those that captured the top ~5% of the total data, representing a set of genes/exons that displayed relatively high fold difference and high significance, as previously recommended (Dalman et al. 2012; Yang et al. 2002).

4.2.2 Genome-Wide Association Study (GWAS)

4.2.2.1 Patient Selection

A discovery meta-analysis of two GWAS comprising individuals from the BRAINS and ISGS/SWISS cohorts was performed to investigate the genetics of carotid artery disease in ischaemic stroke. All study subjects were of self-reported European ancestry and were classified into mutually exclusive etiologic subtypes according to the Trial of Org 10172 in Acute Stroke Treatment (TOAST) (Adams et al. 1993). All subjects were collected prospectively under

protocols approved by the review boards at participating institutions, and all participants gave their written informed consent. Details of the cohorts are given below.

Bio-Repository of DNA in Stroke (BRAINS) cohort

The BRAINS study is an on-going case control study consisting of a UK European cohort (15 UK recruitment centres) and a South Asian cohort (UK South Asians as well as recruitment centres in India and Sri Lanka) (Yadav et al. 2011). Patients are eligible for inclusion if aged over 18 years, with a diagnosis of ischaemic or haemorrhagic stroke confirmed by CT or MRI imaging. Known monogenic causes of stroke are excluded. Upon patient consent, blood samples are obtained for DNA extraction, analysis and storage. A total of 344 genotyped ischaemic stroke patients from BRAINS were included in this study.

Ischemic Stroke Genetics Study (ISGS) cohort

The ISGS is an on-going, case control study of first-ever ischaemic stroke patients recruited from five centres in the US, designed to investigate candidate genes from the thrombosis and haemostasis cascade (Meschia et al. 2003). Ischaemic stroke patients are eligible if over 18 years of age and having symptoms within 30 days of enrolment, as diagnosed by a study neurologist and confirmed by CT or MRI brain imaging. The study excludes cases of post-operative or post-procedural stroke (e.g. carotid stenting, cerebral angiography), stroke related to recent subarachnoid haemorrhage or mechanical heart valve, patients with known inherited stroke syndromes (e.g. CADASIL, homocysteinaemia) and patients with biopsy-proven central nervous system vasculitis or stroke occurring in the setting of active bacterial endocarditis. Once eligibility is confirmed, blood samples are obtained for DNA extraction, analysis, and storage. A total of 272 genotyped ischaemic stroke patients from ISGS were included in this study.

Siblings with Ischemic Stroke Study (SWISS) cohort

SWISS is a family study recruiting probands from 49 centres across the US and Canada, designed to identify genetic risk factors through a genome-wide linkage screen in sibling pairs concordant and discordant for ischemic stroke (Meschia et al. 2006). Probands are adult men and women invited to participate in SWISS if they have a diagnosis of at least one ischaemic stroke confirmed by a study neurologist and CT or MRI imaging, and report having at least one living full sibling with a history of ischaemic stroke. Upon verification of concordance by medical record review, blood samples are obtained from proband, concordant sibling, and, if applicable, discordant sibling, for DNA extraction, analysis, and storage. Patients who are already enrolled in SWISS are not eligible for participation in ISGS. A total of 37 genotyped ischaemic stroke patients from SWISS were included in this study.

4.2.2.2 Evaluation of Carotid Stenosis

Each study cohort evaluated the degree of carotid artery stenosis using high-ultrasound B-mode ultrasonography of the common carotid arterty (CCA), external carotid artery (ECA) and internal carotid artery (ICA). An average stenosis value (%) was calculated from CCA, ECA and ICA stenosis measurements for each patient (as available) and used as the carotid stenosis phenotype. This was based upon the findings of a population-based cohort study, which reported that carotid plaques increase the risk of stroke and cerebral infarction irrespective of their location and with no clear difference in the risk estimates between ipsilateral or contralateral ischaemic strokes (Hollander et al. 2002). Patients with carotid stenosis ≥20% were deemed 'cases', while those with stenosis <20% were deemed controls. This threshold was chosen in concordance with a previous GWAS study of carotid artery disease (Bis et al. 2011), and allowed a relatively equal distribution of the study population between 'cases' and 'controls'.

4.2.2.3 Genotyping and Quality Control

Samples from the BRAINS cohort were genotyped on the Illumina HumanHap 610-quad Beadchip by Dr Sunaina Yadav (Imperial College London) and Miss Kerra Pearce (University College London), according to the manufacturer's protocol. Intensity data was initially examined using the genotyping module in Illumina Genome Studio software. Sample-dependent and sample-independent control probes were checked to investigate possible processing errors. Data for successfully genotyped samples was extracted and SNPs were clustered using the GENCALL algorithm within GENCALL Genome studio V2010.1 Genotyping module software, with plots manually checked for incorrect clustering. Samples from the ISGS/SWISS cohort were genotyped on Illumina HumanHap550K chips and Illumina Bead Studio software was used for the initial sample processing.

Quality control analysis was used performed in PLINK. Sample quality control was performed to remove individuals from BRAINS and ISGS/SWISS GWAS cohorts based upon the following filters: genotyping call rate ≤ 95%, non-European ancestry, outlying autosomal heterozygosity and cryptic relatedness (pi-hat ≥ 0.2). SNP quality control was also performed to remove SNPs with minor allele frequency less than 1% (MAF ≤ 0.01), if the SNP had more than 5% missing data (call frequency ≤ 95%) or if the Hardy-Weinberg p-value was ≥ 10^{-6}.

4.2.2.4 GWAS Imputation, Association Analysis and Meta-Analysis

GWAS analysis was performed by Dr Doug Speed (University College London), Dr Sunaina Yadav (Imperial College London) and Dr Ioana Cotlarciuc (Imperial College London).

Imputation Imputation was performed across the whole genome to allow combination of GWAS cohorts. Imputation aims to predict genotypes at SNPs not directly genotyped in a GWAS by assuming linkage disequilibrium (LD) with experimental SNPs. The experimental SNPs, in conjunction with predicted *in silico* SNPs, expand the number of SNPs that can be tested for association and hence increase the power of a study (Clark & Li 2007; Marchini & Howie 2010). Imputation was performed using IMPUTE version2 software and the 1000 Genomes 06_2010 imputation build. Annotation of the ISGS/SWISS cohort SNPs was based upon the human genome sequence hg18 build, and therefore updated to the hg19 build using the UCSC Genome Browser Liftover tool, for consistency with the BRAINS cohort prior to imputation. Imputed SNPs with correlation of $r^2 < 0.5$ with genotyped SNPs and MAF ≤ 0.01 were excluded.

Association Analysis: Logistic regression analysis was performed independently for each cohort using PLINK software (Purcell et al. 2007) to investigate the association of 5,094,295 genotyped and imputed SNPs with the presence (≥20%) or absence (<20%) of carotid stenosis in ischaemic stroke patients. An additive model was used, with age and gender used as covariates.

Meta-analysis: Results of the association analyses from each cohort were meta-analysed using PLINK software with a fixed effects inverse-variance model, weighted by standard error, to produce a final p-value for each SNP tested. Inter-study heterogeneity was assessed with an I^2 heterogeneity index and a p-value for Cochran's Q statistic.

4.2.2.5 SNP to Gene Annotation

Nicolae et al. (2010) have shown that SNPs associated with complex traits are significantly more likely to be eQTLs than MAF-matched SNPs from high-throughput GWAS platforms. However, the definition of a *cis*-eQTL is open to interpretation. Here, *cis*-eQTLs were defined within a 20kb wingspan, consistent with a gene expression study that showed that the majority of eQTLs lie within 20kb either side of a gene (Veyrieras et al. 2008).

To ensure consistent SNP-to-gene annotation between different analytical tools, independent software, LDAK (www.ldak.org) was used. The complete GWAS meta-analysis SNP list was coded into LDAK by Dr Doug Speed (University College London). Genes were downloaded from UCSC Genome Browser using the latest annotations (hg19), and SNPs were annotated to genes within a 20kb wingspan either side of gene transcription start and stop sites. As longer genes tend to contain more SNPs by chance alone, Bonferroni correction was applied to account for the fact that these longer genes would have smaller p-values. Genes passing a significance threshold of $p<0.01$ after Bonferroni correction were selected as the gene list for further investigation.

4.2.3 Functional Enrichment Analysis

While the Bonferonni method was used for multiple testing adjustment during SNP to gene annotation (Section 4.2.2.5) and protein-protein interaction analysis (Section 4.2.4), FDR correction was applied during functional enrichment analysis. FDR methods aim to reduce the probability of even one false discovery and are therefore less stringent than family-wise error rate procedures such as the Bonferroni method, which aims to reduce the expected proportion of false discoveries. Consequently, FDR methods lead to increased rates of type 1 errors but have greater power.

A number of FDR-based procedures have been proposed, including the Benjamini-Hochberg method (which relies on p-values being uniformly distributed under the null hypothesis; Benjamini & Hochberg 1995); the Benjamini–Hochberg–Yekutieli method (recommended when test statistics have positive regression dependency on each of the test statistics corresponding to the true null hypotheses; Benjamini & Yekutieli 2001); an empirical bayes method (Efron & Tibshirani 2002) and a q-value analog of the p-value method (Storey 2002). The Benjamini-Hochberg is the most widely used FDR-method as it is simple and robust under most circumstances (Kim & van de Wiel 2008). Therefore this method was chosen for multiple testing adjustment throughout the functional enrichment analysis**.**

4.2.3.1 Hypergeometric Testing

Singular Enrichment Analysis (SEA) for gene expression and GWAS data was performed with DAVID Bioinformatics Resources v6.7 (Huang da et al. 2009a; Huang da et al. 2009b). For SEA a subgroup of significantly different genes between two study populations, as selected using a user-defined significance threshold, are uploaded to DAVID. For each pathway or gene set (based upon databases including GO and KEGG), input genes that are part of the pathway or gene set are counted, and the process repeated for an appropriate background list of genes (e.g. all the genes in the genome or all the genes captured on a microarray). Every pathway is then tested for over-representation in the list of input genes.

Over-representation analysis in DAVID assumes a Gaussian hypergeometric probability distribution, which describes the number of successes in a series of n draws without replacement, from a finite population, N. A test against this distribution determines the probability that the number of successes observed could be achieved by chance alone, and thus determines the possibility that a certain GO term or KEGG pathway is over-represented in the inputted gene list, with respect to the background population. In the DAVID annotation system, the hypergeometric distribution is tested by a one-tailed Fisher's Exact test or a modified

variant, the Expression Analysis Systematic Explorer (EASE) score, to measure the gene-enrichment in annotation terms. When genes within two independent groups can fall into one of two mutually exclusive categories, Fisher's Exact test determines whether the proportions of those falling into each category differs by group, taking into account the effects of a finite population that are ignored by other statistical methods such as Chi-square (Hosack et al. 2003). The EASE score is a conservative adjustment to the Fisher's exact probability that weights significance in favour of terms containing more genes, and relies on jack-knifing the probability distribution so as to penalise categories supported by only a few genes. Thus the EASE score prioritises GO terms containing larger number of genes (Hosack et al. 2003).

In this study, DAVID was utilised for over-representation analysis of both gene expression and GWAS data. For analysis of gene expression data, genes passing nominal $p<0.05$ significance and fold difference >2 were uploaded to DAVID and compared to a background population comprising all genes captured on the Affymetrix Human Exon 1.0 ST array. For analysis of GWAS data, SNPs were first annotated to genes using LDAK at 20kb wingspan (as described in Section 4.2.2.5), uploaded to DAVID and compared to a background population comprising all genes in the genome.

4.2.3.2 Gene Set Enrichment Analysis (GSEA)

GSEA of Gene Expression in Partek

GSEA analysis of gene expression data was performed within Partek Genomics Suite v6.6 (Partek Inc., St. Louis, USA). Given pre-defined gene sets (e.g. GO annotations), GSEA determines whether the members of these defined gene sets are distributed randomly throughout an experimental list of genes, or primarily found at the top or bottom of the ranked list (i.e. substantially up- or down-regulated). As signal-to-noise ratio is larger for a set of genes, GSEA is statistically sensitive to small but consistent changes within a gene set, which may be biologically important (Clark & Ma'ayan 2011). Genes are first ranked according to their degree of differential expression between two study groups, then grouped into biological categories as determined by a reference database (e.g. GO). An enrichment score (ES) is calculated to reflect the degree to which these pre-defined gene sets are overrepresented and corresponds to a weighted Kolmogorov-Smirnov-like statistic. The ES is calculated by walking down the entire gene list and increasing a running sum statistic when a gene in the gene set is encountered and decreasing it when the gene is not in the gene set. The increment of increase or decrease is based upon the degree of correlation between the gene and the phenotype. The statistical significance of the ES is estimated by phenotype label permutation (preserving gene-gene correlations and the complex nature of gene expression data). The ES of the permutated data is calculated and the p-value of the experimental ES is calculated relative to this null distribution.

Correction of multiple hypothesis testing is performed by normalising the ES for each gene set to account for the size of the set and then applying an FDR correction (Subramanian et al. 2005). GSEA of gene expression data was performed on Partek using 1000 permutations and GO as a reference database.

GSEA of GWAS in ICSNPathway

GSEA analysis of GWAS data was performed using ICSNPathway (Zhang et al. 2011). ICSNPathway is a web-based tool that uses GWAS data to discover candidate causal SNPs and associated causal pathways. Unlike many platforms that either annotate SNPs to genes to infer causal SNPs or identify causal pathways by pathway analysis, ICSNPathway integrates LD analysis, functional SNP annotation and pathway analysis into one tool.

Firstly, SNPs in LD with the most significant SNPs obtained from GWAS are collated from the Hapmap CEU hg19 build to capture all possible candidate causal SNPs. Candidate causal SNPs are selected based upon a functional SNP that alter protein structure, gene expression or the role of a protein within a molecular pathway. These functional SNPs are defined by the Ensembl Variant Effect Predictor, and include deleterious and non-deleterious non-synonymous SNPs, SNPs leading to gain or loss of stop codon, SNPs resulting in frame shift, and SNPs in transcription factor binding sites, essential splice sites (the first two bp and last two bp of an intron) or regulatory regions.

ICSNPathway performs pathway analysis on the full list of GWAS-obtained SNPs based upon the improved gene set enrichment analysis (*i*-GSEA) method (Zhang et al. 2010). The Ensembl 61 database is used to map each SNP to a gene, with the maximum $-\log_{10}$(p-value) of the SNPs mapped to gene (t) used to represent the gene. All genes are ranked according to their t values and enrichment scores are calculated per pathway by Kolmogorov-Smirnov like running-sum statistic. These enrichment scores are converted to significant proportion based enrichment scores (SPESs) by multiplying by m_1/m_2, where m_1 is the proportion of significant genes (defined as genes mapped with at least one of the 5% most significant SNPs of GWAS) for a pathway, and m_2 is the proportion of significant genes for all the genes in the GWAS. Unlike traditional GSEA that conducts phenotype label permutation, *i*-GSEA implements SNP label permutation and normalisation to generate a distribution of SPES and to correct for gene variation (the bias caused by different genes having different numbers of mapped SNPs) and pathway variation (the bias caused by different pathways comprising different numbers of genes). Based upon SPES distributions generated by permutation, a nominal p-value is calculated and FDR correction performed for multiple testing. By implementing SNP label permutation instead of phenotype label permutation, and by focussing on pathways with high

proportions of significant genes (SPES) rather than relying on the total significance attributed to a few or many significant genes (ES), it has been shown that *i*-GSEA has better sensitivity than traditional GSEA for identifying pathways that represent combined effects of possibly modest SNPs (Zhang et al. 2010).

Here, significant SNPs were defined as those with $p<10^{-5}$, the parameter for searching for LD neighbourhood was restricted to $r^2>0.8$ at a distance of 200kb, and SNPs were annotated to genes within a 20kb wingspan. GO annotations were explored for pathway/gene set analysis, with the minimum number of genes in each pathway being set to 3 and FDR cut-off for pathway analysis set at $p<0.05$.

4.2.4 Protein-Protein Interaction Analysis

The Disease Association Protein-Protein Link Evaluator (DAPPLE; Broad Institute) was used to assess significant physical connectivity among proteins encoded for by genes in loci associated to symptomatic carotid stenosis (Rossin et al. 2011). The hypothesis underlying DAPPLE is that causal genetic variation affects a group of biological mechanism that are detectable by protein-protein interactions. Protein-protein interactions are defined by the InWeb database, which combines information from a number of sources including MINT, BIND, IntAct and KEGG and scores interactions based upon the scale of the experiment in which the interaction was reported and the number of publications in which the interaction had been cited (Lage et al. 2007).

'High confidence' InWeb protein-protein interactions are used to test whether direct and indirect networks built from seed regions are more connected than by chance alone, which is defined by the connectivity expected if connectivity were purely a function of the binding degree of participating proteins. The HapMap CEU hg18 build (www.hapmap.org) is used by DAPPLE for hotspot and LD information, with the wingspan of a SNP defined as the region containing SNPs with $r^2>0.5$ to the associated SNP, and this region being further extended to the nearest recombination hotspot. The hg18 gene list is downloaded from UCSC using Ensemble transcripts, and splice isoforms are collapsed to define the maximum gene footprint from transcription start site to transcription stop site, including flanking regulatory DNA.

DAPPLE builds direct networks from direct interactions between associated proteins according to the InWeb database, as well as indirect networks by considering connections between proteins associated via a protein elsewhere in the genome. Within-degree node-label permutation, comprising a permutation step and a switching step, is used to generate random

protein networks, then an empirical distribution is constructed for each network and used to evaluate the significance of disease networks. Individual proteins within a network are scored using the same permutation method and a subset of candidate proteins per locus is suggested to be harbouring causal variants, with these candidate proteins scores being Bonferroni corrected to adjust for the number of possible candidates within each locus.

For analysis of gene expression data obtained from the transcriptomics study, genes passing a nominal significance of $p<0.05$ and fold difference >2 were input to DAPPLE. Common interactor binding degree cut-off was set at 2, running 1000 permutations and iteration such that if any gene achieves a Bonferroni corrected score of $p<0.05$, that gene was prioritised and the analysis restarted.

4.3 Results

4.3.1 Analysis of Gene Expression Data

4.3.1.1 Sample Selection and Patient Characteristics

Eighteen patients undergoing carotid endarterectomy were selected for the study, of which 9 were symptomatic and 9 asymptomatic. From these 18 endarterectomy specimens, RNA was extracted from 44 samples (27 samples from stenosing plaque and 17 samples from non-stenosing tissue). Of these, 25 samples failed the threshold of RNA integrity required using Nanodrop 260/280nm ratio (accept >1.8) and Agilent Bioanalyser RNA Integrity Number (RIN; accept >3.0), hence only 19 were analysed for global gene expression using the Affymetrix Human Exon 1.0 ST array. Two of these 19 samples failed chip hybridisation control, hence 17 samples were available for analysis, relating to 8 stenosing plaque and 9 non-stenosing tissue samples, taken from 10 patients (7 symptomatic, 3 asymptomatic) (Figure 4.2). Mean age of the population was 67 years (sd = 11.6 years) and mean % carotid stenosis was 72% (sd = 24.2%). The majority of patients were male (80%) and were on statin (70%) and antiplatelet medication (70%). Patient characteristics are detailed in Table 4.2. Power calculations performed on Partek Genomics Suite software found that a sample size of 10 (5 in each analysis arm) delivered 80% power at $p<0.05$ to detect a 2-fold difference in means.

4.3.1.2 Quality Control Analysis

Following RMA normalisation, quality control (QC) parameters were used to investigate reproducibility and consistency of signals across arrays, and to identify any outlier arrays that differed substantially from the majority of the cohort. The following metrics were used to examine sample and hybridisation quality:

Area Under the Curve (AUC): The AUC value is taken from a ROC curve, generated by evaluating how well the probe set signals separate the positive controls from the negative controls with the assumption that the negative controls are a measure of false positives and the positive controls are a measure of true positives. This provides a measurement for overall data quality, with an AUC of 1 representing perfect separation and an AUC value of 0.5 representing no difference between positive and negative controls. The mean score from 17 arrays in this study was 0.82.

Probeset Summarisation Values: The mean signal of all probesets should be consistent between arrays. The mean Probeset Mean in this study cohort was 5.67 (SD=0.14). The mean absolute relative log expression compares probeset signals against the median signal, with higher values

indicating that the signal is different from the others. The mean of the absolute deviation of the residuals from the median (X_mad_residual_mean) produces a measure of how well or poorly all of the probes on a given chip fit the RMA normalisation model, with the difference between the actual value and the predicted value termed the residual. All these metrics provide measures of the overall quality of data that will be used for downstream analysis.

Hybridisation and Labelling Controls: Bacterial spikes are the set of probesets which hybridise to the pre-labelled bacterial spike controls (as BioB, BioC, BioD and Cre). This category is useful in identifying problems with the hybridization and/or chip. PolyA spikes are the set of polyadenylated RNA spikes (Lys, Phe, Thr, and Dap), and can be used to identify problems with the target preparation phase.

Arrays were considered outliers if the metric value was >10% above or below the mean value for the whole cohort. None of the arrays failed on more than two metrics, therefore all arrays were considered in the downstream analysis (Figure 4.3).

Table 4.2: Gene expression study patient characteristics. Demographic, clinical and histological information relating to 10 patients from whom SP and NST carotid endarterectomy tissue samples were used for gene expression analysis using Affymetrix Human Exon 1.0 ST array. **AIS,** acute ischaemic stroke; **IHD**, ischaemic heart disease; **NST**, non-stenosing tissue; **PAD**, peripheral artery disease; **SP**, stenosing plaque, **TIA,** transient ischaemic attack.

Patient ID	Tissue Type	Symptoms	Carotid Stenosis (%)	Plaque Histology	Gender	Age	IHD	PAD	Hypertension	Diabetes Mellitus	Smoking History	Statin Use	Antiplatelet Use
V0012	SP + NST	TIA	90-95	Smooth	F	76	x	x	✓	x	✓	✓	✓
V1236	SP + NST	-	85	Smooth	M	85	x	✓	✓	x	✓	✓	✓
V1299	SP + NST	TIA	90-95	Irregular	M	79	x	x	✓	x	x	✓	x
V1376	SP + NST	AIS + AmFu	90	Smooth	M	60	✓	x	x	✓	✓	✓	✓
V1377	NST	AIS	30	Irregular	M	73	x	x	✓	✓	✓	✓	✓
V1388	NST	-	80	Smooth	F	66	x	x	✓	x	x	✓	x
V1420	SP	AIS	55-65	Irregular	M	51	x	x	✓	x	✓	x	x
V1465	SP + NST	-	75-80	Smooth	M	67	x	x	✓	✓	x	x	✓
V1468	SP + NST	AIS	30	Smooth	M	55	x	x	x	x	x	x	✓
V1477	SP + NST	AmFu	60-75	Irregular	M	54	x	x	✓	x	x	✓	✓

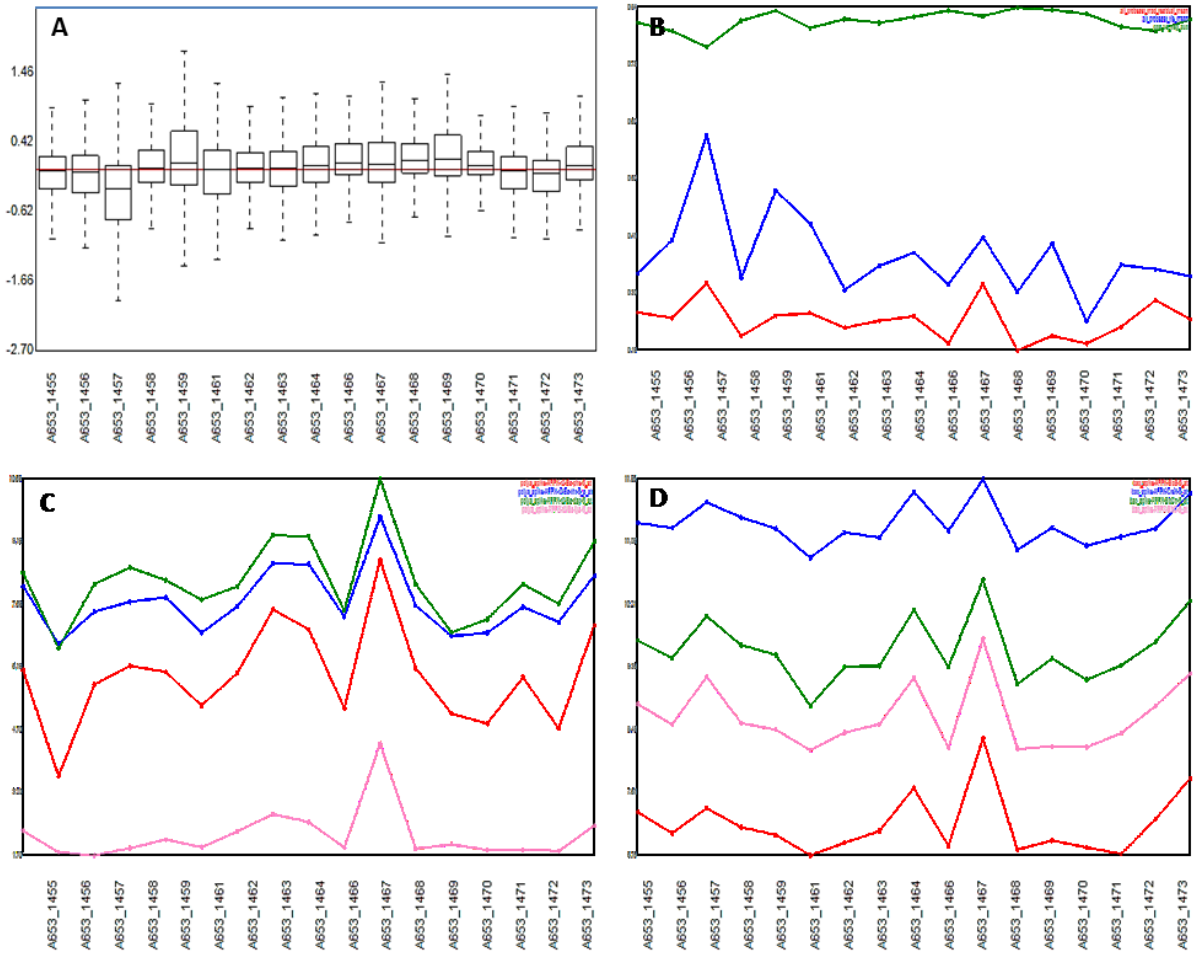


Figure 4.3: Plotted QC metrics to assess array data quality. A: Box plot of the relative log expression signal for all probesets analysed. Deviations from a median value of 0 (e.g. A653_1457, corresponding to patient V1377 NST sample) could indicate skew in the raw intensities for the array that cannot be corrected by RMA normalisation. **B:** Mean absolute deviation of residuals (red), mean absolute relative log expression (blue) and the area under the receiver operator curve for signal discrimination of the positive and negative controls (AUC; green) suggest that array A653_1457 may be an outlier. **C:** PolyA spikes can identify potential problems with the target preparation phase of the experiment. Here, the rank order of the polyA spikes is as expected (lys [pink] <phe [red] <thr [blue] <dap [green]). **D:** Bacterial spikes can identify potential problems with the hybridisation and array chip. The rank order of the bacterial spikes is as expected (BioB [red] < BioC [pink] < BioD [green] < Cre [blue]). Array chip A653_1467 (corresponding to patient V1420 SP sample) shows a compromised signal value with both PolyA and Bacterial Spikes, suggesting hybridisation or target preparation errors. Nonetheless, neither array A653_1457 nor A653_1467 failed (>10% deviation from mean) on more than two metrics, and therefore were not excluded from downstream analysis.

4.3.1.3 Unsupervised Analysis

PCA based on global gene expression demonstrated distinct classification based upon tissue type (non-stenosing tissue (NST) or stenosing plaque (SP)), which was a stronger determining factor of gene expression than any other demographic variable. Furthermore, little differentiation was observed between NST samples from symptomatic and asymptomatic individuals, while SP samples from asymptomatic and symptomatic individuals clustered separately. One outlier (corresponding to patient V1377 NST sample) was identified and removed from further analysis (Figure 4.4).

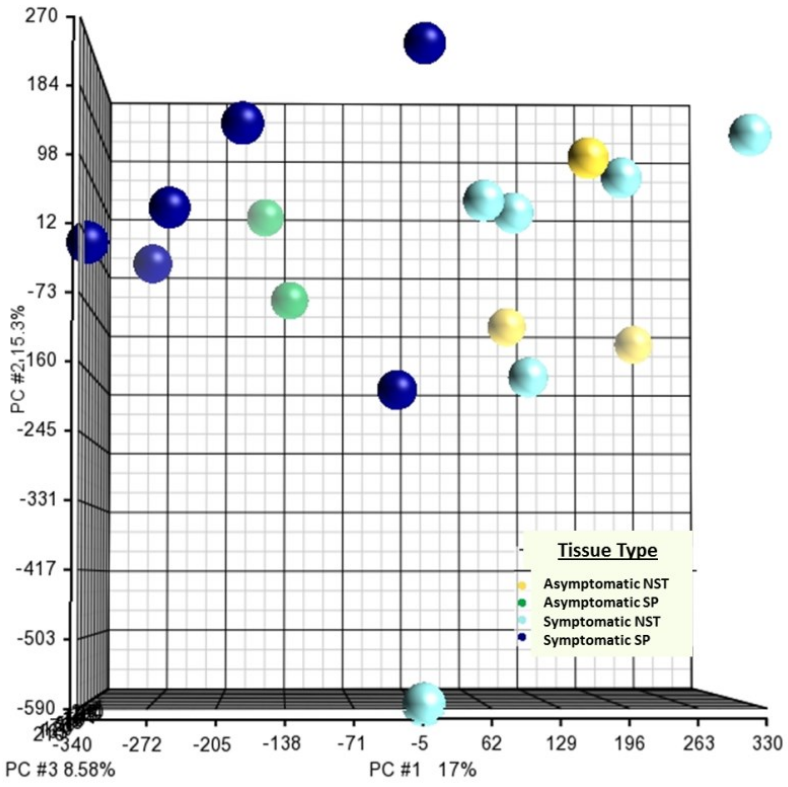


Figure 4.4: Unsupervised analysis of gene expression data by PCA. PCA of the gene expression profiling study (comprising 9 NST and 8 SP carotid endarterectomy tissue samples from 7 symptomatic and 3 asymptomatic individuals) demonstrated a natural ability of global gene expression to separate individuals by tissue type (NST or SP). Furthermore, asymptomatic and symptomatic SP samples were seen to cluster separately. The same analysis performed all other demographic variables did not demonstrate this ability for classification, indicating that global gene expression is useful for tissue and patient classification, independent of other confounding variables. One outlier was identified, corresponding to patient V1377 NST sample. This sample had also been found to be of questionable quality during quality control procedures, and was therefore excluded from downstream analysis. **NST**, non-stenosing tissue; **SP**, stenosing plaque.

4.3.1.4 Supervised Inter-Plaque and Intra-Plaque Analysis

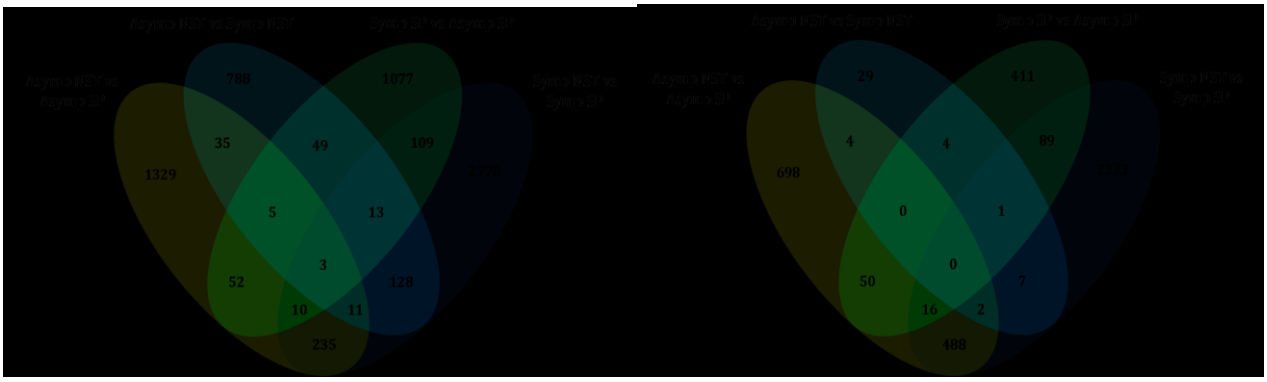
Gene expression analysis was performed based upon the Genbank RNA transcript 'core' gene set of 21,980 genes. Examination of demographic variables found that age, gender and statin use accounted for substantial sources of variation (F-ratios of 2.06, 2.01 and 2.46 respectively; F-ratio of tissue type of 3.78). Therefore these were included as covariates in the ANOVA model.

Differential expression of 2,988 genes (13.5%) was found between all NST (n=9) and SP (n=8) tissue samples ($p<0.05$). None of these genes were found to reach significance of $p<0.05$ after FDR correction. However, 564 genes were differentially expressed at a fold difference >2 (341 up-regulated in SP and 223 down-regulated). Alternative splicing was evident in 17,060 genes (uncorrected $p<0.05$), of which 3121 were found to be alternatively spliced at FDR $p<0.05$. A total of 1620 genes were found to be differentially expressed but not alternatively spliced and 911 showed alternative splicing without being differentially expressed (the 'cassette exon' use case).

Expression and alternative splicing profiles were also compared between NST and SP tissue segments of asymptomatic and symptomatic patients, with SP segments of symptomatic patients deemed vulnerable and unstable (as confirmed by clinical experts and histological examination), and SP segments of asymptomatic patients deemed stable. Four analyses were performed:

Asymptomatic NST (n=3) vs Asymptomatic SP (n=2)
Symptomatic NST (n=5) vs Symptomatic SP (n=6)
Asymptomatic NST (n=3) vs Symptomatic NST (n=5)
Asymptomatic SP (n=2) vs Symptomatic SP (n=6)

The results of these analyses are summarised in Figure 4.5. It was found that tissue type (NST or SP) was a stronger determinant of differential gene expression and alternative splicing than whether a patient was symptomatic or asymptomatic. Nonetheless, the number of differentially expressed genes was more than double between symptomatic NST and SP as compared to asymptomatic NST and SP (2778 genes as compared to 1329 genes, at uncorrected $p<0.05$, fold difference >2). Furthermore, the degree of alternative splicing observed between symptomatic NST and SP samples (2373 genes, FDR p-value<0.05) was substantially higher than in any other analysis.



Gene Expression **Alternative Splicing**

Figure 4.5: Venn diagrams depicting gene expression and alternative splicing profiles in various phenotypes. Independent analyses were performed to explore the effects of tissue type (NST or SP) and symptomatic status on global gene expression, using ANOVA models adjusted for age and gender. Little overlap was noted between genes from the four analyses, indicating the presence of unique pathological processes. Tissue type (NST or SP) was a greater determinant of gene expression and alternative splicing differences than symptomatic status. Interestingly, analysis of differences between symptomatic SP and NST yielded the greatest level of differential gene expression and a marked increase in alternative splicing, suggesting that alternative splicing of genes, rather than gene expression itself, could play a key role in a symptomatic plaque phenotype. For gene expression analysis, genes passing a nominal significance threshold of $p<0.05$, and fold difference >2 are shown. For alternative splicing analysis, genes passing an FDR corrected significance of $p<0.05$ are shown. Asymptomatic NST: n=3, asymptomatic SP: n=2, symptomatic NST: n=5, symptomatic SP: n=6. **NST**, non-stenosing tissue; **SP**, stenosing plaque.

4.3.1.5 Paired Symptomatic NST vs SP Analysis

For comparison with GWAS data, gene expression profiles of paired NST and SP CEA samples from 5 symptomatic individuals (V0012, V1299, V1376, V1468, V1477) were explored. Probes were first interrogated with SNPinProbe1.0 database and Pip Finder tool to remove 17,552 probesets (6.3%) containing known SNPs that could affect probe binding affinity to the microarray, leaving 261,033 probesets for analysis. Differential gene expression and alternative splicing between paired NST and SP were analysed using repeated measures ANOVA, with age, gender and statin use incorporated into the model as covariates.

Differential expression of 2,723 genes (12.4%) was found between paired symptomatic NST and SP tissue samples ($p<0.05$). None of these genes were found to reach significance of $p<0.05$ after FDR correction. However, 368 genes were differentially expressed at a fold difference >2 (162 up-regulated in SP and 206 down-regulated). Furthermore, alternative splicing was evident in 4,573 genes (uncorrected $p<0.05$), of which 2,761 were found to be alternatively spliced at FDR $p<0.05$. A total of 55 genes were found to be differentially expressed but not alternatively spliced and 1714 showed alternative splicing without being differentially expressed. These findings verified earlier indications that alternative splicing of genes, rather than gene expression itself, could play a key role in plaque progression and a symptomatic plaque phenotype.

4.3.2 Analysis of GWAS Data

4.3.2.1 Patient Characteristics

A discovery meta-analysis of two GWAS was performed to investigate the genetics of carotid artery stenosis in ischaemic stroke patients. GWAS meta-analysis comprised 653 patients from the BRAINS (n=344) and ISGS/SWISS cohorts (n=309). Patient demographics are given in Table 4.3. Patients with carotid artery stenosis ≥20% were deemed 'cases' (n=369), and patients with carotid artery stenosis <20% were deemed 'controls' (n=284).

4.3.2.2 Quality Control Analysis

SNP imputation was performed to allow meta-analysis of BRAINS and ISGS/SWISS genotyped data, using IMPUTE2 version2 software and the 1000 Genomes 06_2010 imputation build as a reference haplotype. Logistic regression analysis was performed on the resulting 5,094,295 SNPs independently for both cohorts, with a cross-sectional study design to model the genetic effects on risk for the dichotomous outcome of presence (≥20%) or absence (<20%) of carotid stenosis in an ischaemic stroke population. An additive model was employed to relate genotype dose to carotid stenosis, using age and gender as covariates. Meta-analysis of the association results from each cohort was performed using a fixed-effects inverse variance weighted model.

Probability value quantile-quantile (Q-Q) plots were derived before and after meta-analysis for visual inspection of the quality of GWAS data (Figure 4.6). A Q-Q plot allows interrogation of possible confounding factors (e.g. unrecognised population structure, microarray platforms, genotyping artefacts) affecting GWAS results, by plotting the expected null χ^2 distribution of association test statistics (x-axis) against the observed values (y-axis). Deviation from the central line implies a consistent difference across the genome that may be due to a confounding factor that warrants closer inspection. General concordance between observed and expected values, with sharp deviation from the central line at high p-values suggests a small number of true associations that are more significant than expected under the null hypothesis. A Q-Q plot of the genome-wide distribution of meta-analysis association results demonstrated substantial departure from the null hypothesis of no association. Furthermore, a genomic inflation factor (GIF; λ) value of 0.97 was obtained, indicating low population stratification and therefore no necessity to adjust the logistic regression model for other confounding variables or principal components.

Table 4.3: GWAS cohort patient characteristics. Demographic and clinical information relating to the 653 ischaemic stroke patients comprising the GWAS meta-analysis. Patients with carotid artery stenosis ≥20% were deemed 'cases' (n=369), and patients with carotid artery stenosis <20% were deemed 'controls' (n=284). **BRAINS**, Bio-Repository of DNA in Stroke; **ISGS**, Ischaemic Stroke Genetics Study; **SWISS,** Siblings With Ischemic Stroke Study.

	BRAINS (n=344)	**ISGS (n=272)**	**SWISS (n=37)**
Mean Age (± sd)	71.4 (14.0)	66.0 (14.4)	69.0 (9.9)
Gender	151F, 193M	164F, 109M	20F, 17M
Carotid Stenosis (≥ /< 20%)			
≥20% n (%)	165 (48.0%)	204 (66.0%)	
< 20% n (%)	179 (52.0%)	105 (34.0%)	
Ischaemic Stroke Subytpe (TOAST)			
Cardioembolic n	79	27	19
Small Vessel Disease n	30	15	19
Large Vessel Disease n	42	34	24
Hypertension (%)	71%	64%	73%
Smoking (%)	21%	68%	11%
Diabetes (%)	14%	21%	24%
Hypercholesterolaemia (%)	44%	48%	59%

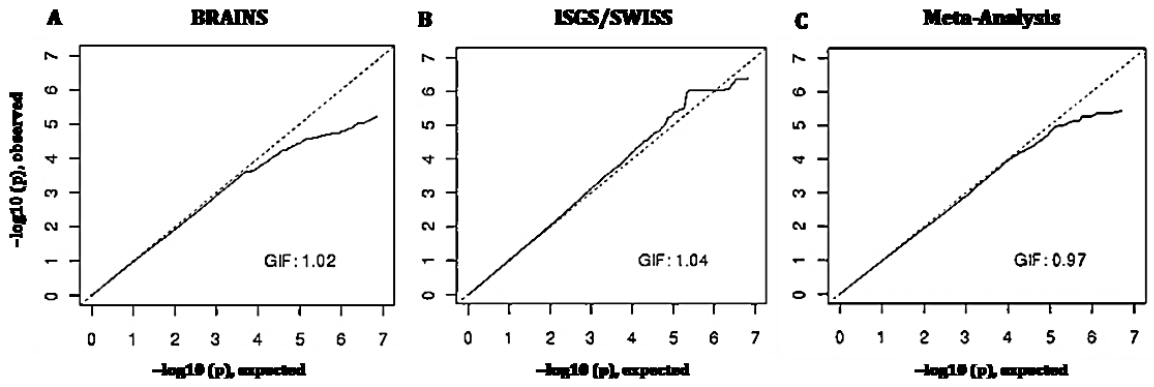


Figure 4.6: Quantile-Quantile (Q-Q) plots of GWAS association results obtained from A) BRAINS cohort B) ISGS/SWISS cohort and C) BRAINS and ISGS/SWISS meta-analysis. Observed association significance is plotted against the expected association significance for each SNP, with strong deviation away from the central line indicating substantial departure from the null hypothesis of no association. Genomic inflation factors (GIF; λ) of ~1.00 were obtained for all three analyses, indicating low population stratification and therefore no necessity to adjust the logistic regression model for other confounding variables or principal components.

4.3.2.3 Genetic Loci Associated with Carotid Stenosis in the Presence of Ischaemic Stroke

No single SNP (either genotyped or imputed) demonstrated a genome-wide significant association with carotid stenosis (≥20%) in the presence of ischaemic stroke, with genome-wide significance being defined as $p<5x10^{-8}$ relating to a Bonferroni correction of $p<0.05$ for approximately 1 million independent tests conducted. Nevertheless, 462 SNPs approached suggestive significance of $p<1x10^{-5}$, and a total of 4039 SNPs reached significance of $p<1x10^{-3}$.

Candidate genomic regions of interest were observed on chromosomes 4, 11 and 20, as illustrated in the Manhattan plot in Figure 4.7 These regions corresponded to SNPs residing in or near to the RasGEF domain family member 1B (*RASGEF1B*) gene, olfactory receptor family 9 subfamily G member 1 (*OR9G1*) gene and kinesin family member 16B (*KIF16B*) gene.

A cluster of 77 SNPs with $p<1x10^{-5}$ were found around the *RASGEF1B* locus (4q21.21), a guanine nucleotide exchange factor with specificity for Ras family proteins. In particular, 13 SNPs with $p<1x10^{-6}$ were identified spanning a 20kb region, in high LD ($r^2>0.95$) with each other and the sentinel SNP rs592145. The SNPs lie 315kb (rs28528190) to 335kb (rs500213) downstream of *RASGEF1B* and were all found to show the same direction of association, with effect sizes ranging from OR 0.509 (95% CI 0.45-0.80, rs592145) to OR 0.524 (95% CI 0.49-0.81, rs611140), as detailed in Table 4.4. Effect sizes in the same direction were almost a necessity of significance in this case, with only two studies in the meta-anaysis. All SNPs exhibited moderate heterogeneity in the meta-analysis (I^2=51-61%).

Seven SNPs with $p<1x10^{-6}$ were identified near the olfactory receptor loci on chromosome 11, a family of G-protein-coupled olfactory receptors. The SNPs demonstrated high LD with each other and the sentinel SNP rs584144 ($r^2>0.95$), and spanned a 14kb region downstream of the *OR9G1* locus (11q11). The SNPs lie 27kb (rs548921) to 42kb (rs513873) downstream of the *OR9G1* gene and demonstrated the same direction of association with OR 0.55, although the SNPs exhibited moderate heterogeneity in the meta-analysis (I^2 ~72%) (Table 4.4). A previous GWAS identified an association between *OR9G1* and hemostatic factors and hematological phenotypes in the Framingham Heart study (Yang et al. 2007). However, investigation of LD between the reported SNP (rs1397048) and the sentinel SNP in this study (rs584144) demonstrated minimal correlation ($r^2>0.059$).

A cluster of five SNPs with $p<1x10^{-6}$ was also identified inside an intron of the *KIF16B* gene (20p11.23), a kinesin-like microtubule dependent motor protein involved in intracellular trafficking. The SNPs spanned a 1.2kb region and demonstrated high LD with each other and the sentinel SNP rs12151916 ($r^2>0.88$) and exhibited the same direction of association, with effect sizes ranging from OR 1.85 (95% CI 1.42-2.37, rs12151916) to OR 1.83 (95% CI 1.38-2.26, rs603444) and low evidence of heterogeneity in the meta-analysis ($I^2<25\%$) (Table 4.4).

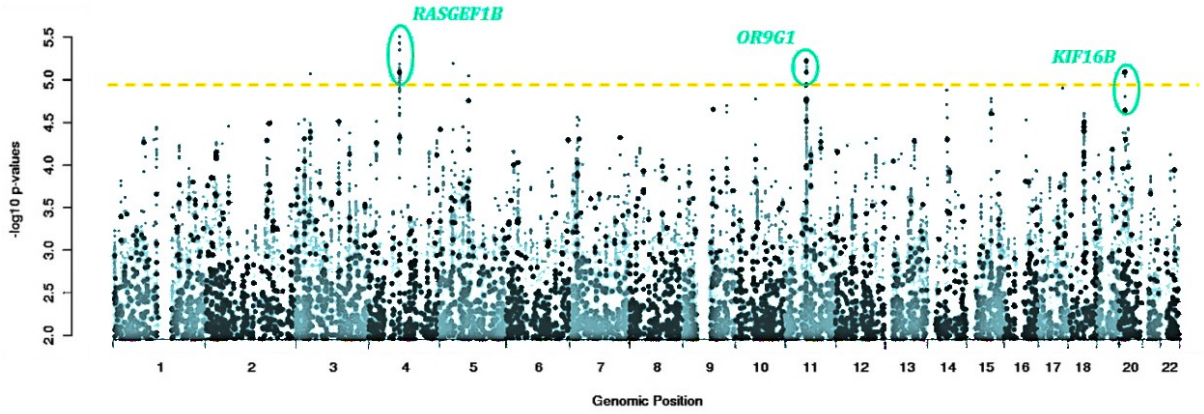


Figure 4.7: Manhattan plot of genome-wide association signals from meta-analysis of GWAS of BRAINS and ISGS/SWISS ischaemic stroke cohorts. Chromosomes are laid down end-to-end along the x-axis and physical position of each association test plotted against the $-\log_{10}$ of the significance p-value. No SNPs passed the genome-wide significance level of $p=5\times10^{-8}$, however the yellow line represents suggestive significance at genome-wide significance level of $p=1\times10^{-5}$. Three genetic loci in or near *RASGEF1B*, *OR9G1* and *KIF16B* genes at significance $p<1\times10^{-5}$ were identified on chromosome 4, 11 and 20 respectively.

Table 4.4: GWAS association results of carotid stenosis ≥20% in ischaemic stroke patients, after meta-analysis of BRAINS and ISGS/SWISS cohorts. Clusters of SNPs with $p<1x10^{-6}$ were identified in three genetic loci corresponding to RASGEF1B, OR9G1 and KIF16B genes. Sentinal SNPs are identified in bold, with odds ratios (OR) describing increase or decrease per copy of the risk allele (A1) as compared to the non-risk allele (A2). **I^2,** I^2 measure of heterogeneity; **KIF16B**, kinesin family member 16B; **LD**, linkage disequilibrium; **OR**, odds ratio; **OR9G1**, olfactory receptor family 9 subfamily G member 1; **Q**, Cochran's Q measure of heterogeneity; **RASGEF1B,** RasGEF domain family member 1B; **SNP**, single nucleotide polymorphism.

Gene	SNP	Chromosome	Position	A1/A2	OR (95% CI)	p-value	Q	I^2 (%)	LD with sentinel SNP, r^2 (D')
RASGEF1B	**rs592145**	**4**	**82726387**	**G/A**	**0.509 (0.45-0.80)**	**3.72×10^{-6}**	**0.14**	**54.9**	**1.00 (1.00)**
	rs669481	4	82727177	T/A	0.513 (0.40-0.73)	4.42×10^{-6}	0.15	51.6	0.95 (1.00)
	rs656129	4	82727881	G/A	0.513 (0.40-0.73)	4.44×10^{-6}	0.15	51.4	0.95 (1.00)
	rs653923	4	82728376	T/G	0.513 (0.40-0.73)	4.44×10^{-6}	0.15	51.4	0.95 (1.00)
	rs509133	4	82724731	T/G	0.516 (0.41-0.73)	5.30×10^{-6}	0.14	54.3	0.95 (1.00)
	rs593528	4	82726035	T/A	0.516 (0.41-0.73)	5.32×10^{-6}	0.14	54.2	0.95 (1.00)
	rs669908	4	82727097	G/A	0.516 (0.41-0.73)	5.32×10^{-6}	0.14	54.2	0.95 (1.00)
	rs555028	4	82727198	C/T	0.516 (0.41-0.73)	5.35×10^{-6}	0.14	54.1	0.95 (1.00)
	rs500213	4	82728569	C/T	0.516 (0.41-0.73)	5.37×10^{-6}	0.14	53.9	0.95 (1.00)
	rs538307	4	82723874	G/A	0.523 (0.44-0.79)	7.83×10^{-6}	0.12	59.2	0.95 (1.00)
	rs510654	4	82708507	A/G	0.523 (0.45-0.78)	8.91×10^{-6}	0.11	60.2	0.95 (1.00)
	rs28528190	4	82707806	G/A	0.522 (0.43-0.76)	9.14×10^{-6}	0.11	61.3	0.95 (1.00)
	rs611140	4	82708464	T/C	0.525 (0.49-0.81)	9.81×10^{-6}	0.11	61.2	0.95 (1.00)
OR9G1	**rs584144**	**11**	**56503728**	**C/T**	**0.552 (1.74-2.25)**	**7.18×10^{-6}**	**0.06**	**72.8**	**1.00 (1.00)**
	rs645359	11	56508191	C/T	0.553 (1.76-2.21)	$7.43E \times 10^{-6}$	0.06	72.5	1.00 (1.00)
	rs504021	11	56496947	G/A	0.553 (1.75-2.27)	7.51×10^{-6}	0.06	72.5	1.00 (1.00)
	rs483798	11	56497255	C/T	0.553 (1.76-2.26)	7.51×10^{-6}	0.06	72.5	1.00 (1.00)
	rs548921	11	56496107	G/A	0.554 (1.79-2.28)	8.44×10^{-6}	0.06	72.0	1.00 (1.00)
	rs679888	11	56496757	G/A	0.555 (1.81-2.29)	9.34×10^{-6}	0.06	71.2	0.96 (1.00)
	rs513873	11	56510623	A/G	0.557 (1.84-2.31)	9.81×10^{-6}	0.05	73.8	1.00 (1.00)
KIF16B	**rs12151916**	**20**	**16270957**	**C/G**	**1.8491 (1.42-2.37)**	**7.87×10^{-6}**	**0.25**	**25.0**	**1.00 (1.00)**
	rs6034443	20	16271601	C/T	1.8294 (1.38-2.26)	9.75×10^{-6}	0.28	12.9	0.88 (0.96)
	rs6043872	20	16272136	G/C	1.8294 (1.38-2.26)	9.75×10^{-6}	0.28	12.9	0.88 (0.96)
	rs6034445	20	16272189	A/G	1.8294 (1.39-2.28)	9.75×10^{-6}	0.28	12.9	0.88 (0.96)
	rs6043869	20	16271690	C/T	1.8367 (1.42-2.34)	9.89×10^{-6}	0.27	18.9	0.88 (0.96)

4.3.2.4 SNP Annotation and *Cis*-eQTL Mapping

SNPs obtained from GWAS meta-analysis following imputation and meta-analysis were annotated to genes within a 20kb wingspan using LDAK software.

A total of 21,911 genes were identified by LDAK, of which 207 achieved a Bonferroni adjusted significance threshold of $p<0.05$. Bonferroni correction was applied to account for the fact that longer genes are likely to contain more SNPs. The SNP and CNV Annotation Database (SCAN; Gamazon et al. 2010) was used to identify eQTL annotations. Annotation confirmation of SNPs reaching a suggestive genome-wide significance of $p<1x10^{-5}$ by SCAN analysis identified 20 loci with SNPs inside genes that achieved Bonferroni adjusted $p<0.05$ in LDAK: *CSF1R*, *ENTPD6*, *FAM5B*, *FMOD*, *GRB14*, *KIF16B*, *LTBP4*, *MAP4K5*, *MEGF11*, *MPND*, *MRGPRD*, *MYO3B*, *OR9G4*, *OSBPL3*, *PTGDR*, *SAV1*, *SIGLEC15*, *SLC25A24*, *SLC6A13*, *UGT3A1*. A further 28 genes were found within a 20kb region of annotated SNPs and annotated as expression genes associated with eQTLs.

4.3.2.5 Comparison of GWAS Results with Previously Reported Associations

A comparative search was performed with previously reported GWAS hits using the gene list obtained from LDAK. Six annotated genes from SNPs reaching significance $p<1x10^{-5}$ from GWAS meta-analysis were reported in GWAS of CVD and metabolic disorders, of which 2 had SNPs within the gene (*GRB14, MYO3B*) and 3 were flanking genes within a 20kb region of the SNP (*AGR2*, *KIF13B*, *SLC14A2*). These genes had been previously reported in disorders including type 2 diabetes, obesity-related traits and response to statin therapy. None of the SNPs identified through the search were found to be the same as those from the current GWAS analysis (Table 4.5).

Table 4.5: Comparison of GWAS results with previously reported associations. Genes found significant in the experimental GWAS of carotid stenosis through LDAK annotations were searched in the literature for reported associations from GWAS of CVD and metabolic disorders. Two genes (*GRB14* and *MYO3B*), with SNPs within genes, were found to be previously reported in GWAS of type 2 diabetes, blood pressure and obesity-related traits. Further three flanking genes (*AGR2, KIF13B* and *SLC14A2*) within 20kb of the sentinel SNP were found to be associated in previous GWAS of type 2 diabetes, response to statin therapy and metabolic levels. However, none of the SNPs within these genes were common between the experimental GWAS and literature, often demonstrated different directions of association, and exhibited weak linkage disequilibrium with the experimental SNP ($r^2<0.4$). **NR**, not reported.

Gene	Region	Experimental				Literature					
		SNP	Risk Allele	p-value	OR (95% CI)	Citation	Disorder	SNP	Risk Allele	p-value	OR (95% CI)
GRB14	2q24.3	rs11895467	T	6.21×10^{-5}	0.6195 (0.42-0.68)	Kooner et al 2011	Type 2 diabetes	rs3923113	A	1×10^{-8}	1.09 (1.06-1.13)
						Kato et al 2011	Blood pressure	rs16849225	C	4×10^{-11}	0.75 (0.53-0.97)
MYO3B	2q31.1	rs2883655	T	3.92×10^{-5}	1.6767 (1.32-2.13)	Comuzzie et al 2012	Obesity-related traits	rs7561268	C	2×10^{-6}	0.03 (NR)
AGR2	7p21	rs17136605	C	9.14×10^{-5}	3.2913 (2.58-4.18)	Pasquale et al 2013	Type 2 diabetes	rs1525739	?	6×10^{-6}	0.44 (NR)
KIF13B	8p12	rs491687	C	7.63×10^{-5}	1.6722 (1.31-2.12)	Barber et al 2010	Response to statin therapy	rs10091038	A	3×10^{-8}	NR
SLC14A2	18q12.3	rs12965695	T	9.57×10^{-5}	2.2873 (1.80-2.92)	Kim et al 2011	Metabolite levels	rs7232775	C	8×10^{-14}	0.0148 [0.011-0.019]

4.3.3 Functional Enrichment Analysis

Gene enrichment analysis of Gene Ontology (GO) terms and Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathways were performed to identify genes of pathophysiological importance to symptomatic plaque progression.

4.3.3.1 Hypergeometric Testing

GWAS

A total of 451 genes passing a Bonferroni-adjusted significance threshold of $p<0.01$ after LDAK annotation were inputted to DAVID for hypergeometric over-representation testing.

Of the databases searched in DAVID (GO, KEGG, Panther, OMIM, DB Disease), no functional annotation terms reached FDR corrected significance of $p<0.05$. A total of 125 GO terms were identified, of which 99 were biological processes (BP), 7 were cellular components (CC) and 19 were molecular functions (MF). Eleven of these passed a suggestive uncorrected significance threshold of $p<0.01$: 6 BP terms (GO:0045621, GO:0045591, GO:0045582, GO:0045589, GO:0045619, GO:0006644); 1 CC term (GO:0005576) and 4 MF terms (GO:0003723, GO:0003836, GO:0051059, GO:0008374). The majority of terms were linked to immune response and regulation of T cell and lymphocyte differentiation in particular (Table 4.6).

Of only two disease terms identified, one DB Disease class, 'Cardiovascular' reached uncorrected $p=0.005$, and one OMIM disease term, 'C4 Deficiency' reached uncorrected $p=0.031$ (consistent with the chromosome region *6p21.3*, previously implicated in GWAS of metabolic disorders; Comuzzie et al. 2012; Fox et al. 2012). This result lends evidence to support the reliability of the analysis, despite no terms reaching FDR significance of $p<0.05$.

Gene Expression Study

From paired analysis of differential gene expression in NST vs SP samples, a total of 368 genes passing FDR correction of $p<0.05$ were inputted to DAVID. Of the 308 GO terms identified, 117 reached a suggestive uncorrected significance threshold of $p<0.01$, and 22 passed FDR corrected significance threshold of $p<0.05$ (11 BP terms, 8 CC terms and 3 MF terms). These terms could be broadly separated into four functional categories: immune response (GO:0006955, GO:0009611, GO:0006954), adhesion (GO:0007155, GO:0022610, GO:0031589), cytoskeletal organisation (GO:0040012, GO:0030334, GO:0007010, GO:0016010, GO:0008092) and extracellular matrix organisation (GO:0007160, GO:0044421, GO:0005576, GO:0031012, GO:0005615, GO:0005578, GO:0001558,) (Table 4.5).

Table 4.6: Functional Enrichment analysis by hypergeometric testing of gene expression and GWAS data. Hypergeometric testing was performed using DAVID software (Huang da et al. 2009a; Huang da et al. 2009b) to explore the enrichment of Gene Ontology (GO) terms in gene expression and GWAS data. Enriched GO terms are shown for those passing significance of FDR p<0.01 in gene expression data, and those passing a nominal significance of p<0.01 in GWAS data (as no GO terms passed FDR corrected p-value). Enriched GO terms were those involved in immune response, protein binding, cytoskeleton organisation and extracellular matrix organisation. One term, GO:0005576~extracellular region, was significant in both gene expression and GWAS data, and contained three common genes: *LTBP4, FMOD* and *ILR7*. The *LTBP4* and *FMOD* genes were also present in a number of other GO terms, highlighted in blue.

GWAS			Gene Expression Study		
GO Term	**p-value***	**FDR p-value**	**GO Term**	**p-value***	**FDR p-value**
GO:0045621~positive regulation of lymphocyte differentiation	4.7E-04	5.3×10^{-1}	GO:0006955~immune response	4.8×10^{-8}	8.3×10^{-5}
GO:0003723~RNA binding	4.9E-04	2.0×10^{-1}	GO:0007155~cell adhesion	6.4×10^{-8}	1.1×10^{-4}
GO:0003836~beta-galactoside alpha-2,3-sialyltransferase activity	1.1E-03	2.2×10^{-1}	GO:0022610~biological adhesion	6.6×10^{-8}	1.1×10^{-4}
GO:0005576~extracellular region	1.2E-03	2.7×10^{-1}	GO:0009611~response to wounding	1.5×10^{-7}	2.6×10^{-4}
GO:0051059~NF-kappaB binding	2.0E-03	2.7×10^{-1}	GO:0007160~cell-matrix adhesion	2.0×10^{-6}	3.4×10^{-3}
GO:0045591~positive regulation of regulatory T cell differentiation	2.1E-03	8.1×10^{-1}	GO:0006954~inflammatory response	3.6×10^{-6}	6.3×10^{-3}
GO:0045582~positive regulation of T cell differentiation	2.9E-03	7.9×10^{-1}	GO:0031589~cell-substrate adhesion	5.1×10^{-6}	8.8×10^{-3}
GO:0045589~regulation of regulatory T cell differentiation	3.4E-03	7.4×10^{-1}	GO:0040012~regulation of locomotion	9.2×10^{-6}	1.6×10^{-2}
GO:0008374~O-acyltransferase activity	6.1E-03	5.1×10^{-1}	GO:0030334~regulation of cell migration	9.5×10^{-6}	1.6×10^{-2}
GO:0045619~regulation of lymphocyte differentiation	6.6E-03	8.8×10^{-1}	GO:0001558~regulation of cell growth	1.4×10^{-5}	2.4×10^{-2}
GO:0006644~phospholipid metabolic process	1.0E-02	9.3×10^{-1}	GO:0007010~cytoskeleton organization	2.8×10^{-5}	4.9×10^{-2}
			GO:0044421~extracellular region part	1.3×10^{-12}	1.8×10^{-9}
			GO:0005576~extracellular region	2.5×10^{-9}	3.3×10^{-6}
			GO:0044459~plasma membrane part	4.3×10^{-7}	5.7×10^{-4}
			GO:0031012~extracellular matrix	5.0×10^{-7}	6.7×10^{-4}
			GO:0005615~extracellular space	1.7×10^{-6}	2.3×10^{-3}
			GO:0005578~proteinaceous extracellular matrix	6.5×10^{-6}	8.6×10^{-3}
			GO:0005886~plasma membrane	9.9×10^{-6}	1.3×10^{-2}
			GO:0016010~dystrophin-associated glycoprotein complex	2.1×10^{-5}	2.8×10^{-2}
			GO:0005515~protein binding	3.9×10^{-8}	5.7×10^{-5}
			GO:0008092~cytoskeletal protein binding	1.8×10^{-5}	2.6×10^{-2}
			GO:0003779~actin binding	2.0×10^{-5}	2.9×10^{-2}

Comparison of Hypergeometric Testing between GWAS and Gene Expression Profiling

Of the GO functional annotation terms obtained from over-representation analysis of GWAS and gene expression profiling data, only one term, GO:0005576~extracellular region, reached adequate significance and was common to both datasets. Investigation of genes annotated to this term revealed 3 genes that were common to both GWAS and gene expression profiling data: *LTBP4, FMOD* and *IL7R*.

All three of these genes were also found in the GO:0005515~protein binding term, which was enriched in the gene expression profiling dataset. *LTBP4* and *FMOD* were annotated to three other GO terms involved with extracellular matrix organisation (GO:0044421~extracellular region part, GO:0031012~extracellular matrix, GO:0005578~proteinaceous extracellular matrix), and *LTBP4* to an additional term GO:0001558~regulation of cell growth.

4.3.3.2 Gene Set Enrichment Analysis (GSEA)

Over-representation analysis by hypergeometric tools uses a threshold to define genes of interest that warrant further investigation. In contrast, GSEA uses the list rank information from all genes examined in a gene expression study or GWAS without the use of a threshold, hence making use of all the data obtained from an experiment (Subramanian et al. 2005).

GSEA analysis of GWAS data in ICSNPathway found enrichment of five GO terms with FDR-corrected significance of $p<0.05$, involved with extra-cellular matrix collagen binding (GO:0005518), cytoskeletal myosin binding (GO:0017022), epithelial differentiation and morphogenesis (GO:0030855, GO:0002009) and RNA processing (GO:0006396). ICSNPathway analysis hypothesised that 4 SNPs (rs2296871, rs543573, rs1183768, rs1185193) in the *SETX* locus contributed to changes in RNA processing, 2 SNPs (rs892089, rs892090) in the *GP6* locus contributed to changes in collagen binding, 1 SNP (rs61929765) in the *KRT3* locus contributed to changes in epithelial differentiation and morphogenesis, and 1 SNP (rs1074407) in the *TRIOBP* locus contributed to changes in myosin binding.

GSEA analysis of gene expression data in Partek uncovered 2 GO terms with FDR corrected significance of $p<0.05$, involving regulation of GTPase activity (GO:0043087) and cell cortex (GO:0005938). A further 143 GO terms were found to be enriched at FDR corrected $p<0.25$, a discovery level of significance suggested by Subramanian et al. (2005). All four genes of interest from ICSNPathway GWAS analysis (*SETX, GP6, KRT3* and *TRIOBP*) were found in GO terms with significance of FDR $p<0.25$ within the gene expression functional enrichment analysis. The most

significant GO term was found to be GO:0017022 myosin binding, containing 15 genes including *TRIOBP* (p=0.025).

4.3.3.3 Selection of Candidate Biomarkers

Functional enrichment analysis with both hypergeometric testing and GSEA highlighted two key processes, cytoskeleton organisation and extracellular matrix remodelling, involved in carotid plaque progression and rupture in ischaemic stroke patients. Hypergeometric testing suggested an important role for *LTBP4* in extracellular matrix organisation. *LTBP4*, found at 19q13.2, also contained two SNPs at suggestive genome-wide significance (rs2303729, $p=7.91x10^{-5}$; rs10403833, $p=9.71x10^{-5}$). GSEA analysis highlighted a role for the non-synonymous SNP rs1074407, within *TRIOBP*, in cytoskeletal organisation. These findings indicated that LTBP4 and TRIOBP warranted further investigation.

4.3.4 Alternative Splicing of TRIOBP

A total of 181 SNPs within a 20kb wingspan were annotated to the *TRIOBP* locus using LDAK software, resulting in the gene achieving a Bonferroni adjusted value of p=0.025. 11 SNPs achieved genome-wide significance of $p<1 \times 10^{-3}$, of which 3 SNPs were found to be genotyped (rather than imputed) and within the *TRIOBP* gene. The sentinel SNP rs1074407 was found to have an association of OR 0.62 (95% CI 0.40-0.72, $p=1.38x10^{-4}$) with symptomatic carotid stenosis ≥20%. The two other genotyped SNPs demonstrated very strong correlation with the sentinel SNP ($r^2>0.90$) and comparable strengths of association (rs4820297: OR 0.63, 95% CI 0.42-0.75, $p=2.45x10^{-4}$; rs12628603: OR 0.65, 95% CI 0.48-0.79, $p=6.55x10^{-4}$).

Investigation of rs1074407 in Ensembl found that the SNP is located in the 3' UTR region of a *TRIOBP* exon, 759bps downstream of the splice region TMP_ESP_22_38155424 (NHLBI Exome Sequencing Project, release 6500). Furthermore, two SNPs in high LD with rs1074407 ($r^2>0.90$), spanning a 39kb region, were also found to be a 3' UTR variant (rs1129448) and an intronic variant predicted to be within an enhancer region (rs13054052). These SNPs were imputed during GWAS analysis, and found to have significant associations with carotid stenosis (rs1129448: OR 0.62, $p=1.45x10^{-4}$; rs13054052: OR 0.64, $p=5.29x10^{-4}$; Table 4.7).

Table 4.7: SNPs in the *TRIOBP* gene. The sentinel SNP rs1074407 and two other genotyped SNPs in high LD with the sentinel SNP were found to have a strong association with symptomatic carotid stenosis ≥20%. The sentinel SNP was found to be located within a 3' UTR region of *TRIOBP* exon, as were two other imputed SNPs in high LD with the sentinel SNP. **A1**, risk allele; **D'**, D-prime measure of linkage disequilibrium; **I^2,** I^2 measure of heterogeneity; **LD**, linkage disequilibrium; **OR,** odds ratio; **Q**, Cochran's Q measure of heterogeneity.

Position	SNP	Genotyped/ Imputed	A1/A2	p-value	OR (95% CI)	Q	I^2	LD with sentinel SNP, r^2 (D')
38156183	**rs1074407**	**Genotyped**	**T/A**	**1.38×10^{-4}**	**0.62 (0.40-0.72)**	**0.85**	**0**	**1.00 (1.00)**
38105981	rs4820297	Genotyped	C/T	2.45×10^{-4}	0.63 (0.42-0.75)	0.85	0	0.93 (1.00)
38119213	rs12628603	Genotyped	G/A	6.55×10^{-4}	0.65 (0.48-0.79)	0.92	0	0.97 (1.00)
38130385	rs13054052	Imputed	A/C	5.29×10^{-4}	0.64 (0.47-0.77)	0.64	0	0.91 (0.97)
38170026	rs1129448	Imputed	G/A	1.45×10^{-4}	0.62 (0.42-0.74)	0.62	0	0.97 (0.98)

To investigate the relationship between rs1074407 and symptomatic carotid stenosis, quantitative trait locus (QTL) analysis was performed in the BRAINS patient cohort. Degree of carotid stenosis was found to increase consecutively with allele count, with the TT genotype of rs1074407 associated with lower degree of stenosis than the AT or AA genotype (mean stenosis: TT=21%, AT=22%, AA=29%; ANOVA p-value: 0.005). This relationship was replicated in the two other genotyped SNPs (rs4820297 and rs12628603) in high LD with rs1074407 (Figure 4.8A).

Analysis of differential gene expression between TRIOBP in paired symptomatic stenosing plaque (SP) and non-stenosing tissue (NST) samples found no significant differential expression of TRIOBP (p=0.209, fold difference=1.20). The unique design of the Affymetrix Human Exon array, with probe sets targeting individual exons, also allowed investigation of regional differences in TRIOBP mRNA splicing (p= 0.0144). This demonstrated a relative reduction in the expression of exons 10 and 11 in symptomatic SP samples when compared with symptomatic NST samples, indicating lower expression of isoforms containing these exons specifically in symptomatic SP (Figure 4.8B). Furthermore, rs1074407 was found to reside in the 3' UTR region of exon 8. This suggests that the rs1074407 T allele may be responsible for the alternative splicing of the TRIOBP mRNA, which in turn influences degree of carotid stenosis and plaque rupture in ischaemic stroke patients. Significant alternative splicing of TRIOBP was not apparent between asymptomatic SP and NST samples (p=0.357) or between asymptomatic NST and symptomatic NST samples (p=0.644). However, significant alternative splicing was observed between asymptomatic SP and symptomatic SP samples (p=0.0435), indicating that lower expression of isoforms containing exons 10 and 11 is specific to symptomatic SP and may be determined by rs1074407.

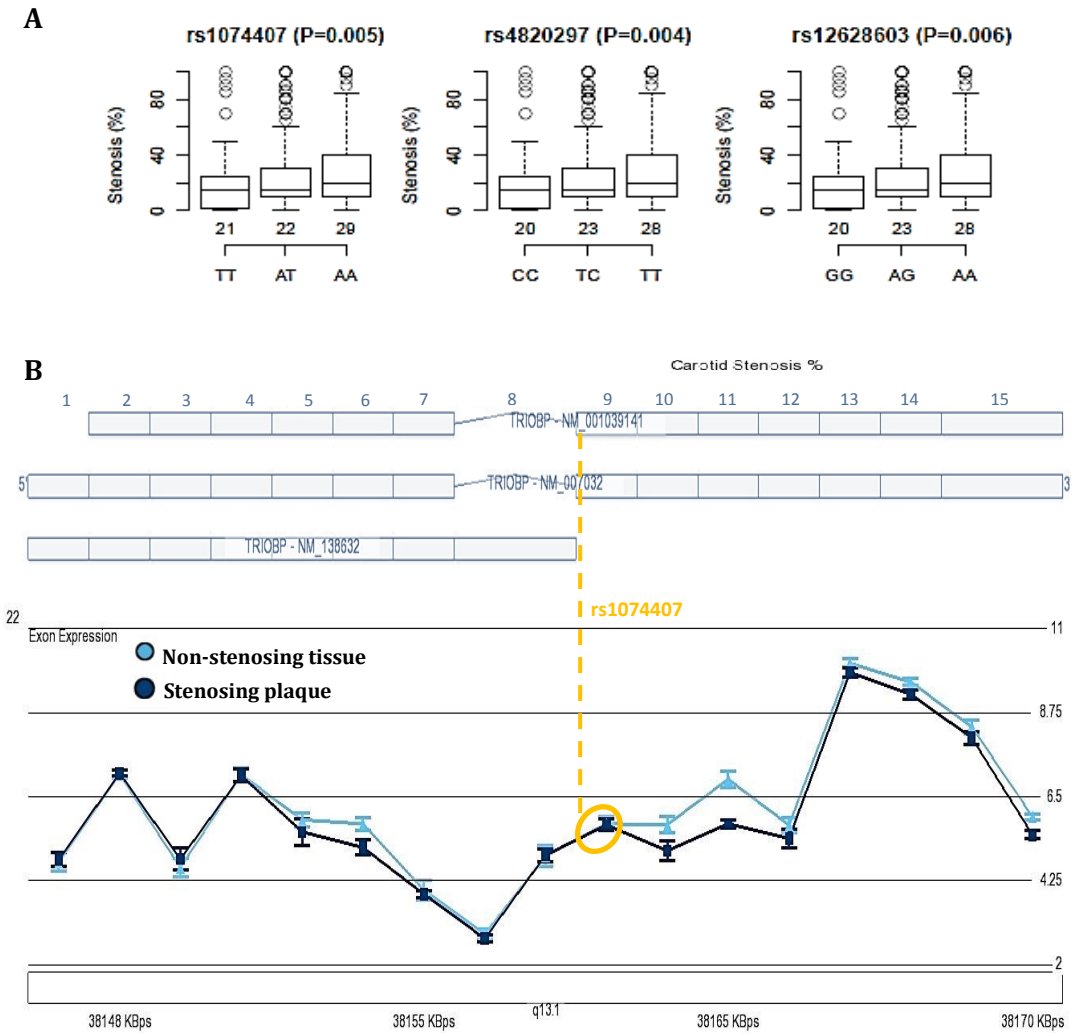


Figure 4.8: Exploring the alternative splicing profile of *TRIOBP*. A) The association between degree of carotid stenosis (%) and genotyped SNPs within the *TRIOBP* gene was investigated in the BRAINS GWAS cohort of 344 individuals with ischaemic stroke to explore quantitative trait loci (QTL). Box and whisker plots are shown for the sentinel SNP rs1074407 and two other genotyped SNPs in high LD with the sentinel SNP (rs4820297 and rs12628603). The horizontal line indicates the median, the box covers the 25–75% percentiles and the maximum length of each whisker is 1.5 times the interquartile range. Mean values of % stenosis are given. These means were used in a non-parametric ANOVA test of variance explained; the pvalues are based on the variance explained by a linear regresson of stenosis on allele count, using two degrees of freedom (the intersect and the gradient, and mean effect of each allele). Individuals with the 'risk' allele T of the sentinel SNP rs1074407 demonstrate a significantly lower degree of carotid stenosis than those with the A allele. A similar relationship is observed with the tested 'risk' alleles of SNPs rs4820297 and rs12628603. **B)** *TRIOBP* exon structure and protein isoforms were explored in the gene expression study data using Partek Genomics Suite software. Three protein isoforms were annotated (NM_001039141, NM_007032, NM_138632), and sentinel SNP rs1074407 was found to lie within the 3' UTR region of exon 8 of the gene, corresponding to Affymetrix exon array probe sets 3944969 and 394470. Least Square (LS) mean expression levels (log_2scale) plotted for each probe set covering the whole *TRIOBP* transcript demonstrated an increased inclusion of exons 10 and 11 (Affymetrix exon array probe sets 3944973-3944982) in non-stenosing tissue regions, as compared with stenosing plaque regions (p= 0.0144). It is hypothesised that SNP rs1074407 may be responsible for this alternative splicing profile, resulting in a protein isoform which is responsible for the stenosing plaque phenotype.

4.3.5 Differential Gene Expression of LTBP4

Annotation of 129 SNPs within a 20kb wingspan of the *LTBP4* locus in LDAK software achieved a Bonferroni adjusted value of p=0.01. 35 SNPs achieved genome-wide significance of $p<1 \times 10^{-3}$, of which 4 SNPs were found to be genotyped (rather than imputed) and within the *LTBP4* gene. The sentinel SNP rs2303729 (risk allele G) demonstrated an association of OR 1.64 (95% CI 1.29-2.08, $p=7.91 \times 10^{-5}$) with carotid stenosis ≥20%. The three other genotyped SNPs demonstrated moderate correlation with the sentinel SNP ($r^2=0.68$-0.71) and similar strengths of association (rs1131620: OR 1.62, 95% CI 1.30-2.08, $p=1.3 \times 10^{-4}$; rs7259067: OR 1.62, 95% CI 1.26-1.97, $p=1.3 \times 10^{-4}$; rs1051303: OR 1.62, 95% CI 1.26-1.97, $p=1.3 \times 10^{-4}$) (Table 4.8).

Examination of rs2303729 using Ensembl found that the SNP is located on the forward DNA strand and encodes an exonic missense purine variant, resulting in a change in amino acid from valine to isoleucine (GTA to ATA, VAR_037119). ENCODE segmentation analysis of mammary epithelial cells (HMEC) and epidermal keratinocytes (NHEK) found that this variant lies in a predicted strong enhancer/*cis*-regulatory element. Furthermore SNP rs10410606, in high LD with rs2303729 ($r^2=0.92$), was also found to lie in a predicted strong enhancer/*cis*-regulatory element. This was an imputed SNP in the GWAS, found to have an association with carotid stenosis of OR 1.63 (95% CI 1.29-2.12; $p=1.09 \times 10^{-4}$) (Table 4.7). Both SNPs were further investigated in is-rSNP, an *in silico* software tool that predicts the possibility of variant being a regulatory SNP affecting transcription factor binding (Macintyre et al. 2010). A number of transcription factors were identified at a Bonferroni adjusted significance of p<0.05, including potentially altered Smad4 and PLZF binding due to rs2303729 (p=0.250 and p=0.0013 respectively) and altered SRY, SOX5 and HNF3A binding due to rs10410606 ($p=2.5 \times 10^{-4}$, p=0.0030 and p=0.0032 respectively).

Table 4.8 SNPs in the *LTBP4* gene. The sentinel SNP rs2303729 and three other genotyped SNPs in moderately high LD with the sentinel SNP were found to have a strong association with symptomatic carotid stenosis ≥20%. The sentinel SNP was found to lie in a predicted strong enhancer/*cis* regulatory element, as was two another imputed SNPs in high LD with the sentinel SNP. **A1**, risk allele; **D'**, D-prime measure of linkage disequilibrium; **I^2,** I^2 measure of heterogeneity; **LD**, linkage disequilibrium; **OR,** odds ratio; **Q**, Cochran's ; Q measure of heterogeneity.

Position	SNP	Genotyped /Imputed	A1/A2	p-value	OR (95% CI)	Q	I^2	LD with sentinel SNP, r^2 (D')
41111069	**rs2303729**	**Genotyped**	**G/A**	**7.91×10^{-5}**	**1.63 (1.29-2.08)**	**0.93**	**0**	**1.00 (1.00)**
41117869	rs1131620	Genotyped	A/G	1.33×10^{-4}	1.63 (1.30-2.08)	0.88	0	0.68 (0.88)
41118056	rs1051303	Genotyped	A/G	1.35×10^{-4}	1.62 (1.26-1.97)	0.88	0	0.68 (0.88)
41124931	rs7259067	Genotyped	C/T	1.34×10^{-4}	1.62 (1.26-1.97)	0.90	0	0.71 (0.92)
41108975	rs10410606	Imputed	A/C	1.09×10^{-4}	1.63 (1.29-2.12)	0.99	0	0.92 (0.93)

QTL analysis was performed in the BRAINS GWAS patient cohort to investigate the relationship between rs2303729 and symptomatic carotid stenosis. Degree of stenosis was shown to increase significantly from the AA genotype to the GG genotype (mean stenosis: AA=20%, AG=23%, GG=29%; ANOVA p-value=0.009), providing evidence for an additive effect of the risk allele G. This relationship was replicated in the three other genotyped SNPs, rs1131620, rs7259067 and rs1051303 (Figure 4.9A).

Analysis of differential gene expression between LTBP4 in paired symptomatic stenosing plaque (SP) and non-stenosing tissue (NST) samples found significantly lower expression of LTBP4 in SP (p= 0.033, fold difference=2.61; Figure 4.9B). No evidence of alternative splicing of LTBP4 was observed (p=0.187, Figure 4.9C). Furthermore, no significant difference in LTBP4 gene expression was observed between asymptomatic SP and NST samples (p=0.319), or between asymptomatic and symptomatic NST samples (p=0.609), but was apparent between asymptomatic and symptomatic SP (p=0.046, fold difference=2.23), indicating that substantial down-regulation of LTBP4 is a specific phenomenon occurring within symptomatic plaques and may be a factor in causing plaque rupture.

4.3.6 LTPB4 as a Biomarker for Symptomatic Carotid Stenosis

The findings above indicate that SNP rs2303729 could represent an eQTL for *LTBP4,* with LTBP4 having potential to be a biomarker capable of identifying plaque progression and rupture. Nonetheless, it is important to note that this relationship has not been formerly tested here, as GWAS and gene expression data come from separate patient cohorts.

The Disease Association Protein-Protein Link Evaluator (DAPPLE) software was next used to investigate protein-protein interaction networks in the paired analysis of gene expression differences between symptomatic stenosing plaque and non-stenosing tissue. The gene list input to DAPPLE was limited to the 368 genes exhibiting a fold difference of >2 at a significance threshold of $p<0.05$, using an iterative process with 1000 network permutations. 97 disease proteins found to participate in the direct network (Figure 4.10). LTBP4 was shown to be a key network node and interact directly with one other protein, CD209, found to be significantly up-regulated in symptomatic stenosing plaque as compared to non-stenosing tissue in the paired gene expression study (p=0.0058, fold difference=2.57). Furthermore, LTBP4 was shown to interact indirectly with 39 other proteins, of which one, FMOD, was earlier annotated in LDAK (Bonferroni adjusted p=0.01), and found to be associated with 11 SNPs with a genome-wide significance of $p<1x10^{-3}$.

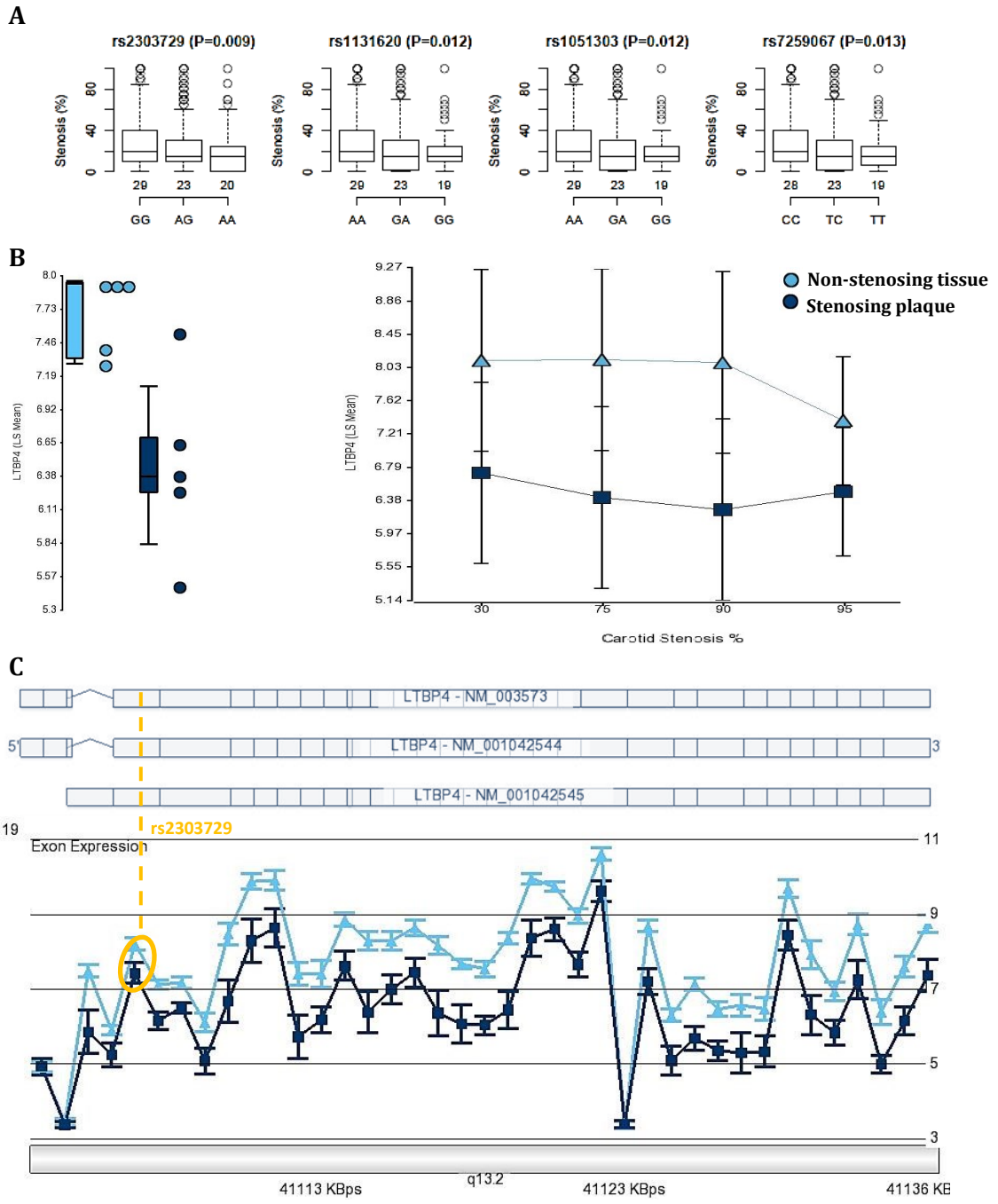


Figure 4.9: Exploring the differential expression of *LTBP4*. A) The association between degree of carotid stenosis (%) and genotyped SNPs within the *LTBP4* gene was investigated in the BRAINS GWAS cohort of 344 individuals with ischaemic stroke to explore quantitative trait loci (QTL). Box and whisker plots are shown for the sentinel SNP rs2303729 and three other genotyped SNPs in moderately high LD with the sentinel SNP (rs1131620, rs7259067 and rs1051303). The horizontal line indicates the median, the box covers the 25–75% percentiles and the maximum length of each whisker is 1.5 times the interquartile range. . Mean values of % stenosis are given. These means were used in a non-parametric ANOVA test of variance explained; the pvalues are based on the variance explained by a linear regresson of stenosis on allele count, using two degrees of freedom (the intersect and the gradient, and mean effect of each allele). Individuals with the risk allele G of the sentinel SNP rs2303729 demonstrate a significantly higher degree of carotid stenosis than those with the A allele. A similar relationship is observed with the tested risk alleles of SNPs rs1131620, rs7259067 and rs1051303. These plots provide a robust assessment of the associations observed. **B)** LTBP4 gene expression levels were next examined in symptomatic patients from the gene expression study (n=5). In differential gene expression analysis, least Square (LS) mean expression levels of LTBP4 were found to be significantly higher in non-stenosing tissue samples as compared to paired stenosing plaque regions (p=0.0334, fold difference=2.61). **C)** *LTBP4* exon structure and protein isoforms were explored in the gene expression study data using Partek Genomics Suite software. Three protein isoforms were annotated (NM_003573, NM_001042544, NM_001042545), and sentinel SNP rs2303729 was found to lie within the predicted strong enhancer/*cis* regulatory element region of the gene, corresponding to Affymetrix exon array probe sets 3833639 and 383640. It is hypothesised that SNP rs2303729 may be an expression quantitative train locus (eQTL) responsible for the differential expression of LTBP4 associated with the stenosing carotid plaque phenotype and plaque rupture.

FMOD was found to be significantly associated with carotid disease in ischaemic stroke through GWAS analysis (top SNP rs2364940, OR: 1.814, $p=1.33x10^{-4}$) and was involved in many GO terms identified through functional enrichment analysis through hypergeometric testing (Table 4.5). The gene was also found to be significantly down-regulated in symptomatic stenosing plaque as compared to non-stenosing tissue in the paired gene expression study (p=0.0039, fold difference=-2.63).

To establish the effects of LTBP4 down-regulation on the molecular network, the LS mean gene expression profiles of LTBP4, CD209 and FMOD were investigated for correlation. LTBP4 and FMOD demonstrated good correlation ($r^2=0.80$), and FMOD and CD209 also demonstrated relatively high correlation ($r^2=0.70$). The correlation between LTBP4 and CD209 levels was found to be $r^2=0.41$; however removal of one outlying data point substantially improved correlation to $r^2=0.72$. The scatterplots of these relationships are illustrated in Figure 4.11. Given that both LTBP4 and FMOD were found to be genes of significance in both the gene expression and GWAS functional enrichment analyses, these findings show that common genetic associations implicate regions encoding proteins that physically interact to determine complex disease processes (Rossin et al. 2011). In particular, it provides evidence for the importance of LTBP4 within a molecular network influencing carotid plaque progression and rupture in ischaemic stroke patients.

Sensitivity and specificity ROC analysis between LTBP4 gene expression in paired symptomatic stenosing plaque and non-stenosing tissue regions yielded an AUC of 0.92, with a sensitivity of 100% and specificity of 80% at a LS mean cutoff of 7.04 (Figure 4.12).

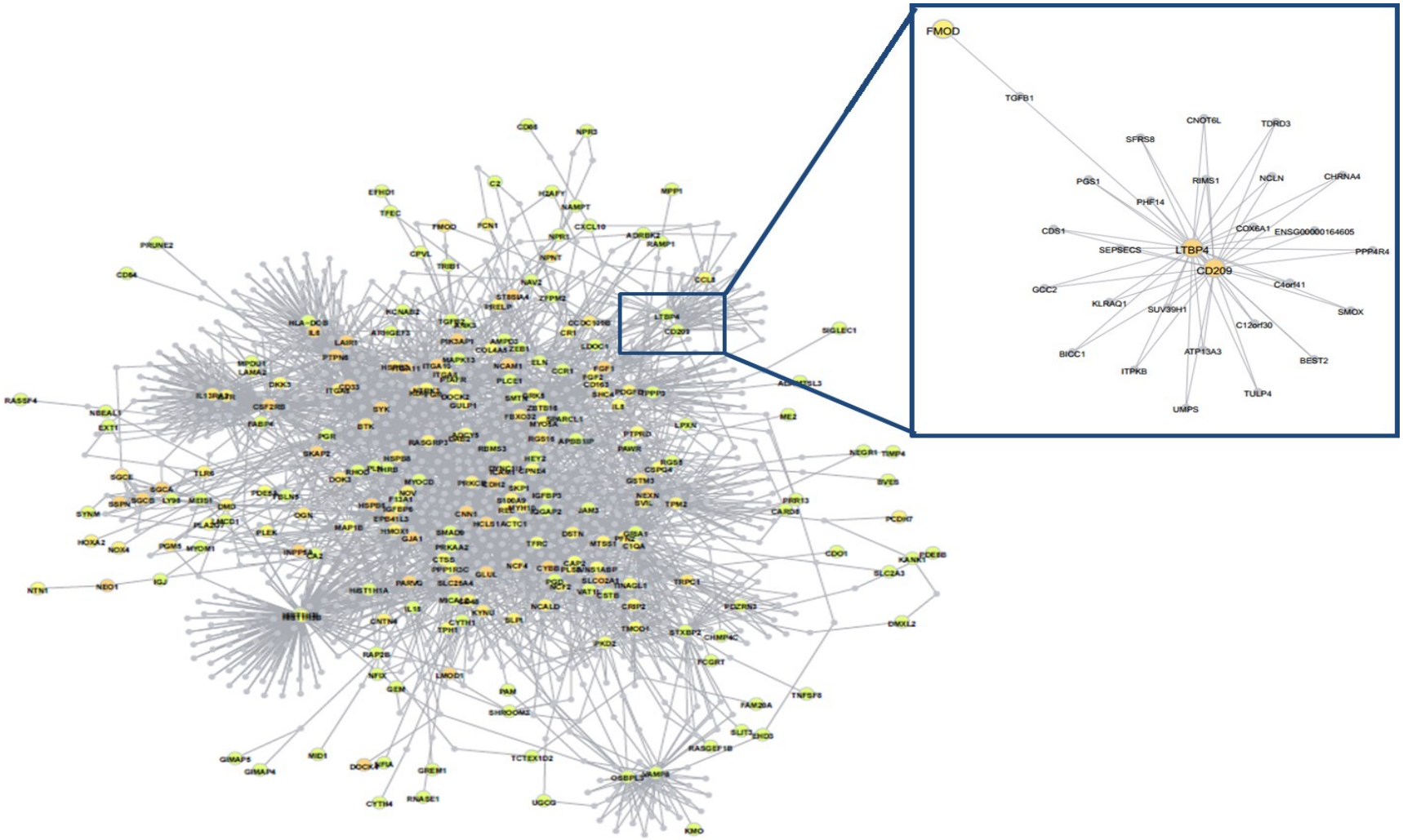


Figure 4.10: Protein-Protein Interaction analysis of LTBP4. The 368 genes exhibiting differential expression in the gene expression study of symptomatic non-stenosing tissue versus stenosing plaque, with a fold difference of >2 at a significance threshold of $p<0.05$, were included as an input to DAPPLE to create a protein-protein interaction network. Associated proteins are represented as nodes connected by an edge if there is *in vitro* evidence for a high-confidence interaction. The large coloured circles represent proteins found to be differentially expressed in the gene expression study, and small circles in grey represent the connected proteins (not known found to be differentially expressed). Each interaction defined in the DAPPLE network is depicted by a line (edge) between nodes. LTBP4 was found to be a key node, directly connected to CD209 and indirectly connected to FMOD.

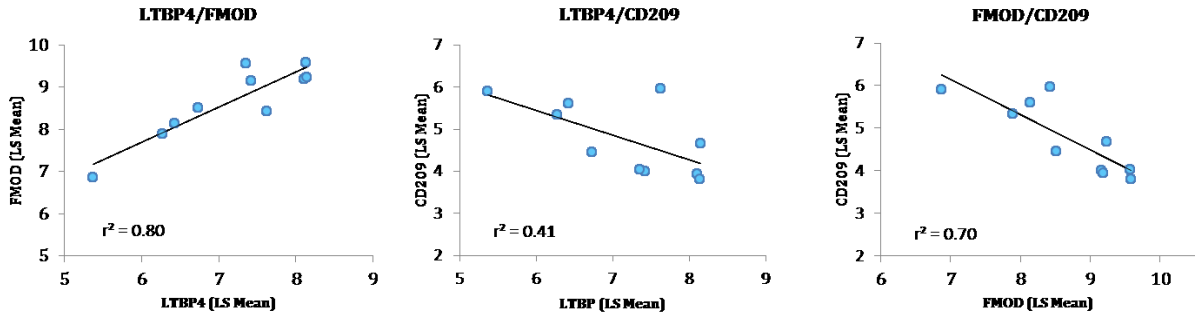


Figure 4.11: Correlations in expression levels between genes involved in the LTBP4 network. The relationship between LTBP4 and interacting proteins was explored using gene expression study data, with the assumption that differential expression of LTBP4 in symptomatic stenosing plaque should similarly affect the expression profiles of interacting genes FMOD and CD209. All three genes demonstrated good correlation (removal of one data point substantially improved correlation of LTBP4 and CD209 from r^2=0.41 to r^2=0.72), indicating that LTBP4 could have a key functional role in symptomatic plaque pathology and could be a promising biomarker of symptomatic plaque instability.

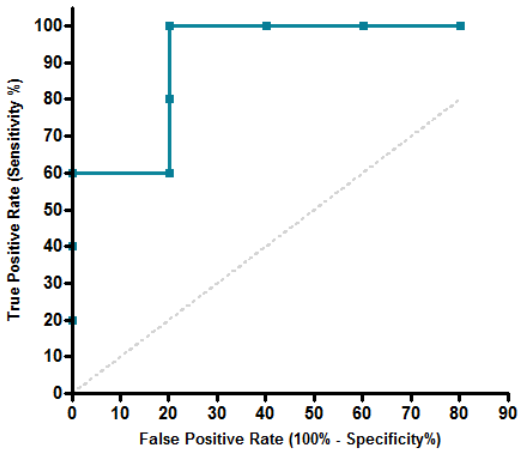


Figure 4.12: Examination of *LTBP4* as a biomarker for predicting carotid plaque rupture. Least square (LS) mean gene expression profiles for LTBP4 were examined in paired stenosing plaque and non-stenosing tissue regions of 5 symptomatic patients. ROC analysis between *LTBP4* expression in these paired samples yielded an AUC of 0.92, with a sensitivity of 100% and specificity of 80% at a LS mean threshold of 7.04, indictating the potential value of LTBP4 as a biomarker for predicting carotid plaque rupture.

4.4 Discussion

The degree of carotid stenosis and characteristics such as thin fibrous cap and immune cell infiltration are associated with symptomatic carotid plaques. However, these features are poor predictors of thromboembolism and it is still unclear why some carotid stenosis progress to occlusion without causing symptoms and others manifest themselves with sudden cerebral infarction. Here, gene expression profiling was undertaken on stenosing and non-stenosing tissue regions of individuals with asymptomatic or symptomatic carotid disease. Results from this gene expression analysis were combined with a GWAS of ischaemic stroke patients with varying degrees of carotid artery stenosis in an integrative genomics approach. The study provides evidence that genetic variation impacts pathological processes underlying symptomatic carotid stenosis and upholds the value of utilising an integrative genomics approach for exploring genetic contributions to carotid disease.

The approach used in this study has been undertaken by other research groups to understand the molecular mechanisms of human cardiovascular disease. Coupling multi-tissue expression gene expression profiles with GWAS data from the CARDIoGRAM study, Huan et al. (2013) discovered a number of co-expression modules involved in B-cell activation, immune response, and ion transport in coronary artery disease, and identified FMOD as a causal differential module upon network analysis. Application of network approaches to gene expression data of carotid endarterectomies from the Biobank of Karolinska Endarterectomies (BiKE) biobank also implicated a key role for ApoC-1 in the pathogenesis of carotid disease (Diez et al. 2010). Nonetheless, the current study is the first of its kind to utilise a pathway-based, integrative genomics approach to understanding the biology underlying symptomatic carotid disease.

Gene Set Enrichment in Symptomatic Stenosing Plaques

High-throughput technologies produce extensive lists of 'genes of interest', however the biological implications of these genes is often difficult to deduce. Pathway and gene set analysis allows identification of biological processes that are likely to be most influential to a disease process. A plethora of these analysis tools have been developed, and as yet there is still no unified method or 'gold standard' (Huang da et al. 2009a). Three major layers are consistent between all these bioinformatics tools: the reference annotation database, algorithm and statistical analysis, and the presentation; the characteristics of each of these layers account for different outputs from each tool (Huang da et al. 2009a). Huang da et al. (2009b) propose that a 'good' gene list should contain a reasonable number of genes, of which most pass a set statistical threshold and explain expected biological processes, with enriched processes being present

more likely than by chance alone and being reproducible between analytical tools and in subsequent wet lab experiments. In this study, hypergeometric testing and GSEA were used for examination of gene expression and GWAS data, under the hypothesis that robust biological processes are likely to withstand analytical differences between tools and be common to both. The current functional enrichment analysis demonstrated intrinsic differences between non-stenosing tissue and unstable plaques in a number of biological processes, in particular extracellular matrix organisation and myosin binding. Further investigation of genes involved in these pathways uncovered two genes, *LTBP4* and *TRIOBP* respectively, which were strongly associated with symptomatic carotid stenosis in GWAS meta-analysis and demonstrated alternative splicing (TRIOBP) or differential gene expression (LTBP4) in microarray analysis of symptomatic stenosing plaques.

Further examination of the GO:0017022 myosin binding term, containing the *TRIOBP* gene and shown to be enriched through GSEA analysis, found it to be a child term of the cytoskeletal protein binding (GO:0008092) and protein binding (GO:0005515) terms. The protein binding term was also found to be enriched in hypergeometric testing of gene expression data, and contained the *LTBP4* and *FMOD* genes. This finding demonstrates the ability of hypergeometric testing and GSEA analysis to identify similar enriched gene sets, and in particular establishes the role of LTBP4 and TRIOBP in related pathological mechanisms within symptomatic stenosing plaques. The cytoskeleton connects to the extracellular matrix via focal adhesions - dynamic protein complexes through which both mechanical and regulatory signals are transmitted. Actomyosin contractility stimulates focal adhesion development (Wolfenson et al. 2011), leading to vascular smooth muscle cell proliferation, differentiation and migration and affecting focal adhesions between smooth muscle cells in the fibrous cap, thus affecting cap integrity and plaque stability (Nikol et al. 1997; O'Sullivan et al. 2011).

Interestingly, exploration of the top hits from GWAS analysis, the *RASGEF1B* and *KIF16B* genes, found them to be in a common GO cellular component term (GO:0005769 early endosome) and both involved in signal transduction (GO:0007165). Additionally, KIF16B plays a key role in microtubule-assisted movement, providing further evidence for the role of the cytoskeleton in plaque progression.

The molecular mechanisms identified through hypergeometric testing and GSEA have previously been implicated in ischaemic stroke carotid disease, hence it could be argued that this study provides little novel insight into the molecular mechanisms underlying plaque progression and rupture. However, one of the biggest strengths of the integrative genomics approach is that it helps to infer casual associations. Hence, the fact that genes in the significant

pathways show evidence of genetic association and appear to be alternatively spliced or differentially expressed makes them highly plausible causal biomarkers for disease.

TRIOBP Alternative Splicing in Symptomatic Stenosing Plaque

GSEA analysis of GWAS data hypothesised that SNP rs1074407 in *TRIOBP*, part of the GO:0017022~myosin binding gene set, was associated with carotid stenosis in ischaemic stroke patients. *TRIOBP* encodes the TRIO And F-Actin Binding Protein, which stabilises F-actin to regulate actin cytoskeleton organisation. It also interacts with trio protein, which promotes the exchange of GDP by GTP via Rho GTPases to coordinate cell-matrix and cytoskeletal rearrangements necessary for cell migration and cell growth (Seipel et al. 2001).

The role of the cytoskeleton in the complex regulation of vascular endothelial repair has been well documented. Upon damage and loss of endothelial integrity, cytoskeletal systems are activated to regulate cell migration and proliferation through re-organisation of F-actin distribution for endothelial repair (Lee & Gotlieb 2003). The loss of structural and functional integrity of the endothelial monolayer, although unlikely to be the cause of initial atherosclerotic lesions, is observed on the surface of lesions leading to ulceration and fissure formation that could promote thrombosis and rupture of the atheroma through the fibrous cap (Virmani et al. 2001).

A major finding in this study was that alternative splicing of mRNAs might be more important than gene expression in determining plaque rupture. This is consistent with a genome-wide screen by Heinzen et al. (2008), who discovered a number of SNPs associated with tissue-specific splicing events. Alternative splicing has been shown to be more frequent than commonly suggested, with every human multi-exon protein-coding gene containing an average of 6.3 alternatively spliced transcripts (Frankish et al. 2012). The ENCODE project reported that while most alternative isoforms do not seem to have a role as functional proteins, they may hold implications for the control of gene regulation and expression. The collaborators suggested that potentially deleterious splice variants lie dormant within the gene and become highly expressed only as a result of a disease event (Tress et al. 2007).

Studies have previously reported that alternative isoforms of TRIOBP can affect the durability and rigidity of F-actin bundling, and could be a cause of hereditary human deafness (Kitajiri et al. 2010). There is also accumulating evidence that polymorphisms in the 3' UTR can alter miRNA binding sites and RNA secondary structure (Yu et al. 2007; Sasaki et al. 2013), suggesting that the *TRIOBP* rs1074407 polymorphism could lead to changes in exon expression. Without data based upon RT-PCR and RNA sequencing, it is difficult to ascertain the identity of

the TRIOBP transcripts related to the alternative splicing profile observed. However, the finding that consecutive probesets within exons 10 and 11 of *TRIOBP* were found to be down-regulated in symptomatic stenosing plaque indicates that these exons are likely to be part of the transcript that is regulated in relation to the rs1074407 genotype (Folkersen et al. 2009).

The finding that TRIOBP is uniquely alternatively spliced in unstable symptomatic plaques could allow for the development of specific therapies targeting this splice variant without affecting the function of normal TRIOBP in other tissue regions. However at this stage a number questions remain unanswered, including the identity of the specific isoform, it's functional role within the plaque and whether the isoform plays a role in normal cytoskeletal function or is a specific biomarker of an unstable plaque state. It is known that 13 alternatively spliced transcript variants exist for TRIOBP, some that may be subject to nonsense-mediated decay. TRIOBP is widely expressed at both the RNA and protein levels, especially within vascular tissue (heart), blood and the placenta, with isoform 3 being uniquely expressed in foetal brain, retina and cochlea (NCBI GEO db). Further experimental investigation of the splice variant uncovered in this present study, and its tissue-specific gene and protein expression is required to assess its value as a therapeutic target for plaque instability.

LTBP4 as a Biomarker for Plaque Rupture

Functional enrichment analysis of GWAS and gene expression data by hypergeometric testing identified extracellular matrix remodelling as a key process involved in carotid plaque instability. This finding is in keeping with a number of previous studies that have implicated the role of extracellular matrix remodelling in fibrous cap thinning and plaque rupture (Heo et al. 2011; Bobryshev et al. 2008). Changes in extracellular matrix organisation have also been reported in microarray studies examining symptomatic carotid plaques versus healthy control tissue (Korol et al. 2011), and symptomatic versus asymptomatic carotid plaques (Saksi et al. 2011; Papaspyridonos et al. 2006).

Two genes involved in extracellular matrix organisation, *LTBP4* and *FMOD*, were found to be differentially expressed between symptomatic stenosing plaque and non-stenosing tissue regions, as well as being associated with carotid stenosis upon GWAS meta-analysis of ischaemic stroke patients. Gene expression studies have previously shown that modules of differentially expressed genes identified through GO form networks of interacting genes (Minguez & Dopazo 2011; Narayanan et al. 2010). Furthermore, a study of GWAS of complex diseases found that genes found harbouring disease-associated SNPs implicate regions encoding proteins that physically interact (Rossin et al. 2011). Through DAPPLE protein-protein

interaction analysis, LTBP4 was found to be a key network node likely to affect the gene expression levels of CD209 (directly) and FMOD (indirectly, via TGF-β1).

LTBP4 is a protein that specifically binds to transforming growth factor beta (TGF-β1) as it is secreted and targeted to the extracellular matrix, thus serving as a molecular modulator for the disposition of TGF-β1 and the assembly of other matrix proteins. Mutations in the *LTBP4* gene have previously been found to effect TGF-β1 signalling and extracellular matrix organisation (Urban et al. 2009). TGF-β1 has been shown to have anti-inflammatory and atheroprotective functions; its production by lipid laden foam cells stimulates collagen production by smooth muscle cells, stabilising fibrous cap formation over the atheromatous plaque (Amento et al 1991). A down-regulation of gene expression was also observed for FMOD, a protein found to regulate the extracellular matrix collagen structure and sequester TGF-β1 into the extracellular matrix, thus affecting the development of atherosclerotic plaque tissue (Hildebrand et al. 1994; Shami et al. 2013). Thus the down-regulation of LTBP4 and FMOD in stenosing plaques, as observed in the current gene expression study, may cause a subsequent decrease in the expression of TGF-β1 and decreased stability of the fibrous cap, leading to plaque rupture. LTBP4 has also been shown to directly interact with CD209, an expression marker found on dendritic cells that have been observed in human atherosclerotic carotid plaques and known to activate the T cell and B cell immune response (Zhang & Zhang 2010). A close relationship exists between the mean signal intensity of dendritic cells in plaques and the symptoms of patients with significant carotid plaques (Kawahara et al. 2007), and CD209+ dendritic cells have been found in rupture-prone unstable plaques (Soilleux et al. 2002). These findings suggest that LTBP4 could be a key player in plaque pathophysiology and rupture and a good target for therapeutic intervention. The functional network developed *in silico* can be tested by knocking down key network nodes such as LTBP4 *in vitro* using RNA interference (RNAi) and studying the effects of loss of gene function on downstream nodes.

QTL mapping uncovered four *LTBP4* genotyped SNPs in high LD (rs2303729, rs1131620, rs1051303 and rs7259067), which affected degree of carotid stenosis in a genotype-dependent manner. All these four SNPs were strongly associated with symptomatic carotid stenosis (genome-wide significance of $p<1x10^{-4}$, OR1.6) and found to be located within the *LTBP4* gene sequence or within a 20kb wingspan of the *LTBP4* gene. The sentinel SNP rs2303729 was found to lie within an exon of *LTBP4*, in a predicted strong enhancer/*cis*-regulatory element. LTBP4 was shown to be uniquely down-regulated in symptomatic stenosing plaque in the gene expression study. It is therefore hypothesised that this SNP could be responsible for the lower expression of LTBP4 observed in symptomatic stenosing plaques within the gene expression study.

QTL mapping and differential expression analysis suggest that *LTBP4* is a gene whose expression may be under genetic control. Genes whose expression is *cis*-regulated, significantly correlated with a quantitative trait and co-localised with a physiological QTL have been postulated to be strong candidate genes for complex diseases (Hubner et al. 2005; Morrissey et al. 2011), demonstrating a robust relationship between the gene and the phenotype and indicating that the gene itself is under regulation by a local SNP. Unfortunately in this study, analysis was based upon different patients from the GWAS and gene expression studies, as no patients were common to both cohorts; investigation of LTBP4 gene expression in genotyped individuals through robust eQTL anlysis will help to confirm the hypothesis that rs2303729 is responsible for differential LTBP4 expression and carotid plaque instability.

The search for therapeutic options to prevent plaque destabilisation is limited by the lack of surrogate markers of disease progression (Peeters et al. 2009). The combination of standard clinical variables and molecular variables influenced by genetic mechanisms could provide an improved tool for ischaemic stroke risk stratification. Using a similar strategy, McGeachie et al. (2009) found that a predictive coronary artery calcification model based upon 13 SNPs and five clinical variables gave a predictive accuracy of 85%, a significant improvement ($p<0.001$) on a model using clinical variables alone.

LTBP4 has previously been implicated in abdominal aortic aneurysm (Thompson et al. 2010) and proposed as a vascular tissue marker of peripheral vascular disease (Buechler 2004). LTBP4 may prove to be a candidate biomarker for identifying vulnerable plaques prone to rupture. Its expression is measurable in blood plasma by a number of commercially available ELISA kits, hence providing an easy and non-invasive blood test for ischaemic stroke risk stratification. Nonetheless, it is important to consider that gene expression levels differ between various tissue types and biofluids (Hardy et al. 2009; Hernandez et al. 2012). Additionally, previous eQTL mapping has found that on average, more loci are associated with the level of a transcript than of its corresponding protein (Ghazalpour et al. 2011), and translation rate constants have been shown to be better predictors of protein levels than mRNA levels, which can explain only approximately 40% of protein level variation (Schwanhausser et al. 2011). These results demonstrate that accurate quantification of LTBP4 blood protein level is required before consideration as a non-invasive risk biomarker for ischaemic stroke.

Limitations of the Study

Technical Challenges

The Affymetrix Human Exon Array 1.0 ST allows interrogation of over a million exons; however it suffers from inherent problems common to all microarrays, such as probe hybridisation behaviour and probe annotation. These issues were minimised in the present study by removing from analysis probes that were predicted to be affected by SNPs, and developing a stringent analysis workflow based on only the 'core' annotated probesets.

The findings of this study are dependent upon factors such as pathway annotations, definition of gene boundaries, and assignment of a p value to a gene which has multiple SNPs, which could substantially alter analysis results and make them unstable (Jia et al. 2011). In an attempt to overcome this, two methods of functional enrichment analysis were undertaken to evaluate the reliability of the results, and preliminary findings were explored further in both the gene expression study and GWAS in order to replicate the conclusions.

In the field of ischaemic stroke and carotid artery disease, GWAS has had limited success in identifying SNPs that have been able to withstand replication in independent cohorts. Genetic associations are most amenable to discovery in highly heritable traits, where a relatively small number of genes exert a relatively large physiological effect. Therefore it is possible that the 'common disease, common variant' GWAS hypothesis may need to be revised, and future study designs need to be able to address the effects of both common and rare variants (Gibson 2011). Exome sequencing could provide a promising tool for exploring rare variants of high penetrance in complex disease such as ischaemic stroke (Cole et al. 2012).

Definition of Carotid Stenosis

Integrative genomics requires a sound understanding of the trait under study, and an appreciation of whether this trait is continuous (e.g. blood pressure) or categorical (e.g. hypertensive vs normotensive). In reality the relationship between the definition of a complex disease and its physiology is not straightforward (Ware et al. 2013). While categorisation of a continuous trait may lead to false associations during genomic analysis, it can also be theorised that genetic affects contribute to a threshold-dependent phenotype that overlies a continuous trait. In this GWAS study, the unavailability of individual patient data hindered the analysis of carotid stenosis as a continuous variable, and plaque was defined as stenosis ≥20%. This is in keeping with a meta-analysis of GWAS from the CHARGE Consortium, in which atherosclerotic thickening of the carotid artery wall was defined in seven of the nine cohorts as either the presence of plaque or the proxy measure of stenosis greater than 25% (Bis et al. 2011). The validity of choosing ≥20% stenosis as a measure of clinically relevant carotid artery disease is

further supported by the gene expression study, in which two out of seven individuals had symptom-causing plaques of <50% stenosis.

Study Design

This study is limited by small sample size, however power calculations in the gene expression study determined that an average sample size of 10 (5 in each analysis arm) delivered 80% power at $p<0.05$ to detect a 2-fold difference in means, providing reliability to the results obtained. The non-stenosing tissue region adjacent to the stenosing plaque was used as a control because of the difficulty in obtaining healthy carotid artery samples, and benefits from eliminating inter-patient heterogeneity - a strategy used in previous studies (Zawadzki et al. 2009; Di Taranto et al. 2012).

The difficulties in conducting a well-powered GWAS of ischaemic stroke have been previously noted (Meschia 2010); the very small sample size utilised in this GWAS casts doubt upon the associations observed between genetic variants and the presence of carotid stenosis in ischaemic stroke patients, and is likely to result in inflated effect sizes. Nonetheless, the effect size observed for *LTBP4* in this GWAS (OR1.64) is similar to effect sizes obtained for other genes found to be associated with large-vessel ischaemic stroke (e.g. *HDAC9*: OR1.42, $p=1.87\text{x}10^{-11}$; *LPA*: OR1.27, $p=6.72\text{x}10^{-4}$; *9p21.3*: OR1.17, $p=2.93\text{x}10^{-5}$; Meschia & Tournier-Lasserve 2013). The limitations of small population cohorts in both the gene expression study and GWAS meta-analysis have been reduced by applying a pathway-based approach to biomarker discovery, rather than relying on single SNP or gene associations. Thus, despite small sample sizes, the combination of both GWAS and gene expression datasets helped to reduce false positive associations and the pathways identified through functional enrichment analysis are plausible based upon current biological knowledge.

Heterogeneity of Carotid Stenosis and Ischaemic Stroke

Basic histological examination of plaques used for gene expression profiling found that symptomatic plaques often exhibited greater irregular surface morphology and ulceration, as compared to asymptomatic plaques. However, variations in plaque histology arising from different ischaemic events (TIA or ischaemic stroke), as previously reported (Redgrave et al. 2006), could not be delineated due to small sample size. Temporal symptomatic plaque remodelling also needs to be assessed before developing targeted therapies or a biomarker-based test for risk prediction (Peeters et al. 2009). Consideration of subject age and gender as covariates in both gene expression and GWAS analysis aimed to improve the reliability of results obtained.

Studies have shown that different carotid artery sites may be influenced by different etiological pathways, with IMT of the common carotid artery linked to total atherosclerotic burden and IMT of the internal carotid artery being attributed to local turbulent haemodynamic flow (Crouse et al. 1996). Possible genetic mechanisms underlying these differences have also been demonstrated in a recent GWAS of common and internal artery IMT (Melton et al. 2013). Due to small sample size and limited availability of phenotypic information from GWAS cohorts, the effects of carotid artery stenosis site could not be explored.

The METASTROKE GWAS study also demonstrated that pathophysiological mechanisms seem to vary between ischaemic stroke subtypes and advocated the need to study these subtypes independently. However, the authors also acknowledged that differing interpretations of TOAST subtypes in the 15 genotyped cohorts may have led to heterogeneity in classification (Traylor et al. 2012). This highlights the difficulty in collecting a 'clean' ischaemic stroke dataset. Large-vessel disease represented the majority of ischaemic stroke cases in this present GWAS and it is therefore postulated that the genetic associations found can be ascribed to this subtype, even though other subtypes were included in the analysis in order to maintain sample size. Nonetheless, these associations require robust validation in an independent homogenous (LAA only) cohort, in which the presence of carotid artery disease can be directly attributed with ipsilateral cerebral ischaemia. Replication of the pathways involved in extracellular matrix organisation and cytoskeleton binding in independent ischaemic stroke and carotid disease GWAS datasets will also help to validate the association of the *LTBP4* and *TRIOBP* with carotid plaque rupture and LAA stroke, similar to the strategy used by Anderson et al. (2013) to establish the association between common genetic variants within the oxidative phosphorylation genes and stroke.

Concluding Remarks

This study provides evidence for the genetic regulation of carotid disease using an integrative genomics approach. Cytoskeletal organisation and extracellular matrix remodelling were found to be key processes determining the stability of the carotid plaque that could be modulated by underlying genetic mechanisms. In particular, *LTBP4*, was found to be a key node gene involved in extracellular matrix organisation. LTBP4 down-regulation in symptomatic stenosing plaque is hypothesised to be affected by SNP rs2303729, a polymorphism found within an exonic region of the gene and located within a predicted enhancer/*cis*-regulatory region. The gene also exhibited high sensitivity and specificity to differentiate between non-stenosing carotid tissue and unstable stenosing plaque. Thus LTBP4 could prove to be a candidate risk biomarker for ischaemic stroke by predicting plaque instability and rupture.

CHAPTER FIVE:

FINAL DISCUSSION

Final Discussion

This thesis set out with the aim of addressing the hypothesis that novel candidate biomarkers for ischaemic stroke exist and are amenable to detection using –omic approaches. This final chapter summarises the work presented thus far, discusses the key findings and recommends future directions of biomarker discovery for ischaemic stroke.

The introductory chapter of the thesis brought to light the limitations of current risk stratification and diagnostic and prognostic evaluation of ischaemic stroke, and discussed the potential of biomarkers to overcome these issues. Biomarkers are used routinely in the management of many disorders. The delineation of biomarkers from normal biology can cause them to be newly expressed or up- or down-regulated, thus they have the potential to serve as diagnostic, prognostic and risk stratifying tools for disease. Blood-based biomarkers are ideal for rapid diagnostic tests for point of care testing, while tissue-based molecules can serve as mechanistic biomarkers that reflect the response of an organism to a local stimulus, rather than the systemic effect of a disease state. Recent advances in genomic, transcriptomic and proteomic technologies, coupled with developments in bioinformatics tools, herald an exciting era for biomarker discovery. While numerous studies have attempted to identify biomarkers for ischaemic stroke using these techniques, to date no biomarkers have demonstrated suitable clinical potential. The work in this thesis aimed to evaluate the current status of diagnostic and prognostic biomarkers for stroke and develop improved –omics workflows for novel biomarker discovery.

With the aim of summarising the state of biomarker research in ischaemic stroke and identifying existing candidate biomarkers that could warrant further investigation, a comprehensive systematic review and meta-analysis were conducted of published literature reporting acute blood-based biomarkers for the diagnosis and prognosis of ischaemic stroke. Review of the entire scientific literature ensured that no potential biomarkers were ignored, and combination of individual studies by meta-analysis allowed quantification of the association between candidate biomarkers and the phenotype of interest, with statistical power improved through increased sample size analysis. The results of the literature review and meta-analyses were presented in Chapter 2.

A total of 137 candidate biomarkers were identified for the diagnosis or prognosis of ischaemic stroke, with biomarkers being either proteins or metabolites broadly involved in lipid metabolism, coagulation and inflammatory processes. Ten candidate biomarkers were found to be significantly associated with acute ischaemic stroke as compared to healthy controls. Nonetheless further investigation of one of these biomarkers, CRP, revealed that significant association with a phenotype

does not necessarily indicate a good candidate biomarker, and CRP distributions in ischaemic stroke and healthy control populations demonstrated poor discriminatory potential of a CRP-based diagnostic test.

Systematic literature review and meta-analysis also identified coagulation factors and metabolites as significant candidate prognostic biomarkers for ischaemic stroke. Admission glucose levels were found to be significantly associated with poor outcome after ischaemic stroke and incidence of post-thrombolytic SICH. However, literature-based meta-analysis of the association of admission glucose levels with SICH was found to be limited by a number of technical challenges, casting doubt upon the conclusions drawn. In order to accurately quantify the risk of SICH associated with admission glucose levels, an individual patient data meta-analysis was conducted by obtaining patient information from eight cohorts, comprising the largest collection of data for analysis to date (>32,000 patients). Contrary to previously published studies, it was shown that hyperglycaemia has minimal association with risk of SICH and that international guidelines for thrombolysis (currently recommending an admission glucose level threshold of 22mmol/L) could be re-evaluated in light of these findings. The study also demonstrated the advantages of performing individual patient data meta-analysis over literature-based meta-analysis when attempting to establish associations between a biomarker and an outcome of interest.

Work in Chapter 2 found 11 candidate biomarkers studied for their potential to differentiate between ischaemic stroke and stroke mimics; however, only one of these, S100B, was found to be significantly associated with ischaemic stroke and to date no biomarkers have successfully progressed to use within a clinical setting, With the aim of identifying novel candidate biomarkers for AIS, a whole proteome approach, using SELDI-TOF MS, was conducted, the results of which were presented in Chapter 3. A pilot study consisting of 13 AIS, 2 TIA and 2 stroke mimic patients established the ability of SELDI-TOF MS and the CM10 ProteinChip to detect a protein profile that is significantly affected during acute cerebral ischaemia. Protein ions detected using SPA matrix were largely limited to the low molecular weight range, although this could be considered an advantage to biomarker discovery as it is a range that is often neglected by other proteomic techniques. It was recognised that the reproducibility of relative protein ion changes was a critical factor for determining candidate biomarker utility, therefore the pilot study also allowed optimisation of the SELDI-TOF MS workflow.

Having established the value of the SELDI-TOF MS approach to diagnostic biomarker discovery, a second patient cohort was obtained for a well-powered study that could overcome common limitations of study design, as identified in Chapter 2. The cohort of 404 subjects was randomised into Training, Test and Validation sets to allow robust interrogation of any putative biomarkers detected. Unsupervised PCA analysis of the Training set data (20 AIS, 10 TIA, 10 stroke mimics) demonstrated

distinct clustering of ischaemic events (AIS and TIA) and non-ischaemic events (stroke mimics) based upon protein ion profiles, that was not observed with other demographic variables. Subsequent DFA of protein ions detected using the pH4 binding condition revealed that the up-regulation of m/z 6640 and down-regulation of m/z 3699 in AIS and TIA could represent a potential dual-marker diagnostic test for acute cerebral ischaemia. Examination of these candidate biomarkers in the Test set (32 AIS, 16 TIA, 16 stroke mimics) demonstrated high sensitivity and specificity of a biomarker test based upon m/z 6640 alone or in conjunction with m/z 3699. As proteins detected through SELDI-TOF MS are not subject to digestion or tandem mass spectrometry, antibody immunodepletion and ELISA were attempted to determine the identity of m/z 6640. Immunodepletion techniques proved unsuccessful due to technical limitations; however ELISA coupled with SELDI-TOF MS revealed the identity of m/z 6640 as ApoC-1. The work undertaken in Chapter 3 demonstrated the ability of SELDI-TOF MS to detect differential plasma protein profiles and highlighted the role of lipid dysregulation in AIS and TIA incidence, in particular identifying ApoC-1 as a novel candidate biomarker for the diagnosis of acute cerebral ischaemia.

Carotid disease and ischaemic stroke share similar environmental and genetic risk factor profiles. Modifiable, non-modifiable and clinical risk factors contribute a 60% population attributable risk for ischaemic stroke, suggesting that the 40% gap in risk could be attributed to genetic variables. In an attempt to identify tissue-based risk biomarkers for large artery ischaemic stroke that could predict carotid plaque destabilisation and rupture, an integrative genomics study was undertaken to combine data from transcriptomic and genomic platforms. The results of this study were presented in Chapter 4.

Transcriptomic profiling was conducted using the Affymetrix Human Exon 1.0 ST array to identify gene expression differences between stenosing carotid plaque and non-stenosing carotid endarterectomy tissue regions taken from 7 symptomatic and 3 asymptomatic individuals. The exon array allows interrogation of both gene-level and exon-level RNA expression, thus allowing discovery of differentially expressed genes as well as characterisation of alternative gene splicing. PCA based upon global gene expression found distinct clustering of stenosing plaque and non-stenosing tissue, and further differentiation between symptomatic and asymptomatic stenosing plaque regions. Analysis of paired stenosing plaque and non-stenosing tissue regions from 5 symptomatic individuals is suggestive of alternative splicing of genes (as well as than gene expression) playing a role in plaque progression and a symptomatic plaque phenotype.

Paired symptomatic patient gene expression data was coupled with the results of GWAS meta-analysis of two cohorts (n=653) investigating the genetics of symptomatic carotid artery disease (>20% stenosis). Despite small sample sizes, functional enrichment analysis of both gene expression and

GWAS data using hypergeometric testing and GSEA analysis implicated the significance of cytoskeletal organisation and extracellular matrix remodelling in carotid plaque destabilisation and rupture.

Genes in these functional pathways common to both gene expression and GWAS were examined in an attempt to uncover differential gene expression or alternative splicing profiles that may be under the control of a genetic variant. The *TRIOBP* gene, involved in cytoskeletal organisation, was found to be significantly alternatively spliced in the gene expression dataset and also contain a number of SNPs significantly associated with symptomatic carotid stenosis upon GWAS meta-analysis. It was hypothesised that the sentinel SNP rs1074407, lying within a 3'UTR exonic region of *TRIOBP*, could be responsible for alternative splicing of the gene at exons 10 and 11.

Furthermore, the *LTBP4* gene, involved in extracellular matrix remodelling, was found to be significantly down-regulated in stenosing carotid plaques and significantly associated with symptomatic carotid stenosis upon GWAS meta-analysis. QTL mapping and differential expression analysis suggested that *LTBP4* is a gene whose expression may be under genetic control. Further suggestion of this hypothesis was obtained by the finding that LTBP4 was a key node in a protein-protein interaction network obtained from genes/transcripts significant in both the GWAS and gene expression studies. Within this network, LTBP4 was highly correlated with the expression of genes that it was directly or indirectly connected to, and could differentiate between non-stenosing tissue and unstable carotid plaque with high sensitivity and specificity. Thus the work in Chapter 4 established a methodology for biomarker discovery using an integrative genomics approach and identified novel candidate risk biomarkers for ischaemic stroke.

5.1 A Journey through the Molecular Mechanisms Underlying Ischaemic Stroke

The work detailed in this thesis has highlighted a number of key processes in the development of ischaemic stroke, from the involvement of cytoskeletal organisation and extracellular matrix remodelling in carotid plaque instability and rupture, to the role of dysregulated lipid metabolism for identifying cerebral ischaemic pathophysiology. While these findings are not in themselves unusual, they emphasise the significance of these mechanisms in the development of ischaemic stroke. More importantly, the work has identified new molecular targets associated with these mechanisms that can be further investigated for their potential as novel biomarkers for risk stratification or diagnosis of ischaemic stroke.

Systematic review and meta-analysis of the literature demonstrated the possibility of proteins involved in lipid metabolism to be used as diagnostic biomarkers for ischaemic stroke, a finding that was replicated with the discovery of the ability of ApoC-1 to differentiate cerebral ischaemic events

from stroke mimics using a proteomics approach. Furthermore, a recent GWAS found a highly significant association of *APOC-1* and *LDLR* with the presence of carotid intima media thickness and plaque (Bis et al. 2011), implicating the potential of proteins involved in lipid metabolism to predict risk of ischaemic stroke. Thus it is interesting to note that in this thesis, neither transcriptomic nor GWAS analysis identified a highly significant association of *APOC-1* or indeed any lipid metabolism-related gene with symptomatic carotid stenosis, even after data being subjected to functional enrichment analysis.

While lipid metabolism plays a key role in carotid plaque development, the concept of atherosclerosis as a passive accumulation of lipid in the arterial wall is outdated, and has been replaced by an understanding of the role of lipids in activating a cascade of inflammatory processes (Adibhatla et al. 2008). Lipoproteins are predominantly synthesised in the liver (rather than locally within an atherosclerotic plaque), and it seems that systemic hyperlipidaemia, rather than a local accumulation of lipids, is associated with progression of atherosclerosis. Hyperlipidaemia and an increase in systemic ApoC-1 result in an accumulation of atherogenic cholesterol-rich remnants in the circulation (Björkegren 2006). Using mouse models, Westerterp et al. (2007) demonstrated that systemic ApoC-1 increased atherosclerosis, while ApoC-1 production by macrophage foam cells within the arterial wall did not affect progression of atherosclerosis *in vivo.*

At the tissue level, lipid-associated proteins could drive the process of plaque destabilisation and rupture. At an early stage within the atherosclerotic lesion, infiltrating monocytes differentiate into $CD68^{+}$ macrophage lipid-rich foam cells. These cells have a high expression of ApoC-1 and ApoE, and are thought to produce proteases that affect the degradation of the extracellular matrix of the fibrous cap and cause plaque instability. As atherosclerosis progresses, vascular smooth muscle cells become senescent and undergo apoptosis, a process also thought to be initiated through ApoC-1 and ApoE expressing foam cells. Histological examination has found a high correlation between the presence of foam cells and the loss of vascular smooth muscle cells near the lipid-rich core, within the shoulder area of the atherosclerotic plaque prone to rupture. Thus local ApoC-1 and ApoE are strongly associated with the expansion of the necrotic core and destabilisation of the plaque through extracellular matrix remodelling (Eo et al. 2011). The failure to find significant associations between lipid-related pathways and symptomatic carotid plaque in the gene expression study undertaken in Chapter 4 may be due to the high number of dead cells present within the stenosing plaque tissue, impeding the examination of high integrity RNA from this region. Furthermore, gene expression analysis was based upon whole tissue sections rather than specific cell types; it is therefore possible that cell-specific mechanisms responsible for plaque rupture (e.g. apolipoprotein up-regulation in foam cells) may have been overlooked. Although GWAS has previously implicated a role for *APOC-1* in carotid disease (Bis et al. 2011), the slightly different study design undertaken in the GWAS performed

in Chapter 4 (studying symptomatic plaque rupture, rather than carotid IMT and plaque progression) may account for the lack of lipid-related pathways detected in the study.

Taken together, the findings suggest that lipid metabolism represents a key process in carotid plaque destabilisation and ischaemic stroke incidence, both at the systemic and tissue level. It can be hypothesised that the detection of ApoC-1 in Chapter 3 as a candidate diagnostic biomarker for cerebral ischaemia represents an underlying systemic hyperlipidaemic state that can differentiate TIA and ischaemic stroke patients from stroke mimics. While local lipid-related mechanisms within the atherosclerotic lesion could instigate plaque destabilisation and rupture, it is possible that at this late stage of atherosclerosis the role of lipid metabolism and apolipoprotein expression is overshadowed by the consequent processes of inflammation, cytoskeletal reorganisation and extracellular matrix remodelling, as demonstrated in Chapter 4.

5.2 A Workflow for Biomarker Discovery From –Omics-Derived Data

The statistical paradigms used to analyse –omics-derived data depend upon the nature of the data generated; for example, the quantitative measures of proteins or gene expression require different statistical treatment to allele calls arising from GWAS microarrays (Ghosh & Poisson 2009). Nonetheless, the work in this thesis has identified a number of common themes that need to be considered in order to facilitate biomarker discovery from –omics data.

5.2.1 Analysis of –Omics-Derived Data

Firstly, raw data requires pre-processing, application of stringent quality control measures and possible normalisation to limit experimental noise. Associations between genes or proteins and the phenotype of interest can be assessed using unsupervised or supervised methods. Unsupervised exploratory methods, such as PCA, utilise pattern discovery algorithms to understand the structure of a dataset, with elements being clustered independent of clinical status and then assessed for phenotypic trends. Semi-supervised clustering algorithms can identify gene expression or protein groups via functional enrichment analysis, in relationship to prior knowledge e.g. Gene Ontology, protein-protein interaction database. These methods can inform on co-expressed genes or co-regulated proteins and also help infer proteomic regulation from genomic regulation (Alawieh et al. 2012).

To directly examine the association between genes or proteins and a phenotype, statistical models and hypothesis testing can be used e.g. via Student's t-test, ANOVA, logistic regression. A number of precautions need to be taken during statistical analysis of –omics data. When each element is test independently, the issue of multiple comparison testing arises, namely that the number of elements is

so large that many spurious statistically significant associations can be expected to occur by chance alone. In order to account for this and reduce the number of false positive results, multiple comparison correction should be performed e.g. by FDR or Bonferroni correction.

The success of initial biomarker discovery studies lies in the depth of analysis and efficient selection of the most promising candidates for investigation. In many biomarker discovery studies, samples are often collected out of convenience rather than through a stringent protocol due to limited sample availability, which may introduce bias and confounding patterns. While statistical analysis may reveal candidate biomarkers of interest, it is imperative that these findings are validated in independent cohorts to establish reproducibility. It is also recommended that statistical models adjust for known clinical confounding parameters e.g. using univariate ANOVA models, to facilitate the conclusion that the candidate biomarker can provide information over and above that which is known by standard clinical factors. In heterogeneous disorders such as ischaemic stroke, the inclusion of comprehensive clinical and imaging patient data is essential to complement –omics profiling to allow the dissection of complex disease mechanisms.

5.2.2 Assessing the Significance of Candidate Biomarkers

In this thesis, candidate biomarkers have been assessed using standardised mean biomarker differences between case and control populations, odds ratios, and sensitivity and specificity analysis. Each of these methods requires careful interpretation in order to understand the true value of the biomarker under investigation.

Pepe et al. (2004) maintain that while most epidemiological studies use odds ratios to assess biomarker association to disease, odds ratios do not themselves characterise the discriminatory capacity of a biomarker. This could explain why, despite strong associations reported in the literature for many ischaemic stroke biomarkers, few were found to be significant upon meta-analysis in Chapter 2. Even markers with strong association e.g. CRP and glucose, did not exhibit adequate classification potential upon closer examination. Similarly, Pepe et al. (2004) describe the relationship between odds ratio and the AUC and demonstrate that an odds ratio of 3 may have little impact on the ROC curve, suggesting that an odds ratio of 16 (corresponding to AUC ~0.84) may be needed to achieve reasonable classification between cases and controls. They stipulate that biomarkers proposed for diagnosis or predicting risk in individual subjects must be held to a much higher standard than mere association with disease, and recommend that biomarker sensitivity and specificity should be assessed.

In Chapter 3 and Chapter 4, the diagnostic value of ApoC-1 and risk prediction value of LTBP4 were examined using ROC curves and AUC statistics. Both were found to demonstrate relatively high

sensitivity and specificity, indicating their clinical potential as biomarkers. Nonetheless, these analyses were based upon limited sample numbers and require robust validation. It is often also taken for granted that the sensitivity and specificity of a biomarker test are not subject to alteration. However, these properties are very much dependent upon the population in which they are utilised. For example, the sensitivity of a test is likely to be greater in severe rather than mild cases of a disease, while specificity is dependent upon characteristics of control cases such as age, gender or the prevalence of associated risk factors (Cook 2007). Similarly, positive and negative predictive values are reliant upon study design (Elwood 2002; Pearl 2002; Rockhill 2002). Hence the value of a biomarker test needs to be evaluated within the context in which it is used.

5.3 Future Directions

Morrow and de Lemos (2007) recommended three main criteria that novel biomarkers should meet: 1) easy measurement by the clinician and 2) addition of valuable new information which 3) allows more effective disease management.

The work detailed in this thesis has led to the identification of candidate biomarkers of ischaemic stroke. However, these findings represent only the first steps of a long journey into the development of clinically relevant biomarker tests. The Early Detection Research Network (Pepe et al. 2001) describes five phases of biomarker development, of which the first phase – exploratory research – has been pursued in this thesis. The next four phases, mirroring a clinical trials process, include developing an assay with clinically reproducible results (phase 2), optimising the time interval in which to screen for the biomarker (phase 3), examining biomarker characteristics (phase 4) and assessing whether the biomarker tangibly improves clinical practice (phase 5). These experimental phases must now be undertaken in order to validate the utility of the candidate biomarkers identified in the thesis.

Identification, qualification, and application of biomarkers remains the holy grail of the current -omics paradigm. It may seem that with the emergence of unbiased, –omics approaches to experimental design, traditional hypothesis-driven, single molecule-based approaches will soon become outdated. On the contrary, the molecular mechanisms underlying candidate biomarkers identified through -omics approaches require robust validation using candidate-based techniques and can only be understood through hypothesis-based investigation. Thus the two designs should be viewed as complimentary tools for biomarker discovery.

Although genomic, transcriptomic and proteomic techniques have been successfully employed in this thesis to identify novel candidate biomarkers for ischaemic stroke, the role of these gene, transcript and protein components on the organism as a whole remains unclear. Large population-based studies

have established that chronic diseases such as cerebrovascular disease occur in a context of interaction between multiple genes, environmental risk factors and epigenetic changes (Knox 2010). New awareness of the non-linear, complex dynamics of living systems has led to the evolution of systems biology, whereby genomic, transcriptomic and proteomic information can be combined to understand disease pathophysiology at the level of molecular pathways, regulatory networks, cells, tissues, organs and ultimately the whole organism. Thus –omics provide a bottom-up approach for the identification and global measurement of molecular components, while systems biology modelling provides a top-down approach to form integrative models of disease (Butcher et al. 2004). This holistic approach, integrating massive molecular, clinical and pathological data with novel statistical and bioinformatics approaches, could have a major impact on the management of ischaemic stroke and could pave the way for individualised therapy in the future (Dominiczak et al. 2010).

A rudimentary attempt at systems biology was undertaken in Chapter 4 of this thesis. Nonetheless, a true systems biology approach requires a much higher level of computational sophistication. It must also be noted that at present, successful biomarker discovery from systems biology approaches requires resolution of conceptual as well as practical challenges. Findings from the international Encyclopedia of DNA Elements (ENCODE) project have overturned the classic definition of the gene and Francis Crick's 'central dogma', as well as highlighting the necessity to appreciate the complexity of the genome in experimental study design (Gerstein et al. 2007; Pearson 2006). Protein-coding genes comprise less than 2% of the eukaryotic genome and ~98% of transcriptional output represents non-coding RNAs (Vemuganti 2010). Furthermore, recent literature suggests that <40% of cellular protein levels can be predicted from mRNA measurements, supporting the conclusion that mRNA levels should not be used as surrogates for corresponding protein levels (Kendrick 2012). Given that proteins are viewed as the functional molecules primarily driving disease pathogenesis, this information queries the use of mRNA expression levels as biomarkers, and emphasises the limitations associated with the assumptions made when using an integrative genomics approach. While the integration of genetic and protein biomarkers is a promising strategy for stroke management (Piran et al. 2012), until the relationships between the genome, transcriptome and phenotype are better understood it can be argued that proteins represent the most reliable molecular elements for investigating disease association and biomarker discovery.

5.4 Final Remarks

The work in this thesis has confirmed the hypothesis that novel candidate biomarkers for ischaemic stroke are amenable to detection, using innovative –omics approaches for biomarker discovery. The studies have generated new hypotheses to test the value of these candidates as relevant risk prediction and diagnostic biomarkers for ischaemic stroke. In particular, the work contributes the following to the field of biomarker research for ischaemic stroke:

- Highlights existing candidate diagnostic and prognostic biomarkers that demonstrate significant association with ischaemic stroke and require additional investigation to evaluate clinical utility.
- Identifies common limitations of biomarker study design and provides guidelines for improved biomarker discovery using –omics approaches.
- Demonstrates that admission glucose level is not a valuable prognostic biomarker for the prediction of symptomatic intracerebral haemorrhage following intravenous rt-PA thrombolysis for ischaemic stroke, and advocates the extension of the international glucose threshold for thrombolysis to beyond 22mmol/L, thus allowing a greater number of patients to be considered for treatment.
- Ascertains the ability to detect novel blood-and tissue-based biomarkers of ischaemic stroke by utilising -omics-based approaches to biomarker discovery, and highlights the complexity underlying carotid atherosclerosis and ischaemic stroke pathophysiology at both the gene and protein level.
- Reaffirms a key role for lipid metabolism in the development of ischaemic stroke and the function of cytoskeletal organisation and extracellular matrix remodelling in determining plaque instability.
- Establishes the need for deeper investigation of alternative splicing mechanisms in driving carotid plaque progression and instability.
- Identifies novel candidate biomarkers that can be employed in single- or multi-panel tests to distinguish acute cerebral ischaemia from stroke mimics or predict carotid plaque instability and rupture.

In conclusion, this thesis provides the scientific community with new recommendations and methodologies for biomarker discovery for ischaemic stroke and identifies novel candidate biomarkers for diagnosis and risk stratification. Research is now warranted to further characterise these candidate biomarkers and assess their clinical potential.

REFERENCES

Abdul-Salam, V.B. et al., 2010. Identification and assessment of plasma lysozyme as a putative biomarker of atherosclerosis. *Arterioscler Thromb Vasc Biol*, 30(5), pp.1027–1033.

Abo-Zaid, G. et al., 2013. Individual participant data meta-analyses should not ignore clustering. *J Clin Epidemiol*, 66(8), pp.865–873 e4.

Adams Jr., H.P. et al., 1993. Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. *Stroke*, 24(1), pp.35–41.

Adamson, J., Beswick, A. & Ebrahim, S., 2004. Is stroke the most common cause of disability? *J Stroke Cerebrovasc Dis*, 13(4), pp.171–177.

Adibhatla, R.M., Dempsy, R. & Hatcher, J.F., 2008. Integration of cytokine biology and lipid metabolism in stroke. *Frontiers in bioscience : a journal and virtual library*, 13, pp.1250–70.

Aebersold, R. & Mann, M., 2003. Mass spectrometry-based proteomics. *Nature*, 422(6928), pp.198–207.

Affymetrix, 2006. GeneChip Exon Array System for Human, Mouse and Rat. Data Sheet. Part Number 702005 Rev.2. , 2013.

Agardh, H.E. et al., 2011. Expression of fatty acid-binding protein 4/aP2 is correlated with plaque instability in carotid atherosclerosis. *Journal of internal medicine*, 269(2), pp.200–10.

Ahmed, R.M. et al., 2010. Carotid endarterectomy for symptomatic, but "haemodynamically insignificant" carotid stenosis. *Eur J Vasc Endovasc Surg*, 40(4), pp.475–482.

Ahn, J.H. et al., 2011. The usefulness of albumin-adjusted ischemia-modified albumin index as early detecting marker for ischemic stroke. *Neurological sciences : official journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology*, 32(1), pp.133–8.

Airas, L., Lindsberg, P.J., Karjalainen-Lindsberg, M.L., et al., 2008. Vascular adhesion protein-1 in human ischaemic stroke. *Neuropathology and Applied Neurobiology*, 34(4), pp.394–402.

Airas, L., Lindsberg, P.J., Karjalainen-Lindsberg, M.-L., et al., 2008. Vascular adhesion protein-1 in human ischaemic stroke. *Neuropathology and applied neurobiology*, 34(4), pp.394–402.

Alaoui-Jamali, M.A. & Xu, Y.-J., 2006. Proteomic technology for biomarker profiling in cancer: an update. *Journal of Zhejiang University Science B*, 7(6), pp.411–420.

Alawieh, A. et al., 2012. Systems biology, bioinformatics, and biomarkers in neuropsychiatry. *Front Neurosci*, 6, p.187.

Albers Caplan, L., Easton,J. D., Fayad, P. B., Mohr, J. P., Saver, J. L., Sherman, D. G. for the TIA Worknig Group, G., 2002. Transient Ischemic Attack — Proposal for a New Definition. *New England Journal of Medicine*, 348(16), pp.1607–1609.

Allard, L. et al., 2004. ApoC-I and ApoC-III as potential plasmatic markers to distinguish between ischemic and hemorrhagic stroke. *Proteomics*, 4(8), pp.2242–2251.

Allard, L., Discovery of four potential markers for the early diagnosis of stroke. Presented at 21st l'ACORATA Biology Conference, Paris, October 2004. *Clinical Chemistry*, pp.1–1.

Allard, L., Burkhard, P.R., Lescuyer, P., Burgess, J.A., et al., 2005. PARK7 and nucleoside diphosphate kinase A as plasma markers for the early diagnosis of stroke. *Clinical Chemistry*, 51(11), pp.2043–2051.

Allard, L., Burkhard, P.R., Lescuyer, P., Burgess, J. a, et al., 2005. PARK7 and nucleoside diphosphate kinase A as plasma markers for the early diagnosis of stroke. *Clinical chemistry*, 51(11), pp.2043–51.

Allison, P.D., 1999. *Logistic Regression Using the ASA System: Theory and Application*, Cary, NC. : SAS Institute Inc.

Altman, D.G. & Bland, J.M., 1996. Detecting skewness from summary information. *BMJ*, 313(7066), p.1200.

Alvarez-Sabin, J. et al., 2003. Effects of admission hyperglycemia on stroke outcome in reperfused tissue plasminogen activator--treated patients. *Stroke*, 34(5), pp.1235–1241.

Alvarez-Sabin, J. et al., 2004. Impact of admission hyperglycemia on stroke outcome after thrombolysis: risk stratification in relation to time to reperfusion. *Stroke*, 35(11), pp.2493–2498.

Amarenco, P. et al., 2009. New approach to stroke subtyping: the A-S-C-O (phenotypic) classification of stroke. *Cerebrovascular diseases (Basel, Switzerland)*, 27(5), pp.502–8.

Anderson, C.D. et al., 2013. Common variants within oxidative phosphorylation genes influence risk of ischemic stroke and intracerebral hemorrhage. *Stroke*, 44(3), pp.612–619.

Anderson, N.L. & Anderson, N.G., 2002. The human plasma proteome: history, character, and diagnostic prospects. *Molecular cellular proteomics MCP*, 1(11), pp.845–867.

Aoki, T. et al., 2002. Blood-brain barrier disruption and matrix metalloproteinase-9 expression during reperfusion injury: mechanical versus embolic focal ischemia in spontaneously hypertensive rats. *Stroke: A Journal of Cerebral Circulation*, 33(11), pp.2711–2717.

Arrell, D.K., Neverova, I. & Van Eyk, J.E., 2001. Cardiovascular Proteomics : Evolution and Potential. *Circulation Research*, 88(8), pp.763–773.

Arsava, E.M. et al., 2010. The Causative Classification of Stroke system: an international reliability and optimization study. *Neurology*, 75(14), pp.1277–1284.

Ashburner, M. et al., 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*, 25(1), pp.25–29.

Asplund, K., 2005. What MONICA told us about stroke. *Lancet Neurol*, 4(1), pp.64–68.

Ay, H. et al., 2007. A computerized algorithm for etiologic classification of ischemic stroke: the Causative Classification of Stroke System. *Stroke; a journal of cerebral circulation*, 38(11), pp.2979–84.

Babuin, L. & Jaffe, A.S., 2005. Troponin: the biomarker of choice for the detection of cardiac injury. *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne*, 173(10), pp.1191–202.

Baker, S.G., Kramer, B.S. & Srivastava, S., 2002. Markers for early detection of cancer: statistical guidelines for nested case-control studies. *BMC Med Res Methodol*, 2, p.4.

Baker, W.H. et al., 2000. Effect of contralateral occlusion on long-term efficacy of endarterectomy in the asymptomatic carotid atherosclerosis study (ACAS). ACAS Investigators. *Stroke*, 31(10), pp.2330–2334.

Bamford, J. et al., 1991. Classification and natural history of clinically identifiable subtypes of cerebral infarction. *Lancet*, 337(8756), pp.1521–6.

Barber, M.J. et al., 2010. Genome-Wide Association of Lipid-Lowering Response to Statins in Combined Study Populations. *PLoS One*, 5(3), p.e9763.

Barres, R. et al., 2012. Acute exercise remodels promoter methylation in human skeletal muscle. *Cell Metab*, 15(3), pp.405–411.

Baselga, J. et al., 1998. Recombinant humanized anti-HER2 antibody (Herceptin) enhances the antitumor activity of paclitaxel and doxorubicin against HER2/neu overexpressing human breast cancer xenografts. *Cancer Research*, 58(13), pp.2825–2831.

Beaussier, H. et al., 2008. Carotid plaque, arterial stiffness gradient, and remodeling in hypertension. *Hypertension*, 52(4), pp.729–736.

Benjamini, Y. & Hochberg, Y., 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1), pp.289–300.

Benjamini, Y. & Yekutieli, D., 2001. The Control of the False Discovery Rate in Multiple Testing under Dependency. *The Annals of Statistics*, 29(4), pp.1165–1188.

Bentley, P. et al., 2010. Causal relationship of susceptibility genes to ischemic stroke: comparison to ischemic heart disease and biochemical determinants. *PLoS One*, 5(2), p.e9136.

Berglund, L.M. et al., 2012. NFAT regulates the expression of AIF-1 and IRT-1: yin and yang splice variants of neointima formation and atherosclerosis. *Cardiovasc Res*, 93(3), pp.414–423.

Bevan, S. et al., 2012. Genetic heritability of ischemic stroke and the contribution of previously reported candidate gene and genomewide associations. *Stroke*, 43(12), pp.3161–3167.

Bidzhekov, K. et al., 2012. microRNA expression signatures and parallels between monocyte subsets and atherosclerotic plaque in humans. *Thromb Haemost*, 107(4), pp.619–625.

Biomarkers Definitions Working Group, 2001. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clinical Pharmacology & Therapeutics*, 69(3), pp.89–95.

Bio-Rad Laboratories Inc., 2008. *Biomarker discovery to using SELDI technology: a guide to successful study and experimental design. Bulletin 5642 US/EG Rev A.*,

Bis, J.C. et al., 2011. Meta-analysis of genome-wide association studies from the CHARGE consortium identifies common variants associated with carotid intima media thickness and plaque. *Nat Genet*, 43(10), pp.940–947.

Bjorkegren, J. et al., 2002. Postprandial enrichment of remnant lipoproteins with apoC-I in healthy normolipidemic men with early asymptomatic atherosclerosis. *Arterioscler Thromb Vasc Biol*, 22(9), pp.1470–1474.

Björkegren, J., 2006. Dual roles of apolipoprotein CI in the formation of atherogenic remnants. *Current atherosclerosis reports*, 8(1), pp.1–2.

Blanco-Colio, L.M. et al., 2006. Biology of atherosclerotic plaques: what we are learning from proteomic analysis. *Cardiovascular Research*, 72(1), pp.18–29.

Bobryshev, Y. V et al., 2008. Matrix vesicles in the fibrous cap of atherosclerotic plaque: possible contribution to plaque rupture. *J Cell Mol Med*, 12(5B), pp.2073–2082.

De Bock, M. et al., 2010. Challenges for Biomarker Discovery in Body Fluids Using SELDI-TOF-MS. *Journal of biomedicine biotechnology*, 2010, p.906082.

Bolstad, B.M. et al., 2003. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics (Oxford, England)*, 19(2), pp.185–93.

Bons, J. et al., 2006. Application of SELDI-TOF-MS in protein profiling: Promises and pitfalls. *Jugoslovenska Medicinska Biohemija-yugoslav Medical Biochemistry*, 25(3), pp.201–210.

Borlongan, C. V et al., 2012. Permeating the blood brain barrier and abrogating the inflammation in stroke: implications for stroke therapy. *Curr Pharm Des*, 18(25), pp.3670–3676.

Bossuyt, P.M. & Leeflang, M.M., 2008. Chapter 6: Developing Criteria for Including Studies. In T. C. Collaboration, ed. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*.

Bossuyt, P. M. & Leeflang, M.M., 2008. Chapter 6: Developing Criteria for Including Studies. In *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 0.4. The Cochrane Collaboration*.

Brass, L.M. et al., 1992. A study of twins and stroke. *Stroke*, 23(2), pp.221–223.

Brazzelli, M. et al., 2009. Magnetic resonance imaging versus computed tomography for detection of acute vascular lesions in patients presenting with stroke symptoms. *Cochrane Database Syst Rev*, (4), p.CD007424.

Brea, D. et al., 2011. Proteomic analysis shows differential protein expression in endothelial progenitor cells between healthy subjects and ischemic stroke patients. *Neurological research*, 33(10), pp.1057–63.

Brea, D. et al., 2009. Usefulness of haptoglobin and serum amyloid A proteins as biomarkers for atherothrombotic ischemic stroke diagnosis confirmation. *Atherosclerosis*, 205(2), pp.561–7.

Brott, T. et al., 1989. Measurements of acute cerebral infarction: a clinical examination scale. *Stroke: A Journal of Cerebral Circulation*, 20(7), pp.864–870.

Brown, D.L. et al., 2005. Survey of emergency physicians about recombinant tissue plasminogen activator for acute ischemic stroke. *Annals of Emergency Medicine*, 46(1), pp.56–60.

Brown, R.D. et al., 1996. Stroke incidence, prevalence, and survival: secular trends in Rochester, Minnesota, through 1989. *Stroke*, 27(3), pp.373–380.

Bruno, A. et al., 2002. Admission glucose level and clinical outcomes in the NINDS rt-PA Stroke Trial. *Neurology*, 59(5), pp.669–74.

Buechler CA, US), Maisel, Alan (Del Mar, CA, US), Anderberg, Joseph Michael (Encinitas, CA, US), Mcpherson, Paul H. (Encinitas, CA, US), Dahlen, Jeffrey R. (San Diego, CA, US), Kirchick, Howard J. (San Diego, CA, US), K.F. (Rancho S.F., 2004. Markers for differential diagnosis and methods of use thereof. , (20040253637).

Burke, H.B., 2005. Proteomics: analysis of spectral data. *Cancer Inform*, 1, pp.15–24.

Bush, W.S. & Haines, J., 2010. Overview of linkage analysis in complex traits. *Current protocols in human genetics / editorial board, Jonathan L. Haines ... [et al.]*, Chapter 1, p.Unit 1.9.1–18.

Bush, W.S. & Moore, J.H., 2012. Chapter 11: Genome-wide association studies. *PLoS Comput Biol*, 8(12), p.e1002822.

Butcher, E.C., Berg, E.L. & Kunkel, E.J., 2004. Systems biology in drug discovery. *Nat Biotechnol*, 22(10), pp.1253–1259.

Cagnin, S. et al., 2009. Reconstruction and functional analysis of altered molecular pathways in human atherosclerotic arteries. *BMC Genomics*, 10, p.13.

Cantor, R.M., Lange, K. & Sinsheimer, J.S., 2010. Prioritizing GWAS results: A review of statistical methods and recommendations for their application. *Am J Hum Genet*, 86(1), pp.6–22.

Capes, S.E. et al., 2001. Stress Hyperglycemia and Prognosis of Stroke in Nondiabetic and Diabetic Patients: A Systematic Overview. *Stroke*, 32(10), pp.2426–2432.

Chang, C.T. et al., 2013. Oxidized ApoC1 on MALDI-TOF and glycated-ApoA1 band on gradient gel as potential diagnostic tools for atherosclerotic vascular disease. *Clin Chim Acta*, 420, pp.69–75.

Chen, P.H. et al., 2012. Classifying Ischemic Stroke, from TOAST to CISS. *CNS Neurosci Ther*, 18(6), pp.452–456.

Chinn, S., 2000. A simple method for converting an odds ratio to effect size for use in meta-analysis. *Statistics in Medicine*, 19(22), pp.3127–3131.

Clark, A.G. & Li, J., 2007. Conjuring SNPs to detect associations. *Nature genetics*, 39(7), pp.815–6.

Clark, N.R. & Ma'ayan, A., 2011. Introduction to statistical methods to analyze large data sets: principal components analysis. *Sci Signal*, 4(190), p.tr3.

Cline, M.S. et al., 2005. ANOSVA: a statistical method for detecting splice variation from expression data. *Bioinformatics*, 21 Suppl 1, pp.i107–15.

Cocho, D. et al., 2006. Pretreatment hemostatic markers of symptomatic intracerebral hemorrhage in patients treated with tissue plasminogen activator. *Stroke: A Journal of Cerebral Circulation*, 37(4), pp.996–999.

Cohn, J.S. et al., 2003. Plasma concentration and lipoprotein distribution of ApoC-I is dependent on ApoE genotype rather than the Hpa I ApoC-I promoter polymorphism. *Atherosclerosis*, 169(1), pp.63–70.

Cole, J.W. et al., 2012. Rare Variants in Ischemic Stroke: An Exome Pilot Study. *PLoS One*, 7(4), p.e35591.

Comuzzie, A.G. et al., 2012. Novel Genetic Loci Identified for the Pathophysiology of Childhood Obesity in the Hispanic Population. *PLoS One*, 7(12), p.e51954.

Congrains, A. et al., 2012. Genetic variants at the 9p21 locus contribute to atherosclerosis through modulation of ANRIL and CDKN2A/B. *Atherosclerosis*, 220(2), pp.449–455.

Cook, N.R., 2007a. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation*, 115(7), pp.928–935.

Cook, N.R., 2007b. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation*, 115(7), pp.928–935.

Cookson, W. et al., 2009. Mapping complex disease traits with global gene expression. *Nat Rev Genet*, 10(3), pp.184–194.

Counsell, C. et al., 2002. Predicting outcome after acute and subacute stroke: development and validation of new prognostic models. *Stroke*, 33(4), pp.1041–1047.

Counsell, C., Dennis, M. & McDowall, M., 2004. Predicting functional outcome in acute stroke: comparison of a simple six variable model with other predictive systems and informal clinical prediction. *J Neurol Neurosurg Psychiatry*, 75(3), pp.401–405.

Crouse, J.R. et al., 1996. Risk factors and segment-specific carotid arterial enlargement in the Atherosclerosis Risk in Communities (ARIC) cohort. *Stroke*, 27(1), pp.69–75.

Cruchaga, C. et al., 2013. GWAS of cerebrospinal fluid tau levels identifies risk variants for Alzheimer's disease. *Neuron*, 78(2), pp.256–68.

Cuadrado, E. et al., 2010. The proteome of human brain after ischemic stroke. *Journal of neuropathology and experimental neurology*, 69(11), pp.1105–15.

Cuadrado, E. et al., 2009. Vascular MMP-9/TIMP-2 and neuronal MMP-10 up-regulation in human brain after stroke: a combined laser microdissection and protein array study. *Journal of proteome research*, 8(6), pp.3191–7.

Cucchiara, B. et al., 2008. A risk score to predict intracranial hemorrhage after recombinant tissue plasminogen activator for acute ischemic stroke. *J Stroke Cerebrovasc Dis*, 17(6), pp.331–333.

Cucchiara, B. et al., 2009. Factors Associated With Intracerebral Hemorrhage After Thrombolytic Therapy for Ischemic Stroke Pooled Analysis of Placebo Data From the Stroke-Acute Ischemic NXY Treatment (SAINT) I and SAINT II Trials. *Stroke: A Journal of Cerebral Circulation*, 40(9), pp.3067–3072.

Dahl, T.B. et al., 2007. Increased expression of visfatin in macrophages of human unstable carotid and coronary atherosclerosis: possible role in inflammation and plaque destabilization. *Circulation*, 115(8), pp.972–980.

Dalman, M.R. et al., 2012. Fold change and p-value cutoffs significantly alter microarray interpretations. *BMC bioinformatics*, 13 Suppl 2, p.S11.

Dambinova, S.A. et al., 2012. Diagnostic potential of the NMDA receptor peptide assay for acute ischemic stroke. *PLoS One*, 7(7), p.e42362.

Dambinova, S.A., Khounteev, G.A. & Skoromets, A.A., 2002. Multiple panel of biomarkers for TIA/stroke evaluation. *Stroke: A Journal of Cerebral Circulation*, 33(5), pp.1181–1182.

Dawson, J. et al., 2012. Urinary proteomics to support diagnosis of stroke. *PloS one*, 7(5), p.e35879.

Dayon, L. et al., 2011. Brain extracellular fluid protein changes in acute stroke patients. *Journal of proteome research*, 10(3), pp.1043–51.

Dayon, L. et al., 2008. Relative quantification of proteins in human cerebrospinal fluids by MS/MS using 6-plex isobaric tags. *Analytical Chemistry*, 80(8), pp.2921–2931.

Deeks JL, Altman DG, B.M., 2001. *Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. In: Egger M, Smith GD, Altman DG, editors. Systematic Reviews in Health Care. Meta-analysis in Context.*, London, UK: BMJ Publishing Group.

Degraba, T.J. et al., 2011. Biomarker discovery in serum from patients with carotid atherosclerosis. *Cerebrovasc Dis Extra*, 1(1), pp.115–129.

Delano D. Galver L., Rosenow C., E.M., 2010. Array difference in genomic coverage and data quality impact GWAS success. Illumina White Paper: DNA Analysis 2010. *Illumina White Paper: DNA Analysis*.

DeLong, E.R., DeLong, D.M. & Clarke-Pearson, D.L., 1988. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, 44(3), pp.837–45.

Demchuk, a. M. et al., 1999. Serum Glucose Level and Diabetes Predict Tissue Plasminogen Activator Related Intracerebral Hemorrhage in Acute Ischemic Stroke. *Stroke*, 30(1), pp.34–39.

Demchuk, A.M. et al., 1999. Serum glucose level and diabetes predict tissue plasminogen activator-related intracerebral hemorrhage in acute ischemic stroke. *Stroke: A Journal of Cerebral Circulation*, 30(1), pp.34–39.

Derex, L. & Nighoghossian, N., 2008. Intracerebral haemorrhage after thrombolysis for acute ischaemic stroke: an update. *Journal of Neurology, Neurosurgery & Psychiatry*, 79(10), pp.1093–1099.

DerSimonian, R. & Laird, N., 1986. Meta-analysis in clinical trials. *Control Clin Trials*, 7(3), pp.177–188.

Dharmasaroja, P.A. et al., 2012. Intracerebral hemorrhage following intravenous thrombolysis in Thai patients with acute ischemic stroke. *J Clin Neurosci*, 19(6), pp.799–803.

Diamandis, E.P., 2002. Proteomic patterns in serum and identification of ovarian cancer. *The Lancet*, 360(9327), p.170.

Diez, D. et al., 2010. The use of network analyses for elucidating mechanisms in cardiovascular disease. *Mol Biosyst*, 6(2), pp.289–304.

Ding, H. et al., 2010. Confirmation of genomewide association signals in Chinese Han population reveals risk loci for ischemic stroke. *Stroke; a journal of cerebral circulation*, 41(1), pp.177–80.

Dirnagl, U., Iadecola, C. & Moskowitz, M.A., 1999. Pathobiology of ischaemic stroke: an integrated view. *Trends Neurosci*, 22(9), pp.391–397.

Doehner, W. et al., 2012. Elevated plasma levels of neuropeptide proenkephalin a predict mortality and functional outcome in ischemic stroke. *Journal of the American College of Cardiology*, 60(4), pp.346–54.

Doig, D. & Brown, M.M., 2012a. Carotid Stenting Versus Endarterectomy. *Annual Review of Medicine*.

Doig, D. & Brown, M.M., 2012b. Carotid stenting versus endarterectomy. *Annual review of medicine*, 63, pp.259–76.

Dominiczak, A.F. et al., 2010. Systems biology to battle vascular disease. *Nephrol Dial Transplant*, 25(4), pp.1019–1022.

Easton, J.D. et al., 2009. Definition and evaluation of transient ischemic attack: a scientific statement for healthcare professionals from the American Heart Association/American Stroke Association Stroke Council; Council on Cardiovascular Surgery and Anesthesia; Council on Cardio. *Stroke*, 40(6), pp.2276–2293.

Efron, B. & Tibshirani, R., 2002. Empirical bayes methods and false discovery rates for microarrays. *Genetic epidemiology*, 23(1), pp.70–86.

Egger, M. et al., 1997. Bias in meta-analysis detected by a simple, graphical test. *BMJ*, 315(7109), pp.629–634.

Elbers, C.C. et al., 2009. Using genome-wide pathway analysis to unravel the etiology of complex diseases. *Genet Epidemiol*, 33(5), pp.419–431.

Elwood, M., 2002. Proteomic patterns in serum and identification of ovarian cancer. *Lancet*, 360(9327), p.170; author reply 170–1.

Emerging Risk Factors Collaboration et al., 2010. Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies. *Lancet*, 375(9733), pp.2215–2222.

Eo, H.-S. et al., 2011. Association with inflammatory cells and apolipoproteins to the progression of atherosclerosis. *Journal of the Korean Surgical Society*, 80(4), pp.289–96.

European Carotid Surgery Trialists' Collaborative Group [ECST], 1991. MRC European Carotid Surgery Trial: interim results for symptomatic patients with severe (70-99%) or with mild (0-29%) carotid stenosis. *Lancet*, 337(8752), pp.1235–1243.

Feigin, V.L. et al., 2009. Worldwide stroke incidence and early case fatality reported in 56 population-based studies: a systematic review. *Lancet Neurol*, 8(4), pp.355–369.

Feuerstein, G.Z. & Chavez, J., 2009. Translational medicine for stroke drug discovery: the pharmaceutical industry perspective. *Stroke: A Journal of Cerebral Circulation*, 40(3 Suppl), pp.S121–S125.

Florian-Kujawski, M. et al., 2004. Biomarker profiling of plasma from acute coronary syndrome patients. Application of ProteinChip Array analysis. *Int Angiol*, 23(3), pp.246–254.

Foerch, C. et al., 2012. Diagnostic accuracy of plasma glial fibrillary acidic protein for differentiating intracerebral hemorrhage and cerebral ischemia in patients with symptoms of acute stroke. *Clinical chemistry*, 58(1), pp.237–45.

Folkersen, L. et al., 2009. Relationship between CAD Risk Genotype in the Chromosome 9p21 Locus and Gene Expression. Identification of Eight New <italic>ANRIL</italic> Splice Variants. *PLoS One*, 4(11), p.e7677.

Forrest, S.T. et al., 2004. Intron retention generates a novel Id3 isoform that inhibits vascular lesion formation. *J Biol Chem*, 279(31), pp.32897–32903.

Fox, C.S. et al., 2012. Genome-wide association for abdominal subcutaneous and visceral adipose reveals a novel locus for visceral fat in women. *PLoS Genet*, 8(5), p.e1002695.

Frankish, A. et al., 2012. The importance of identifying alternative splicing in vertebrate genome annotation. *Database*, 2012.

Furlan, M. et al., 1996. Spontaneous neurological recovery after stroke and the fate of the ischemic penumbra. *Ann Neurol*, 40(2), pp.216–226.

Gamazon, E.R. et al., 2010. SCAN: SNP and copy number annotation. *Bioinformatics (Oxford, England)*, 26(2), pp.259–62.

Gasecki, A.P. et al., 1995. Long-term prognosis and effect of endarterectomy in patients with symptomatic severe carotid stenosis and contralateral carotid stenosis or occlusion: results from NASCET. North American Symptomatic Carotid Endarterectomy Trial (NASCET) Group. *J Neurosurg*, 83(5), pp.778–782.

Gerstein, M.B. et al., 2007. What is a gene, post-ENCODE? History and updated definition. *Genome Res*, 17(6), pp.669–681.

Ghazalpour, A. et al., 2011. Comparative analysis of proteome and transcriptome variation in mouse. *PLoS Genet*, 7(6), p.e1001393.

Ghosh, D. & Poisson, L.M., 2009. "Omics" data and levels of evidence for biomarker discovery. *Genomics*, 93(1), pp.13–6.

Gibson, G., 2011. Rare and common variants: twenty arguments. *Nat Rev Genet*, 13(2), pp.135–145.

Giles, M.F. et al., 2010. Addition of brain infarction to the ABCD2 Score (ABCD2I): a collaborative analysis of unpublished data on 4574 patients. *Stroke; a journal of cerebral circulation*, 41(9), pp.1907–13.

Gillette, M.A., Mani, D.R. & Carr, S.A., 2005. Place of pattern in proteomic biomarker discovery. *J Proteome Res*, 4(4), pp.1143–1154.

Glickman, S.W. et al., 2011. Discriminative capacity of biomarkers for acute stroke in the emergency department. *The Journal of emergency medicine*, 41(3), pp.333–9.

Golder, S., Loke, Y.K. & Bland, M., 2011. Meta-analyses of adverse effects data derived from randomised controlled trials as compared to observational studies: methodological overview. *PLoS medicine*, 8(5), p.e1001026.

Gonzalez-Garcia, S. et al., 2012. Serum neuron-specific enolase and S100 calcium binding protein B biomarker levels do not improve diagnosis of acute stroke. *J R Coll Physicians Edinb*, 42(3), pp.199–204.

Grau, A.J. et al., 2001. Risk factors, outcome, and treatment in subtypes of ischemic stroke: the German stroke data bank. *Stroke*, 32(11), pp.2559–2566.

Greenland, P. & O'Malley, P.G., 2005. When is a new prediction marker useful? A consideration of lipoprotein-associated phospholipase A2 and C-reactive protein for stroke risk. *Archives of internal medicine*, 165(21), pp.2454–6.

Gretarsdottir, S. et al., 2008. Risk variants for atrial fibrillation on chromosome 4q25 associate with ischemic stroke. *Ann Neurol*, 64(4), pp.402–409.

Gretarsdottir, S. et al., 2003. The gene encoding phosphodiesterase 4D confers risk of ischemic stroke. *Nat Genet*, 35(2), pp.131–138.

Grotta, J.C. et al., 2001. Clinical deterioration following improvement in the NINDS rt-PA Stroke Trial. *Stroke: A Journal of Cerebral Circulation*, 32(3), pp.661–668.

Gschwendtner, A. et al., 2009. Sequence variants on chromosome 9p21.3 confer risk for atherosclerotic stroke. *Ann Neurol*, 65(5), pp.531–539.

Guerrera, I.C. & Kleiner, O., 2005. Application of mass spectrometry in proteomics. *Bioscience Reports*, 25(1-2), pp.71–93.

Hacke, W. et al., 2004. Association of outcome with early stroke treatment: pooled analysis of ATLANTIS, ECASS, and NINDS rt-PA stroke trials. *Lancet*, 363(9411), pp.768–74.

Hall, H.S. & Bassiouny, H.&, 2012. Pathophysiology of Carotid Atherosclerosis. In A. Nicolaides Beach, K. W., Kyriacou, E., Pattichis C. S., ed. *Ultrasound and Carotid Bifurcation Atherosclerosis*. New York: Springer.

Halliday, A. et al., 2010. 10-year stroke prevention after successful carotid endarterectomy for asymptomatic stenosis (ACST-1): a multicentre randomised trial. *Lancet*, 376(9746), pp.1074–1084.

Hamberger, A. et al., 1990. The contribution from the choroid plexus and the periventricular CNS to amino acids and proteins in the human CSF. *Neurochemical Research*, 15(3), pp.307–312.

Hand, P.J. et al., 2006. Distinguishing between stroke and mimic at the bedside: the brain attack study. *Stroke: A Journal of Cerebral Circulation*, 37(3), pp.769–775.

Harbison, J. et al., 2003. Diagnostic accuracy of stroke referrals from primary care, emergency room physicians, and ambulance staff using the face arm speech test. *Stroke*, 34(1), pp.71–76.

Hardy, J., Trabzuni, D. & Ryten, M., 2009. Whole genome expression as a quantitative trait. *Biochem Soc Trans*, 37(Pt 6), pp.1276–1277.

Hasan, N., 2013. Keeping a finger on the pulse: Cardiovascular disease rate as a measure of sustainable development. *JRSM Cardiovascular Disease*, 2, p.2048004013491731.

Haver, V.G. et al., 2010. Rupture of vulnerable atherosclerotic plaques: microRNAs conducting the orchestra? *Trends Cardiovasc Med*, 20(2), pp.65–71.

Heinzen, E.L. et al., 2008. Tissue-specific genetic control of splicing: implications for the study of complex traits. *PLoS biology*, 6(12), p.e1.

Heiss, G. et al., 1991. Carotid atherosclerosis measured by B-mode ultrasound in populations: associations with cardiovascular risk factors in the ARIC study. *Am J Epidemiol*, 134(3), pp.250–256.

Helgadottir, A. et al., 2004. The gene encoding 5-lipoxygenase activating protein confers risk of myocardial infarction and stroke. *Nat Genet*, 36(3), pp.233–239.

Heo, S.H. et al., 2011. Plaque rupture is a determinant of vascular events in carotid artery atherosclerotic disease: involvement of matrix metalloproteinases 2 and 9. *J Clin Neurol*, 7(2), pp.69–76.

Heritier, S. & Ronchetti, E., 2004. Robust binary regression with continuous outcomes. *Canadian Journal of Statistics*, 32(3), pp.239–249.

Hernandez, D.G. et al., 2012. Integration of GWAS SNPs and tissue specific expression profiling reveal discrete eQTLs for human traits in blood and brain. *Neurobiol Dis*, 47(1), pp.20–28.

Higgins, J.P.T. & Green, S., 2011. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0* T. C. Collaboration, ed.,

Hildebrand, A. et al., 1994. Interaction of the small interstitial proteoglycans biglycan, decorin and fibromodulin with transforming growth factor beta. *Biochem J*, 302 (Pt 2, pp.527–534.

Hill, M.D., Buchan, A.M. & Canadian Alteplase for Stroke Effectiveness Study, I., 2005. Thrombolysis for acute ischemic stroke: results of the Canadian Alteplase for Stroke Effectiveness Study. *CMAJ*, 172(10), pp.1307–1312.

Hirschhorn, J.N., 2009. Genomewide association studies--illuminating biologic pathways. *The New England journal of medicine*, 360(17), pp.1699–701.

Hollander, M. et al., 2002. Carotid plaques increase the risk of stroke and subtypes of cerebral infarction in asymptomatic elderly: the Rotterdam study. *Circulation*, 105(24), pp.2872–7.

Holliday, E.G. et al., 2012a. Common variants at 6p21.1 are associated with large artery atherosclerotic stroke. *Nat Genet*, 44(10), pp.1147–1151.

Holliday, E.G. et al., 2012b. Common variants at 6p21.1 are associated with large artery atherosclerotic stroke. *Nat Genet*, 44(10), pp.1147–1151.

Holmans, P. et al., 2009. Gene ontology analysis of GWA study data sets provides insights into the biology of bipolar disorder. *Am J Hum Genet*, 85(1), pp.13–24.

Hortin, G.L. & Sviridov, D., 2010. The dynamic range problem in the analysis of the plasma proteome. *Journal of proteomics*, 73(3), pp.629–636.

Hosack, D.A. et al., 2003. Identifying biological themes within lists of genes with EASE. *Genome Biol*, 4(10), p.R70.

Hozo, S.P., Djulbegovic, B. & Hozo, I., 2005. Estimating the mean and variance from the median, range, and the size of a sample. *BMC Medical Research Methodology*, 5(1), p.13.

Huan, T. et al., 2013. A systems biology framework identifies molecular underpinnings of coronary heart disease. *Arterioscler Thromb Vasc Biol*, 33(6), pp.1427–1434.

Huang da, W., Sherman, B.T. & Lempicki, R.A., 2009a. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*, 37(1), pp.1–13.

Huang da, W., Sherman, B.T. & Lempicki, R.A., 2009b. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*, 4(1), pp.44–57.

Huang, P. et al., 2009. Serum free hemoglobin as a novel potential biomarker for acute ischemic stroke. *Journal of neurology*, 256(4), pp.625–31.

Hubner, N. et al., 2005. Integrated transcriptional profiling and linkage analysis for identification of genes underlying disease. *Nat Genet*, 37(3), pp.243–253.

Humphries, S.E. & Morgan, L., 2004. Genetic risk factors for stroke and carotid atherosclerosis: insights into pathophysiology from candidate gene approaches. *Lancet Neurol*, 3(4), pp.227–235.

Hutchens, T.W. & Yip, T.T., 1993. New desorption strategies for the mass spectrometric analysis of macromolecules. *Rapid Commun Mass Spectrom*, 7(7), pp.576–580.

Ijas, P. et al., 2007. Microarray analysis reveals overexpression of CD163 and HO-1 in symptomatic carotid plaques. *Arterioscler Thromb Vasc Biol*, 27(1), pp.154–160.

Ikram, M.A. et al., 2009. Genomewide association studies of stroke. *The New England Journal of Medicine*, 360(17), pp.1718–1728.

Ingall, T.J. et al., 2004. Findings from the reanalysis of the NINDS tissue plasminogen activator for acute ischemic stroke treatment trial. *Stroke: A Journal of Cerebral Circulation*, 35(10), pp.2418–2424.

Intercollegiate Stroke Working Party, 2012. *National Clinical Guideline for Stroke* 4th Editio., Royal College of Physicians.

International HapMap Consortium, 2005. A haplotype map of the human genome. *Nature*, 437(7063), pp.1299–320.

International Stroke Genetics, C. et al., 2012. Genome-wide association study identifies a variant in HDAC9 associated with large vessel ischemic stroke. *Nat Genet*, 44(3), pp.328–333.

International Stroke Genetics, C. & Wellcome Trust Case-Control, C., 2010. Failure to validate association between 12p13 variants and ischemic stroke. *N Engl J Med*, 362(16), pp.1547–1550.

Inzitari, D. et al., 2000. The causes and risk of stroke in patients with asymptomatic internal-carotid-artery stenosis. North American Symptomatic Carotid Endarterectomy Trial Collaborators. *The New England journal of medicine*, 342(23), pp.1693–700.

Ioannidis, J.P. & Panagiotou, O.A., 2011. Comparison of effect sizes associated with biomarkers reported in highly cited individual articles and in subsequent meta-analyses. *JAMA*, 305(21), pp.2200–2210.

Irizarry, R.A., Hobbs, B., et al., 2003. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics (Oxford, England)*, 4(2), pp.249–64.

Irizarry, R.A., Bolstad, B.M., et al., 2003. Summaries of Affymetrix GeneChip probe level data. *Nucleic acids research*, 31(4), p.e15.

Jensen, L.J. et al., 2009. STRING 8--a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res*, 37(Database issue), pp.D412–6.

Jensen, M.B. et al., 2008. The promise and potential pitfalls of serum biomarkers for ischemic stroke and transient ischemic attack. *The neurologist*, 14(4), pp.243–6.

Jia, P. et al., 2011. Pathway-based analysis of GWAS datasets: effective but caution required. *The international journal of neuropsychopharmacology / official scientific journal of the Collegium Internationale Neuropsychopharmacologicum (CINP)*, 14(4), pp.567–72.

Jin, R., Yang, G. & Li, G., 2010. Molecular insights and therapeutic targets for blood-brain barrier disruption in ischemic stroke: critical role of matrix metalloproteinases and tissue-type plasminogen activator. *Neurobiology of Disease*, 38(3), pp.376–385.

Jin, Y. & Manabe, T., 2005. Direct targeting of human plasma for matrix-assisted laser desorption/ionization and analysis of plasma proteins by time of flight-mass spectrometry. *ELECTROPHORESIS*, 26(14), pp.2823–2834.

Johnson, B.F. et al., 1995. Clinical outcome in patients with mild and moderate carotid artery stenosis. *J Vasc Surg*, 21(1), pp.120–126.

Johnson, R.C. et al., 2010. Accounting for multiple comparisons in a genome-wide association study (GWAS). *BMC Genomics*, 11, p.724.

Johnston, S.C. et al., 2007. Validation and refinement of scores to predict very early stroke risk after transient ischaemic attack. *Lancet*, 369(9558), pp.283–292.

Jong, M.C., Hofker, M.H. & Havekes, L.M., 1999. Role of ApoCs in lipoprotein metabolism: functional differences between ApoC1, ApoC2, and ApoC3. *Arterioscler Thromb Vasc Biol*, 19(3), pp.472–484.

Kanehisa, M. & Goto, S., 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*, 28(1), pp.27–30.

Karvunidis, T. et al., 2009. Recent progress of proteomics in critical illness. *Shock Augusta Ga*, 31(6), pp.545–552.

Kastrup, A. et al., 2008. Early disruption of the blood-brain barrier after thrombolytic therapy predicts hemorrhage in patients with acute stroke. *Stroke: A Journal of Cerebral Circulation*, 39(8), pp.2385–2387.

Kato, N. et al., 2011. Meta-analysis of genome-wide association studies identifies common variants associated with blood pressure variation in east Asians. *Nat Genet*, 43(6), pp.531–538.

Katzan, I.L. et al., 2000. Use of tissue-type plasminogen activator for acute ischemic stroke: the Cleveland area experience. *Jama The Journal Of The American Medical Association*, 283(9), pp.1151–1158.

Kavalci, C. et al., 2011. Value of biomarker-based diagnostic test in differential diagnosis of hemorrhagic-ischemic stroke. *Bratislavske Lekarske Listy*, 112(7), pp.398–401.

Kawahara, I. et al., 2007. The expression of vascular dendritic cells in human atherosclerotic carotid plaques. *Hum Pathol*, 38(9), pp.1378–1385.

Kendrick, N., A gene's mRNA level measured by microarrays or RT-PCR does not necessarily predict its protein level: R2 (or Rs) for plots of mRNA versus protein is < 0.4.

Kennedy, R.E. et al., 2012. Association between family risk of stroke and myocardial infarction with prevalent risk factors and coexisting diseases. *Stroke*, 43(4), pp.974–979.

Khatri, P., Sirota, M. & Butte, A.J., 2012. Ten years of pathway analysis: current approaches and outstanding challenges. *PLoS Comput Biol*, 8(2), p.e1002375.

Kidwell, C.S. et al., 2002. Late secondary ischemic injury in patients receiving intraarterial thrombolysis. *Annals of Neurology*, 52(6), pp.698–703.

Kiers, L. et al., 1992. Stroke topography and outcome in relation to hyperglycaemia and diabetes. *Journal of Neurology, Neurosurgery & Psychiatry*, 55(4), pp.263–270.

Kim, K.I. & van de Wiel, M.A., 2008. Effects of dependence in high-dimensional multiple testing problems. *BMC bioinformatics*, 9, p.114.

Kim, Y.J. et al., 2011. Large-scale genome-wide association studies in East Asians identify new genetic loci influencing metabolic traits. *Nat Genet*, 43(10), pp.990–995.

Kingsmore, S.F. et al., 2007. Genome-wide association studies: progress in identifying genetic biomarkers in common, complex diseases. *Biomarker insights*, 2, pp.283–92.

Kisialiou, A. et al., 2012. Blood biomarkers role in acute ischemic stroke patients: higher is worse or better? *Immun Ageing*, 9(1), p.22.

Kitajiri, S. et al., 2010. Actin-bundling protein TRIOBP forms resilient rootlets of hair cell stereocilia essential for hearing. *Cell*, 141(5), pp.786–798.

Klein, R.J. et al., 2005. Complement factor H polymorphism in age-related macular degeneration. *Science (New York, N.Y.)*, 308(5720), pp.385–9.

Kleinig, T.J. & Vink, R., 2009. Suppression of inflammation in ischemic and hemorrhagic stroke: therapeutic options. *Current Opinion in Neurology*, 22(3), pp.294–301.

Knauer, C. et al., 2012. A biochemical marker panel in MRI-proven hyperacute ischemic stroke-a prospective study. *BMC Neurol*, 12, p.14.

Knox, S.S., 2010. From “omics” to complex disease: a systems biology approach to gene-environment interactions in cancer. *Cancer Cell Int*, 10, p.11.

Ko N. Ander B. P., Stamova B., Jickling G., Zhan X., Tian Y., Kim H.,Young W. L., Sharp F.R., X.H., 2011. Exon Level Expression Profiles That Associate with Angiographic Vasospasm in Patients with Aneurysmal Subarachnoid Hemorrhage. *International Stroke Conference.* .

Kodali, P. et al., 2012. Multiple liquid chromatography separations and nanoESI-ion trap detection of plasma proteins in search of stroke biomarkers: A pilot study. *Journal of separation science*, 35(17), pp.2153–61.

Kolmakova, A. et al., 2004. Apolipoprotein C-I Induces Apoptosis in Human Aortic Smooth Muscle Cells via Recruiting Neutral Sphingomyelinase. *Arterioscler Thromb Vasc Biol*, 24(2), pp.264–269.

Kooner, J.S. et al., 2011. Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci. *Nat Genet*, 43(10), pp.984–989.

Van Kooten, F. et al., 1993. Hyperglycemia in the acute phase of stroke is not caused by stress. *Stroke: A Journal of Cerebral Circulation*, 24(8), pp.1129–1132.

Korol, R.M. et al., 2011a. Detection of altered extracellular matrix in surface layers of unstable carotid plaque: an optical spectroscopy, birefringence and microarray genetic analysis. *Photochem Photobiol*, 87(5), pp.1164–1172.

Korol, R.M. et al., 2011b. Detection of altered extracellular matrix in surface layers of unstable carotid plaque: an optical spectroscopy, birefringence and microarray genetic analysis. *Photochem Photobiol*, 87(5), pp.1164–1172.

Kraft, P. & Hunter, D.J., 2009. Genetic risk prediction--are we there yet? *The New England journal of medicine*, 360(17), pp.1701–3.

Lage, K. et al., 2007. A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nat Biotechnol*, 25(3), pp.309–316.

Lander, E.S. & Schork, N.J., 1994. Genetic dissection of complex traits. *Science (New York, N.Y.)*, 265(5181), pp.2037–48.

Lanfranconi, S. & Markus, H.S., 2012. Stroke subtyping for genetic association studies? A comparison of the CCS and TOAST classifications. *International journal of stroke : official journal of the International Stroke Society*.

Lanktree, M.B. et al., 2009. Multi-ethnic genetic association study of carotid intima-media thickness using a targeted cardiovascular SNP microarray. *Stroke; a journal of cerebral circulation*, 40(10), pp.3173–9.

Lansberg, M.G. et al., 2007. Risk factors of symptomatic intracerebral hemorrhage after tPA therapy for acute stroke. *Stroke*, 38(8), pp.2275–2278.

Lansberg, M.G. et al., 2007. Risk factors of symptomatic intracerebral hemorrhage after tPA therapy for acute stroke. *Stroke; a journal of cerebral circulation*, 38(8), pp.2275–8.

De las Fuentes, L. et al., 2012. Pathway-based genome-wide association analysis of coronary heart disease identifies biologically important gene sets. *Eur J Hum Genet*, 20(11), pp.1168–1173.

Laskowitz, D.T. et al., 2009. Clinical usefulness of a biomarker-based diagnostic test for acute stroke: the Biomarker Rapid Assessment in Ischemic Injury (BRAIN) study. *Stroke; a journal of cerebral circulation*, 40(1), pp.77–85.

Laskowitz, D.T. et al., 2005. Panel of biomarkers predicts stroke. *Annals Of The New York Academy Of Sciences*, 1053, p.30.

Lee, J.S. & Gotlieb, A.I., 2003. Understanding the role of the cytoskeleton in the complex regulation of the endothelial repair. *Histol Histopathol*, 18(3), pp.879–887.

Lee, S., Shafe, A.C.E. & Cowie, M.R., 2011. UK stroke incidence, mortality and cardiovascular risk management 1999-2008: time-trend analysis from the General Practice Research Database. *BMJ open*, 1(2), p.e000269.

Lescuyer, P. et al., 2004. Identification of post-mortem cerebrospinal fluid proteins as potential biomarkers of ischemia and neurodegeneration. *Proteomics*, 4(8), pp.2234–2241.

Li, J. et al., 2002. Proteomics and bioinformatics approaches for identification of serum biomarkers to detect breast cancer. *Clin Chem*, 48(8), pp.1296–1304.

Li, M., Li, C. & Guan, W., 2008. Evaluation of coverage variation of SNP chips for genome-wide association studies. *Eur J Hum Genet*, 16(5), pp.635–643.

Liu, J. et al., 2012. Matrix metalloproteinase-2-mediated occludin degradation and caveolin-1-mediated claudin-5 redistribution contribute to blood-brain barrier damage in early ischemic stroke stage. *J Neurosci*, 32(9), pp.3044–3057.

Lopez, M.F. et al., 2012. Discrimination of ischemic and hemorrhagic strokes using a multiplexed, mass spectrometry-based assay for serum apolipoproteins coupled to multi-marker ROC algorithm. *Proteomics Clin Appl*, 6(3-4), pp.190–200.

Lopez, M.F. et al., 2012. Discrimination of ischemic and hemorrhagic strokes using a multiplexed, mass spectrometry-based assay for serum apolipoproteins coupled to multi-marker ROC algorithm. *Proteomics. Clinical applications*, 6(3-4), pp.190–200.

Lorenz, M.W. et al., 2012. Carotid intima-media thickness progression to predict cardiovascular events in the general population (the PROG-IMT collaborative project): a meta-analysis of individual participant data. *Lancet*, 379(9831), pp.2053–2062.

Lorgis, L. et al., 2009. Prognostic value of N-terminal pro-brain natriuretic peptide in elderly people with acute myocardial infarction: prospective observational study. *BMJ*, 338.

Lou, M. et al., 2008. The HAT Score: a simple grading scale for predicting hemorrhage after thrombolysis. *Neurology*, 71(18), pp.1417–1423.

Love, S. & Barber, R., 2001. Expression of P-selectin and intercellular adhesion molecule-1 in human brain after focal infarction or cardiac arrest. *Neuropathol Appl Neurobiol*, 27(6), pp.465–473.

Luftner, D., Luke, C. & Possinger, K., 2003. Serum HER-2/neu in the management of breast cancer patients. *Clin Biochem*, 36(4), pp.233–240.

Lusis, A.J., 2000. Atherosclerosis. *Nature*, 407(6801), pp.233–241.

Lynch, J.R. et al., 2004. Novel diagnostic test for acute stroke. *Stroke*, 35(1), pp.57–63.

Lynch, J.R. et al., 2004. Novel diagnostic test for acute stroke. *Stroke; a journal of cerebral circulation*, 35(1), pp.57–63.

Macintyre, G. et al., 2010. is-rSNP: a novel technique for in silico regulatory SNP detection. *Bioinformatics*, 26(18), pp.i524–30.

Madden, J.A., 2012. Role of the vascular endothelium and plaque in acute ischemic stroke. *Neurology*, 79(13 Suppl 1), pp.S58–62.

Mancia, G., 2010. Effects of intensive blood pressure control in the management of patients with type 2 diabetes mellitus in the Action to Control Cardiovascular Risk in Diabetes (ACCORD) trial. *Circulation*, 122(8), pp.847–849.

Marchini, J. & Howie, B., 2010. Genotype imputation for genome-wide association studies. *Nature reviews. Genetics*, 11(7), pp.499–511.

Markus, H.S. et al., 2013. Evidence HDAC9 genetic variant associated with ischemic stroke increases risk via promoting carotid atherosclerosis. *Stroke; a journal of cerebral circulation*, 44(5), pp.1220–5.

Markus, H.S., 2011. Stroke genetics. *Hum Mol Genet*, 20(R2), pp.R124–31.

Markus, H.S., 2011. Stroke genetics. *Human molecular genetics*, 20(R2), pp.R124–31.

Martinet, W. et al., 2002. Gene expression profiling of apoptosis-related genes in human atherosclerosis: upregulation of death-associated protein kinase. *Arteriosclerosis thrombosis and vascular biology*, 22(12), pp.2023–2029.

Matarin, M. et al., 2007. A genome-wide genotyping study in patients with ischaemic stroke: initial analysis and data release. *Lancet Neurol*, 6(5), pp.414–420.

Maurer, M.H. et al., 2003. The proteome of human brain microdialysate. *Proteome Science*, 1(1), p.7.

Mayeux, R., 2004. Biomarkers: potential uses and limitations. *NeuroRx*, 1(2), pp.182–188.

Mazya, M. et al., 2012. Predicting the risk of symptomatic intracerebral hemorrhage in ischemic stroke treated with intravenous alteplase: safe Implementation of Treatments in Stroke (SITS) symptomatic intracerebral hemorrhage risk score. *Stroke; a journal of cerebral circulation*, 43(6), pp.1524–31.

Mazya, M. V et al., 2013. External validation of the SEDAN score for prediction of intracerebral hemorrhage in stroke thrombolysis. *Stroke*, 44(6), pp.1595–1600.

McGeachie, M. et al., 2009. Integrative predictive model of coronary artery calcification in atherosclerosis. *Circulation*, 120(24), pp.2448–2454.

McShane, L.M. et al., 2005. Reporting recommendations for tumor marker prognostic studies (REMARK). *J Natl Cancer Inst*, 97(16), pp.1180–1184.

Melton, P.E. et al., 2013. Genetic architecture of carotid artery intima-media thickness in Mexican Americans. *Circ Cardiovasc Genet*, 6(2), pp.211–221.

Meschia, J.F. et al., 2011. Genomic risk profiling of ischemic stroke: results of an international genome-wide association meta-analysis. *PLoS One*, 6(9), p.e23161.

Meschia, J.F., 2010. Stroke genome-wide association studies: the large numbers imperative. *Stroke*, 41(4), pp.579–580.

Meschia, J.F. et al., 2003. The Ischemic Stroke Genetics Study (ISGS) Protocol. *BMC Neurol*, 3, p.4.

Meschia, J.F. et al., 2006. The Siblings With Ischemic Stroke Study (SWISS): a progress report. *Clin Med Res*, 4(1), pp.12–21.

Meschia, J.F. & Tournier-Lasserve, E., 2013. Advances in stroke: genetics 2012. *Stroke*, 44(2), pp.309–310.

Meurer, W.J. et al., 2010. Lack of association between hyperglycaemia at arrival and clinical outcomes in acute stroke patients treated with tissue plasminogen activator. *Int J Stroke*, 5(3), pp.163–166.

Michel, P. & Bogousslavsky, J., 2006. Shared mechanisms of ischaemic and haemorrhagic stroke: still a lot to learn. *Journal of Neurology, Neurosurgery & Psychiatry*, 77(1), p.1.

Miller, M. et al., 2011. Triglycerides and cardiovascular disease: a scientific statement from the American Heart Association. *Circulation*, 123(20), pp.2292–2333.

Minguez, P. & Dopazo, J., 2011. Assessing the Biological Significance of Gene Expression Signatures and Co-Expression Modules by Studying Their Network Properties. *PLoS One*, 6(3), p.e17474.

Mohammad, Y., Qattan, M. & Prabhakaran, S., 2010. Epidemiology and Pathophysiology of Intracranial Large Artery Stenosis. *The Open Atherosclerosis & Thrombosis Journal*, 3, pp.3–7.

Moldes, O. et al., 2008. High serum levels of endothelin-1 predict severe cerebral edema in patients with acute ischemic stroke treated with t-PA. *Stroke: A Journal of Cerebral Circulation*, 39(7), pp.2006–2010.

Montaner Delgado, P., Purroy, F., Quintana, M., Penalba, A., Rosell, A., Chacon, P., Molina, C. A. & Alvarez-Sabin, J., J., 2005. Biochemical diagnosis of acute stroke using a panel of plasma biomarkers. . *Cerebrovasc Dis*, 19(Suppl 2), p.47.

Montaner, J., Mendioroz, M., Delgado, P., García-Berrocoso, T., Giralt, D., Merino, C., Ribó, M., Rosell, A., Penalba, A., Fernández-Cadenas, I., Romero, F., Molina, C. & Alvarez-Sabín, J., 2012. Differentiating ischemic from hemorrhagic stroke using plasma biomarkers: The S100B/RAGE pathway. *Journal of proteomics*, pp.1–8.

Montaner, J., Mendioroz, M., Delgado, P., García-Berrocoso, T., Giralt, D., Merino, C., Ribó, M., Rosell, A., Penalba, A., Fernández-Cadenas, I., Romero, F., Molina, C., Alvarez-Sabín, J., et al., 2012. Differentiating ischemic from hemorrhagic stroke using plasma biomarkers: the S100B/RAGE pathway. *Journal of proteomics*, 75(15), pp.4758–65.

Montaner, J. et al., 2003. *Matrix metalloproteinase-9 pretreatment level predicts intracranial hemorrhagic complications after thrombolysis in human stroke.*,

Morris, A.P. et al., 2012. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet*, 44(9), pp.981–990.

Morrissey, C. et al., 2011. Integrated genomic approaches to identification of candidate genes underlying metabolic and cardiovascular phenotypes in the spontaneously hypertensive rat. *Physiol Genomics*, 43(21), pp.1207–1218.

Morrow, D.A. & de Lemos, J.A., 2007. Benchmarks for the assessment of novel cardiovascular biomarkers. *Circulation*, 115(8), pp.949–52.

Moskau, S. et al., 2005. Heritability of carotid artery atherosclerotic lesions: an ultrasound study in 154 families. *Stroke*, 36(1), pp.5–8.

Mukherjee, D. & Patil, C.G., 2011. Epidemiology and the global burden of stroke. *World neurosurgery*, 76(6 Suppl), pp.S85–90.

Muller, J.E. & Tofler, G.H., 1992. Triggering and hourly variation of onset of arterial thrombosis. *Ann Epidemiol*, 2(4), pp.393–405.

Muthaian, R., Minhas, G. & Anand, A., 2012. Pathophysiology of stroke and stroke-induced retinal ischemia: emerging role of stem cells. *Journal of cellular physiology*, 227(3), pp.1269–79.

Naganuma, M. et al., 2011. Reduced estimated glomerular filtration rate is associated with stroke outcome after intravenous rt-PA: the Stroke Acute Management with Urgent Risk-Factor Assessment and Improvement (SAMURAI) rt-PA registry. *Cerebrovasc Dis*, 31(2), pp.123–129.

Naghavi, M. et al., 2003a. From vulnerable plaque to vulnerable patient: a call for new definitions and risk assessment strategies: Part I. *Circulation*, 108(14), pp.1664–1672.

Naghavi, M. et al., 2003b. From vulnerable plaque to vulnerable patient: a call for new definitions and risk assessment strategies: Part II. *Circulation*, 108(15), pp.1772–1778.

Di Napoli, M., 2001. C reactive protein and acute phase of ischaemic stroke. *BMJ*, 322(7302), pp.1605–1606.

Narayanan, M. et al., 2010. Simultaneous Clustering of Multiple Gene Expression and Physical Interaction Datasets. *PLoS Comput Biol*, 6(4), p.e1000742.

Naruko, T. et al., 2002. Neutrophil infiltration of culprit lesions in acute coronary syndromes. *Circulation*, 106(23), pp.2894–2900.

National Audit Office, 2005. *Reducing Brain Damage: Faster Access to Better Stroke Care.*, Department of Health.

National Stroke Strategy Report, 2007. *National Stroke Strategy Report*,

Nicolae, D.L. et al., 2010. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet*, 6(4), p.e1000888.

Nicolaides, A.N. et al., 2005. Severity of asymptomatic carotid stenosis and risk of ipsilateral hemispheric ischaemic events: results from the ACSRS study. *Eur J Vasc Endovasc Surg*, 30(3), pp.275–284.

Nikol, S. et al., 1997. Differential expression of nonmuscle myosin II isoforms in human atherosclerotic plaque. *Atherosclerosis*, 130(1-2), pp.71–85.

Nilsen, T.W. & Graveley, B.R., 2010. Expansion of the eukaryotic proteome by alternative splicing. *Nature*, 463(7280), pp.457–63.

Ning, M. et al., 2010. Proteomic Protease Substrate Profiling of tPA Treatment in Acute Ischemic Stroke Patients: A Step Toward Individualizing Thrombolytic Therapy at the Bedside. *Translational Stroke Research*, 1(4), pp.268–275.

Nishiwaki, A. et al., 2003. Upregulation of P-selectin and intercellular adhesion molecule-1 after retinal ischemia-reperfusion injury. *Invest Ophthalmol Vis Sci*, 44(11), pp.4931–4935.

North American Symptomatic Carotid Endarterectomy Trial Collaborators [NASCET], 1991. Beneficial effect of carotid endarterectomy in symptomatic patients with high-grade carotid stenosis. *N Engl J Med*, 325(7), pp.445–453.

Noto, A.T. et al., 2008. The ApoC-I content of VLDL particles is associated with plaque size in persons with carotid atherosclerosis. *Lipids*, 43(7), pp.673–679.

O'Donnell, M.J. et al., 2010. Risk factors for ischaemic and intracerebral haemorrhagic stroke in 22 countries (the INTERSTROKE study): a case-control study. *The Lancet*, 376(9735), pp.112–123.

O'Sullivan, J.F., Martin, K. & Caplice, N.M., 2011. Microribonucleic acids for prevention of plaque rupture and in-stent restenosis: "a finger in the dam." *J Am Coll Cardiol*, 57(4), pp.383–389.

Oates, C.P. et al., 2009. Joint recommendations for reporting carotid ultrasound investigations in the United Kingdom. *Eur J Vasc Endovasc Surg*, 37(3), pp.251–261.

Omenn, G.S. et al., 2005. Overview of the HUPO Plasma Proteome Project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. *Proteomics*, 5(13), pp.3226–45.

Ovbiagele, B. et al., 2013. Forecasting the Future of Stroke in the United States: A Policy Statement From the American Heart Association and American Stroke Association. *Stroke*.

Ovbiagele, B. et al., 2013. Forecasting the Future of Stroke in the United States: A Policy Statement From the American Heart Association and American Stroke Association. *Stroke*.

Paciaroni, M. et al., 2009. Acute hyperglycemia and early hemorrhagic transformation in ischemic stroke. *Cerebrovasc Dis*, 28(2), pp.119–123.

Pagani, F. & Baralle, F.E., 2004. Genomic variants in exons and introns: identifying the splicing spoilers. *Nat Rev Genet*, 5(5), pp.389–396.

Papaspyridonos, M. et al., 2006. *Novel candidate genes in unstable areas of human atherosclerotic plaques.*,

Park, S.Y. et al., 2013. Plasma heart-type fatty acid binding protein level in acute ischemic stroke: comparative analysis with plasma S100B level for diagnosis of stroke and prediction of long-term clinical outcome. *Clin Neurol Neurosurg*, 115(4), pp.405–410.

Pasquale, L.R. et al., 2013. Exploring genome-wide - dietary heme iron intake interactions and the risk of type 2 diabetes. *Front Genet*, 4, p.7.

Passador-Gurgel, G. et al., 2007. Quantitative trait transcripts for nicotine resistance in Drosophila melanogaster. *Nature genetics*, 39(2), pp.264–8.

Pearl, D.C., 2002. Proteomic patterns in serum and identification of ovarian cancer. *Lancet*, 360(9327), pp.161–169.

Pearson, H., 2006. Genetics: what is a gene? *Nature*, 441(7092), pp.398–401.

Pearson, T.A. et al., 2003. Markers of inflammation and cardiovascular disease: application to clinical and public health practice: A statement for healthcare professionals from the Centers for Disease Control and Prevention and the American Heart Association. *Circulation*, 107(3), pp.499–511.

Peeters, W. et al., 2009. Carotid atherosclerotic plaques stabilize after stroke: insights into the natural process of atherosclerotic plaque stabilization. *Arterioscler Thromb Vasc Biol*, 29(1), pp.128–133.

Pelisek, J., Eckstein, H.H. & Zernecke, A., 2012. Pathophysiological mechanisms of carotid plaque vulnerability: impact on ischemic stroke. *Arch Immunol Ther Exp (Warsz)*, 60(6), pp.431–442.

Pepe, M.S. et al., 2004. Limitations of the Odds Ratio in Gauging the Performance of a Diagnostic , Prognostic , or Screening Marker. *American Journal of Epidemiology*, 159(9), pp.882–890.

Pepe, M.S., 2004. Limitations of the Odds Ratio in Gauging the Performance of a Diagnostic, Prognostic, or Screening Marker. *American Journal of Epidemiology*, 159(9), pp.882–890.

Pepe, M.S. et al., 2001. Phases of biomarker development for early detection of cancer. *Journal of the National Cancer Institute*, 93(14), pp.1054–61.

Perisic, L. et al., 2013. Profiling of Atherosclerotic Lesions by Gene and Tissue Microarrays Reveals PCSK6 as a Novel Protease in Unstable Carotid Atherosclerosis. *Arterioscler Thromb Vasc Biol*.

Petretto, E. et al., 2010. New insights into the genetic control of gene expression using a Bayesian multi-tissue approach. *PLoS Comput Biol*, 6(4), p.e1000737.

Petricoin, E.F. et al., 2002. Proteomic patterns in serum and identification of ovarian cancer. *The Lancet*, 360(9327), pp.170–171.

Petricoin, E.F. et al., 2002. Use of proteomic patterns in serum to identify ovarian cancer. *Lancet*, 359(9306), pp.572–577.

Petzold, A. et al., 2008. Glial and axonal body fluid biomarkers are related to infarct volume, severity, and outcome. *Journal of stroke and cerebrovascular diseases the official journal of National Stroke Association*, 17(4), pp.196–203.

Pezzini, A. et al., 2009. Common genetic markers and prediction of recurrent events after ischemic stroke in young adults. *Neurology*, 73(9), pp.717–723.

Piran, S. et al., 2012. Where genome meets phenome: rationale for integrating genetic and protein biomarkers in the diagnosis and management of dilated cardiomyopathy and heart failure. *J Am Coll Cardiol*, 60(4), pp.283–289.

Purcell, S. et al., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*, 81(3), pp.559–575.

Putaala, J. et al., 2011. Post-thrombolytic hyperglycemia and 3-month outcome in acute ischemic stroke. *Cerebrovasc Dis*, 31(1), pp.83–92.

Qui Qi, X., Liu, J., Zhang, W. & Wang, Q., F., 2008. Relationship between subtypes of MRI-confirmed acute ischemic stroke based on OCSP and TOAST classifications. *Neural Regeneration Research*, 3(7), pp.798–801.

Rai, A.J. et al., 2005. HUPO Plasma Proteome Project specimen collection and handling: towards the standardization of parameters for plasma proteome samples. *Proteomics*, 5(13), pp.3262–3277.

Rainer, T.H. et al., 2007. Comparison of plasma beta-globin DNA and S-100 protein concentrations in acute stroke. *Clinica chimica acta international journal of clinical chemistry*, 376(1-2), pp.190–196.

Redgrave, J.N. et al., 2008. Critical cap thickness and rupture in symptomatic carotid plaques: the oxford plaque study. *Stroke*, 39(6), pp.1722–1729.

Redgrave, J.N. et al., 2006. Histological assessment of 526 symptomatic carotid plaques in relation to the nature and timing of ischemic symptoms: the Oxford plaque study. *Circulation*, 113(19), pp.2320–2328.

Reynolds, M.A. et al., 2003. Early biomarkers of stroke. *Clinical Chemistry*, 49(10), pp.1733–1739.

Ribo, M. et al., 2005. Acute hyperglycemia state is associated with lower tPA-induced recanalization rates in stroke patients. *Stroke: A Journal of Cerebral Circulation*, 36(8), pp.1705–1709.

Ribo, M. et al., 2004. Admission fibrinolytic profile is associated with symptomatic hemorrhagic transformation in stroke patients treated with tissue plasminogen activator. *Stroke: A Journal of Cerebral Circulation*, 35(9), pp.2123–2127.

Riggs, J., 1996. Tissue-type plasminogen activator should not be used in acute ischemic stroke. *Arch Neurol.*, 53(12), pp.1306–8.

Righetti, P.G. et al., 2005. Prefractionation techniques in proteome analysis: the mining tools of the third millennium. *Electrophoresis*, 26(2), pp.297–319.

Riley, R.D., Lambert, P.C. & Abo-Zaid, G., 2010. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ*, 340, p.c221.

Rockhill, B., 2002. Proteomic patterns in serum and identification of ovarian cancer. *The Lancet*, 360(9327), p.169.

Ronaldson, P.T. & Davis, T.P., 2012. Blood-brain barrier integrity and glial support: mechanisms that can be targeted for novel therapeutic approaches in stroke. *Curr Pharm Des*, 18(25), pp.3624–3644.

Ross, R., 1999. Atherosclerosis--an inflammatory disease. *N Engl J Med*, 340(2), pp.115–126.

Rossin, E.J. et al., 2011. Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. *PLoS Genet*, 7(1), p.e1001273.

Rothwell, P.M. et al., 2004. Change in stroke incidence, mortality, case-fatality, severity, and risk factors in Oxfordshire, UK from 1981 to 2004 (Oxford Vascular Study). *Lancet*, 363(9425), pp.1925–1933.

Rothwell, P.M. et al., 2000. Evidence of a chronic systemic cause of instability of atherosclerotic plaques. *Lancet*, 355(9197), pp.19–24.

Rothwell, P.M., 2005. External validity of randomised controlled trials: "to whom do the results of this trial apply?" *Lancet*, 365(9453), pp.82–93.

Rothwell, P.M., 2010. Limitations of the usual blood-pressure hypothesis and importance of variability, instability, and episodic hypertension. *Lancet*, 375(9718), pp.938–948.

Rothwell, P.M., Villagra, R. & Donders, R., 1996. The role of carotid atherosclerosis in the aetiology of ischaemic stroke. *Cerebrovasc Dis*, 6(Suppl 2), p.1.

Rothwell, P.M. & Warlow, C.P., 2005. Timing of TIAs preceding stroke: time window for prevention is very short. *Neurology*, 64(5), pp.817–820.

Royston, P., Altman, D.G. & Sauerbrei, W., 2006. Dichotomizing continuous predictors in multiple regression: a bad idea. *Statistics in medicine*, 25(1), pp.127–41.

Rozek, A., Buchko, G.W. & Cushley, R.J., 1995. Conformation of two peptides corresponding to human apolipoprotein C-I residues 7-24 and 35-53 in the presence of sodium dodecyl sulfate by CD and NMR spectroscopy. *Biochemistry*, 34(22), pp.7401–7408.

Saksi, J. et al., 2011. Gene expression differences between stroke-associated and asymptomatic carotid plaques. *J Mol Med (Berl)*, 89(10), pp.1015–1026.

Sanson-Fisher, R.W. et al., 2007. Limitations of the randomized controlled trial in evaluating population-based health interventions. *American Journal of Preventive Medicine*, 33(2), pp.155–61.

Sasaki, S. et al., 2013. Variants in the 3' UTR of General Transcription Factor IIF, polypeptide 2 affect female calving efficiency in Japanese Black cattle. *BMC Genet*, 14, p.41.

Saver, J.L., 2007. Hemorrhage after thrombolytic therapy for stroke: the clinically relevant number needed to harm. *Stroke*, 38(8), pp.2279–2283.

Saver, J.L., 2006. Time is brain--quantified. *Stroke*, 37(1), pp.263–266.

Schaub, S. et al., 2004. Urine protein profiling with surface-enhanced laser-desorption/ionization time-of-flight mass spectrometry. *Kidney Int*, 65(1), pp.323–332.

Schriger, D.L. et al., 1998. Cranial computed tomography interpretation in acute stroke: physician accuracy in determining eligibility for thrombolytic therapy. *JAMA*, 279(16), pp.1293–1297.

Schwanhausser, B. et al., 2011. Global quantification of mammalian gene expression control. *Nature*, 473(7347), pp.337–342.

Seipel, K. et al., 2001. Tara, a novel F-actin binding protein, associates with the Trio guanine nucleotide exchange factor and regulates actin cytoskeletal organization. *J Cell Sci*, 114(Pt 2), pp.389–399.

Seshadri, S. et al., 2010. Parental occurrence of stroke and risk of stroke in their children: the Framingham study. *Circulation*, 121(11), pp.1304–1312.

Shami, A. et al., 2013. Fibromodulin deficiency reduces low-density lipoprotein accumulation in atherosclerotic plaques in apolipoprotein E-null mice. *Arterioscler Thromb Vasc Biol*, 33(2), pp.354–361.

Shinton, R. & Beevers, G., 1989. Meta-analysis of relation between cigarette smoking and stroke. *BMJ*, 298(6676), pp.789–794.

Sladek, R. et al., 2007. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature*, 445(7130), pp.881–5.

Sluimer, J.C. et al., 2007. Dead or alive: gene expression profiles of advanced atherosclerotic plaques from autopsy and surgery. *Physiol Genomics*, 30(3), pp.335–341.

Soilleux, E.J. et al., 2002. Human atherosclerotic plaques express DC-SIGN, a novel protein found on dendritic cells and macrophages. *J Pathol*, 198(4), pp.511–516.

Spencer, C.C. et al., 2009. Designing genome-wide association studies: sample size, power, imputation, and the choice of genotyping chip. *PLoS Genet*, 5(5), p.e1000477.

Staals, J. et al., 2008. A SELDI-TOF-MS study in Lacunar Stroke with Subsequent Haptoglobin Phenotyping. *Curr Neurovasc Res*, 5(2), pp.93–98.

Sterne, J.A. et al., 2011. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ*, 343, p.d4002.

Storey, J.D., 2002. A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(3), pp.479–498.

Strazzullo, P. et al., 2010. Excess body weight and incidence of stroke: meta-analysis of prospective studies with 2 million participants. *Stroke*, 41(5), pp.e418–26.

Strbian, D., Sairanen, T., Rantanen, K., et al., 2011. Characteristics and outcome of ischemic stroke patients who are free of symptoms at 24 hours following thrombolysis. *Cerebrovasc Dis*, 31(1), pp.37–42.

Strbian, D., Sairanen, T., Meretoja, A., et al., 2011. Patient outcomes from symptomatic intracerebral hemorrhage after stroke thrombolysis. *Neurology*, 77(4), pp.341–348.

Strbian, D. et al., 2012. Symptomatic intracranial hemorrhage after stroke thrombolysis: the SEDAN score. *Annals of neurology*, 71(5), pp.634–41.

Strbian, D., Sairanen, T. & Meretoja, A., 2012. Patient outcomes from symptomatic intracerebral hemorrhage after stroke thrombolysis. *Online*.

Streiner, D.L., 2002. Breaking up is hard to do: the heartbreak of dichotomizing continuous data. *Canadian journal of psychiatry Revue canadienne de psychiatrie*, 47(3), pp.262–266.

Strong, K., Mathers, C. & Bonita, R., 2007. Preventing stroke: saving lives around the world. *Lancet Neurol*, 6(2), pp.182–187.

Subramanian, A. et al., 2005. Gene set enrichment analysis : A knowledge-based approach for interpreting genome-wide.

Subramanian, A. et al., 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*, 102(43), pp.15545–15550.

Suissa, S. & Blais, L., 1995. Binary regression with continuous outcomes. *Statistics in Medicine*, 14(3), pp.247–255.

Takano, S. et al., 2008. Apolipoprotein C-1 maintains cell survival by preventing from apoptosis in pancreatic cancer cells. *Oncogene*, 27(20), pp.2810–2822.

Tam, S.W., Pirro, J. & Hinerfeld, D., 2004. Depletion and fractionation technologies in plasma proteomic analysis. *Expert Review Of Proteomics*, 1(4), pp.411–420.

Tamam, Y., Iltumur, K. & Apak, I., 2005. Assessment of acute phase proteins in acute ischemic stroke. *Tohoku J Exp Med*, 206(2), pp.91–98.

Tanne, D. et al., 2002a. Markers of increased risk of intracerebral hemorrhage after intravenous recombinant tissue plasminogen activator therapy for acute ischemic stroke in clinical practice: the Multicenter rt-PA Stroke Survey. *Circulation*, 105(14), pp.1679–1685.

Tanne, D. et al., 2002b. Markers of increased risk of intracerebral hemorrhage after intravenous recombinant tissue plasminogen activator therapy for acute ischemic stroke in clinical practice: the Multicenter rt-PA Stroke Survey. *Circulation*, 105(14), pp.1679–1685.

Di Taranto, M.D. et al., 2012a. Altered expression of inflammation-related genes in human carotid atherosclerotic plaques. *Atherosclerosis*, 220(1), pp.93–101.

Di Taranto, M.D. et al., 2012b. Altered expression of inflammation-related genes in human carotid atherosclerotic plaques. *Atherosclerosis*, 220(1), pp.93–101.

Thompson, A.R. et al., 2010. Assessment of the association between genetic polymorphisms in transforming growth factor beta, and its binding protein (LTBP), and the presence, and expansion, of Abdominal Aortic Aneurysm. *Atherosclerosis*, 209(2), pp.367–373.

Thompson, J.R., Attia, J. & Minelli, C., 2011. The meta-analysis of genome-wide association studies. *Brief Bioinform*, 12(3), pp.259–269.

Tobin, W.O. et al., 2009. Identification of stroke mimics in the emergency department setting. *J Brain Dis*, 1, pp.19–22.

Townsend N. Bhatnagar P., Smolina K., Nichols M., Leal J., Luengo-Fernandez R., Rayner M., W.K., 2012. *Coronary Heart Disease Statistics* B. H. Foundation, ed., London.

Traylor, M. et al., 2012. Genetic risk factors for ischaemic stroke and its subtypes (the METASTROKE collaboration): a meta-analysis of genome-wide association studies. *Lancet Neurol*, 11(11), pp.951–962.

Tress, M.L. et al., 2007. The implications of alternative splicing in the ENCODE protein complement. *Proceedings of the National Academy of Sciences of the United States of America*, 104(13), pp.5495–500.

Tuttolomondo, A. et al., 2009. Immuno-inflammatory and thrombotic/fibrinolytic variables associated with acute ischemic stroke diagnosis. *Atherosclerosis*, 203(2), pp.503–508.

Tuttolomondo, A. et al., 2008. Predictors of outcome in acute ischemic cerebrovascular syndromes: The GIFA study. *International Journal of Cardiology*, 125(3), pp.391–396.

Tworoger, S.S. & Hankinson, S.E., 2006. Use of biomarkers in epidemiologic studies: minimizing the influence of measurement error in the study design and analysis. *Cancer Causes Control*, 17(7), pp.889–899.

Urban, Z. et al., 2009. Mutations in LTBP4 cause a syndrome of impaired pulmonary, gastrointestinal, genitourinary, musculoskeletal, and dermal development. *Am J Hum Genet*, 85(5), pp.593–605.

Uyttenboogaart, M. et al., 2008a. Lipid profile, statin use, and outcome after intravenous thrombolysis for acute ischaemic stroke. *J Neurol*, 255(6), pp.875–880.

Uyttenboogaart, M. et al., 2008b. Lipid profile, statin use, and outcome after intravenous thrombolysis for acute ischaemic stroke. *Journal of neurology*, 255(6), pp.875–80.

Vemuganti, R., 2010. The MicroRNAs and Stroke: No Need to be Coded to be Counted. *Transl Stroke Res*, 1(3), pp.158–160.

Vemuganti, R., 2010. The MicroRNAs and Stroke: No Need to be Coded to be Counted. *Translational stroke research*, 1(3), pp.158–160.

Vemuganti, R. & Dempsey, R.J., 2006. Increased expression of genes that control ionic homeostasis, second messenger signaling and metabolism in the carotid plaques from patients with symptomatic stroke. *Journal of Neurochemistry*, 97 Suppl 1, pp.92–96.

Veyrieras, J.B. et al., 2008. High-resolution mapping of expression-QTLs yields insight into human gene regulation. *PLoS Genet*, 4(10), p.e1000214.

Virmani, R. et al., 2000. Lessons from sudden coronary death: a comprehensive morphological classification scheme for atherosclerotic lesions. *Arterioscler Thromb Vasc Biol*, 20(5), pp.1262–1275.

Virmani, R., Burke, A.P. & Farb, A., 2001. Sudden cardiac death. *Cardiovasc Pathol*, 10(5), pp.211–218.

Voight, B.F. et al., 2010. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat Genet*, 42(7), pp.579–589.

Wang, E.T. et al., 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature*, 456(7221), pp.470–6.

Wang, J. et al., 2012. The expression of IGFs and IGF binding proteins in human carotid atherosclerosis, and the possible role of IGF binding protein-1 in the regulation of smooth muscle cell proliferation. *Atherosclerosis*, 220(1), pp.102–109.

Wang, J. & Doré, S., 2007. Inflammation after intracerebral hemorrhage. *Journal of cerebral blood flow and metabolism official journal of the International Society of Cerebral Blood Flow and Metabolism*, 27(5), pp.894–908.

Wang, K., Li, M. & Bucan, M., 2007. Pathway-based approaches for analysis of genomewide association studies. *Am J Hum Genet*, 81(6), pp.1278–1283.

Wang, Q., Tang, X.N. & Yenari, M.A., 2007. The inflammatory response in stroke. *J Neuroimmunol*, 184(1-2), pp.53–68.

Wannamethee, S.G., Shaper, A.G. & Perry, I.J., 1997. Serum creatinine concentration and risk of cardiovascular disease: a possible marker for increased risk of stroke. *Stroke*, 28(3), pp.557–563.

Wardlaw, J.M., 2009. Thrombolysis for acute ischaemic stroke. *Cochrane Database Syst Rev.*, 7(4), p.CD000213.

Ware, J.S., Petretto, E. & Cook, S.A., 2013. Integrative genomics in cardiovascular medicine. *Cardiovasc Res*, 97(4), pp.623–630.

Warlow, C. et al., 2003. Stroke. *Lancet*, 362(9391), pp.1211–1224.

De Weerd, M. et al., 2010. Prevalence of asymptomatic carotid artery stenosis in the general population: an individual participant data meta-analysis. *Stroke*, 41(6), pp.1294–1297.

Wellcome Trust Case Control, C., 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, 447(7145), pp.661–678.

Welsh, P. et al., 2009. Associations of inflammatory and haemostatic biomarkers with poor outcome in acute ischaemic stroke. *Cerebrovascular diseases (Basel, Switzerland)*, 27(3), pp.247–53.

Westerterp, M. et al., 2007. Apolipoprotein CI aggravates atherosclerosis development in ApoE-knockout mice despite mediating cholesterol efflux from macrophages. *Atherosclerosis*, 195(1), pp.e9–16.

Westerterp, M. et al., 2007. Apolipoprotein CI aggravates atherosclerosis development in ApoE-knockout mice despite mediating cholesterol efflux from macrophages. *Atherosclerosis*, 195(1), pp.e9–16.

Whisnant, J.P., 1997. Modeling of risk factors for ischemic stroke. The Willis Lecture. *Stroke*, 28(9), pp.1840–1844.

White, M.Y., Gundry, R.L. & Cordwell, S.J., 2006. When does a fingerprint constitute a diagnostic? *Lance*, 368(9540), pp.971–973.

Whiteley, W., Tseng, M.-C. & Sandercock, P., 2008. Blood biomarkers in the diagnosis of ischemic stroke: a systematic review. *Stroke; a journal of cerebral circulation*, 39(10), pp.2902–9.

Whiteley, W.N. et al., 2012. Risk factors for intracranial hemorrhage in acute ischemic stroke patients treated with recombinant tissue plasminogen activator: a systematic review and meta-analysis of 55 studies. *Stroke*, 43(11), pp.2904–2909.

Wiesner, A., 2004. Detection of tumor markers with ProteinChip technology. *Curr Pharm Biotechnol*, 5(1), pp.45–67.

Wolfenson, H. et al., 2011. Actomyosin-generated tension controls the molecular kinetics of focal adhesions. *J Cell Sci*, 124(Pt 9), pp.1425–1432.

Woodside, K., 2003. Differential gene expression in primary and recurrent carotid stenosis. *Biochemical and Biophysical Research Communications*, 302(3), pp.509–514.

Woolas, R.P. et al., 1995. Combinations of multiple serum markers are superior to individual assays for discriminating malignant from benign pelvic masses. *Gynecol Oncol*, 59(1), pp.111–116.

World Health Organisation MONICA Project, 1990. *MONICA Manual:Monitoring of Trends and Determinants in Cardiovascular Disease.* 1990:part ed. W. H. O. Cardiovascular Diseases Unit, ed., Geneva, Switzerland.

Wunderlich, M.T. et al., 2005. Release of brain-type and heart-type fatty acid-binding proteins in serum after acute ischaemic stroke. *Journal of Neurology*, 252(6), pp.718–724.

Xi, G., Keep, R.F. & Hoff, J.T., 2006. Mechanisms of brain injury after intracerebral haemorrhage. *The Lancet*, 5(1), pp.53–63.

Xue, A. et al., 2010. Discovery of serum biomarkers for pancreatic adenocarcinoma using proteomic analysis. *Br J Cancer*, 103(3), pp.391–400.

Yadav, S. et al., 2011. Bio-Repository of DNA in stroke (BRAINS): a study protocol. *BMC Med Genet*, 12, p.34.

Yang, I. V et al., 2002. Within the fold: assessing differential expression measures and reproducibility in microarray assays. *Genome biology*, 3(11), p.research0062.

Yang, Q. et al., 2007. Genome-wide association and linkage analyses of hemostatic factors and hematological phenotypes in the Framingham Heart Study. *BMC Med Genet*, 8 Suppl 1, p.S12.

You, J.-S. et al., 2005. The impact of blood contamination on the proteome of cerebrospinal fluid. *Proteomics*, 5(1), pp.290–296.

Yu, Z. et al., 2007. Aberrant allele frequencies of the SNPs located in microRNA target sites are potentially associated with human cancers. *Nucleic Acids Res*, 35(13), pp.4535–4541.

Yuan, C. et al., 2001. Carotid atherosclerotic plaque: noninvasive MR characterization and identification of vulnerable lesions. *Radiology*, 221(2), pp.285–299.

Yusuf, S. et al., 2004. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. *Lancet*, 364(9438), pp.937–952.

Zawadzki, C. et al., 2009. Tissue factor pathway inhibitor-2 gene methylation is associated with low expression in carotid atherosclerotic plaques. *Atherosclerosis*, 204(2), pp.e4–14.

Zhang, K. et al., 2011. ICSNPathway: identify candidate causal SNPs and pathways from genome-wide association study by one analytical framework. *Nucleic Acids Res*, 39(Web Server issue), pp.W437–43.

Zhang, K. et al., 2010. i-GSEA4GWAS: a web server for identification of pathways/gene sets associated with traits by applying an improved gene set enrichment analysis to genome-wide association study. *Nucleic Acids Res*, 38(Web Server issue), pp.W90–5.

Zhang, R.L. et al., 1999. Thrombolysis with tissue plasminogen activator alters adhesion molecule expression in the ischemic rat brain. *Stroke*, 30(3), pp.624–629.

Zhang, X. et al., 2008a. Potential biomarkers of acute cerebral infarction detected by SELDI-TOF-MS. *Am J Clin Pathol*, 130(2), pp.299–304.

Zhang, X. et al., 2008b. Potential biomarkers of acute cerebral infarction detected by SELDI-TOF-MS. *Am J Clin Pathol*, 130(2), pp.299–304.

Zhang, Y. & Zhang, C., 2010. Role of dendritic cells in cardiovascular diseases. *World J Cardiol*, 2(11), pp.357–364.

Ziegler, A., König, I.R. & Thompson, J.R., 2008. Biostatistical aspects of genome-wide association studies. *Biometrical journal. Biometrische Zeitschrift*, 50(1), pp.8–28.

Zimmermann, K. & Leser, U., 2010. Analysis of Affymetrix Exon Arrays.

Zimmermann-Ivol, C.G. et al., 2004. Fatty acid binding protein as a serum marker for the early diagnosis of stroke: a pilot study. *Molecular cellular proteomics MCP*, 3(1), pp.66–72.

Zinkstok, S.M. et al., 2013. Safety of thrombolysis in stroke mimics: results from a multicenter cohort study. *Stroke*, 44(4), pp.1080–1084.

Del Zoppo, G.J. et al., 2009. Expansion of the time window for treatment of acute ischemic stroke with intravenous tissue plasminogen activator: a science advisory from the American Heart Association/American Stroke Association. *Stroke*, 40(8), pp.2945–2948.

References for Chapter 2 (Tables 2.1, 2.2 & 2.3)

1. Castellanos M, Castillo J, García MM, Leira R, Serena J, Chamorro A, et al. Inflammation-mediated damage in progressing lacunar infarctions: a potential therapeutic target. Stroke: A Journal of Cerebral Circulation. 2002;33(4):982–7.

2. Shyu KG, Chang H, Lin CC, Kuan P. Circulating intercellular adhesion molecule-1 and E-selectin in patients with acute coronary syndrome. Chest. 1996;109(6):1627–30.

3. Supanc V, Biloglav Z, Kes VB, Demarin V. Role of cell adhesion molecules in acute ischemic stroke. Annals of Saudi Medicine. Medknow Publications; 2011;31(4):365–70.

4. Lu L-F, Yang S-S, Wang C-P, Hung W-C, Yu T-H, Chiu C-A, et al. Elevated visfatin/pre-B-cell colony-enhancing factor plasma concentration in ischemic stroke. Journal of stroke and cerebrovascular diseases the official journal of National Stroke Association. 2009;18(5):354–9.

5. Rajeshwar K, Kaul S, Al-Hazzani A, Babu MS, Balakrishna N, Sharma V, et al. C-reactive protein and nitric oxide levels in ischemic stroke and its subtypes: correlation with clinical outcome. Inflammation. 2012 Jun;35(3):978–84.

6. Beg M, Nizami A, Singhal KC, Mohammed J, Gupta A, Azfar SF. Role of serum fibrinogen in patients of ischemic cerebrovascular disease. Nepal Medical College journal NMCJ. 2007;9(2):88–92.

7. Cha J-K, Jeong M-H, Kim E-K, Lim Y-J, Ha B-R, Kim S-H, et al. Surface expression of P-selectin on platelets is related with clinical worsening in acute ischemic stroke. Journal of Korean Medical Science. Korean Academy of Medical Sciences; 2002;17(6):811–6.

8. Dai D-F, Thajeb P, Tu C-F, Chiang F-T, Chen C-H, Yang R-B, et al. Plasma concentration of SCUBE1, a novel platelet protein, is elevated in patients with acute coronary syndrome and ischemic stroke. Journal of the American College of Cardiology. 2008 Jun;51(22):2173–80.

9. Weikert C, Westphal S, Berger K, Dierkes J, Möhlig M, Spranger J, et al. Plasma resistin levels and risk of myocardial infarction and ischemic stroke. The Journal of Clinical Endocrinology & Metabolism. 2008;93(7):2647–53.

10. Youssef MYZ, Mojiminiyi OA, Abdella NA. Plasma concentrations of C-reactive protein and total homocysteine in relation to the severity and risk factors for cerebrovascular disease. Translational research the journal of laboratory and clinical medicine. 2007;150(3):158–63.

11. Tsai T-H, Chen Y-L, Lin H-S, Liu C-F, Chang H-W, Lu C-H, et al. Link between lipoprotein-associated phospholipase A2 gene expression of peripheral-blood mononuclear cells and prognostic outcome after acute ischemic stroke. Journal of atherosclerosis and thrombosis. 2012 Jan;19(6):523–31.

12. Lukasik M, Dworacki G, Michalak S, Kufel-Grabowska J, Watala C, Kozubski W. Chronic hyper-reactivity of platelets resulting in enhanced monocyte recruitment in patients after ischaemic stroke. Platelets. 2012 Jan;23(2):132–42.

13. Chen S, Martens-Lobenhoffer J, Weissenborn K, Kielstein JT, Lichtinghagen R, Deb M, et al. Association of dimethylarginines and mediators of inflammation after acute ischemic stroke. Journal of neuroinflammation. 2012 Jan;9:251.

14. Ben Assayag E, Shenhar-Tsarfaty S, Ofek K, Soreq L, Bova I, Shopin L, et al. Serum cholinesterase activities distinguish between stroke patients and controls and predict 12-month mortality. Molecular medicine (Cambridge, Mass.). 2010;16(7-8):278–86.

15. Cheung EYL, Uitte De Willige S, Vos HL, Leebeek FWG, Dippel DWJ, Bertina RM, et al. Fibrinogen gamma' in ischemic stroke: a case-control study. Stroke: A Journal of Cerebral Circulation. 2008;39(3):1033–5.

16. De Lau LML, Cheung EYL, Kluft C, Leebeek FWG, Meijer P, Laterveer R, et al. Strongly increased levels of fibrinogen elastase degradation products in patients with ischemic stroke. British journal of haematology. 2008 Dec;143(5):734–7.

17. Elkind MS, Coates K, Tai W, Paik MC, Boden-Albala B, Sacco RL. Levels of acute phase proteins remain stable after ischemic stroke. BMC Neurology. BioMed Central; 2006;6:37.

18. Waje-Andreassen U, Kråkenes J, Ulvestad E, Thomassen L, Myhr K-M, Aarseth J, et al. IL-6: an early marker for outcome in acute ischemic stroke. Acta Neurologica Scandinavica. 2005;111(6):360–5.

19. Yip H-K, Lai S-L, Lan M-Y, Chang W-N, Liu JS, Kao Y-F, et al. Time course of platelet activation and von Willebrand factor in patients with non-valvular atrial fibrillation after ischemic stroke. Circulation journal : official journal of the Japanese Circulation Society. 2007 Mar;71(3):321–6.

20. Perini F, Morra M, Alecci M, Galloni E, Marchi M, Toso V. Temporal profile of serum anti-inflammatory and pro-inflammatory interleukins in acute ischemic stroke patients. Neurological sciences official journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology. 2001;22(4):289–96.

21. Cojocaru IM, Cojocaru IM CMBCAA, Cojocaru M, Burcin C, Atanasiu A. Detection of neopterin as parameter of potential monocyte activation in patients with acute ischemic stroke. Romanian journal of internal medicine = Revue roumaine de médecine interne. 2007 Jan;45(4):365–9.

22. Grau AJ, Reis A, Buggle F, Al-Khalaf A, Werle E, Valois N, et al. Monocyte function and plasma levels of interleukin-8 in acute ischemic stroke. Journal of the Neurological Sciences. 2001;192(1-2):41–7.

23. Seneş M, Kazan N, Coşkun O, Zengi O, Inan L YD. Oxidative and nitrosative stress in acute ischaemic stroke. Annals of Clinical Biochemistry. 2007;44(Pt 1):43–7.

24. Herisson F, Delaroche O, Auffray-Calvier E, Duport BD, Guillon B. Ischemia-modified albumin and heart fatty acid-binding protein: could early ischemic cardiac biomarkers be used in acute stroke management? Journal of stroke and cerebrovascular diseases : the official journal of National Stroke Association. Elsevier Ltd; 2010;19(4):279–82.

25. Nadar SK, Lip GYH, Blann AD. Platelet morphology, soluble P selectin and platelet P-selectin in acute ischaemic stroke. The West Birmingham Stroke Project. Thrombosis and haemostasis. 2004;92(6):1342–8.

26. Zimmermann-Ivol CG, Burkhard PR, Le Floch-Rohr J, Allard L, Hochstrasser DF, Sanchez J-C. Fatty acid binding protein as a serum marker for the early diagnosis of stroke: a pilot study. Molecular cellular proteomics MCP. 2004;3(1):66–72.

27. Oberheiden T, Nguyen XD, Fatar M, Elmas E, Blahak C, Morper N, et al. Platelet and monocyte activation in acute ischemic stroke--is there a correlation with stroke etiology? Clinical and applied thrombosis/hemostasis : official journal of the International Academy of Clinical and Applied Thrombosis/Hemostasis. 2012;18(1):87–91.

28. Cano CP, Bermúdez VP, Atencio HE, Medina MT, Anilsa A, Souki A, et al. Increased serum malondialdehyde and decreased nitric oxide within 24 hours of thrombotic stroke onset. American journal of therapeutics. 2003;10(6):473–6.

29. Dambinova S a, Khounteev G a, Izykenova G a, Zavolokov IG, Ilyukhina AY, Skoromets A a. Blood test detecting autoantibodies to N-methyl-D-aspartate neuroreceptors for evaluation of patients with transient ischemic attack and stroke. Clinical chemistry. 2003 Oct;49(10):1752–62.

30. Khan U, Crossley C, Kalra L, Rudd A, Wolfe CDA, Collinson P, et al. Homocysteine and its relationship to stroke subtypes in a UK black population: the south London ethnicity and stroke study. Stroke: A Journal of Cerebral Circulation. 2008;39(11):2943–9.

31. Basic Kes V, Simundic A-M, Nikolac N, Topic E, Demarin V. Pro-inflammatory and anti-inflammatory cytokines in acute ischemic stroke and their relation to early neurological deficit and stroke outcome. Clinical biochemistry. The Canadian Society of Clinical Chemists; 2008 Nov;41(16-17):1330–4.

32. Alvarez-Perez FJ, Castelo-Branco M, Alvarez-Sabin J. Usefulness of measurement of fibrinogen, D-dimer, D-dimer/fibrinogen ratio, C reactive protein and erythrocyte sedimentation rate to assess the pathophysiology and mechanism of ischaemic stroke. Journal of neurology, neurosurgery, and psychiatry. 2011 Sep;82(9):986–92.

33. Rainer TH, Wong KS, Lam W, Lam NYL, Graham CA, Lo YMD. Comparison of plasma beta-globin DNA and S-100 protein concentrations in acute stroke. Clinica chimica acta international journal of clinical chemistry. 2007;376(1-2):190–6.

34. Glickman SW, Phillips S, Anstrom KJ, Laskowitz DT, Cairns CB. Discriminative capacity of biomarkers for acute stroke in the emergency department. The Journal of emergency medicine. Elsevier Inc.; 2011 Sep;41(3):333–9.

35. Dvorak F, Haberer I, Sitzer M, Foerch C. Characterisation of the diagnostic window of serum glial fibrillary acidic protein for the differentiation of intracerebral haemorrhage and ischaemic stroke. Cerebrovascular diseases (Basel, Switzerland). 2009 Jan;27(1):37–41.

36. Foerch C, Niessner M, Back T, Bauerle M, De Marchis GM, Ferbert A, et al. Diagnostic accuracy of plasma glial fibrillary acidic protein for differentiating intracerebral hemorrhage and cerebral ischemia in patients with symptoms of acute stroke. Clinical chemistry. 2012 Jan;58(1):237–45.

37. Undén J, Strandberg K, Malm J, Campbell E, Rosengren L, Stenflo J, et al. Explorative investigation of biomarkers of brain damage and coagulation system activation in clinical stroke differentiation. Journal of neurology. 2009 Jan;256(1):72–7.

38. Modrego PJ, Boned B, Berlanga JJ, Serrano M. Plasmatic B-Type Natriuretic Peptide and C-Reactive Protein in Hyperacute Stroke as Markers of Ct-Evidence of Brain Edema. International Journal of Medical Sciences. Ivyspring International Publisher; 2008;5(1):18–23.

39. Fonseca AC, Matias JS, Pinho e Melo T, Falcão F, Canhão P, Ferro JM. N-terminal probrain natriuretic peptide as a biomarker of cardioembolic stroke. International journal of stroke : official journal of the International Stroke Society. 2011 Oct;6(5):398–403.

40. Montaner J, Mendioroz M, Delgado P, García-Berrocoso T, Giralt D, Merino C, et al. Differentiating ischemic from hemorrhagic stroke using plasma biomarkers: The S100B/RAGE pathway. Journal of proteomics. Elsevier B.V.; 2012 Feb 11;1–8.

41. Kavalci C, Genchallac H, Durukan P, Cevik Y. Value of biomarker-based diagnostic test in differential diagnosis of hemorrhagic-ischemic stroke. Bratislavske Lekarske Listy. 2011;112(7):398–401.

42. Castellanos M, Leira R, Serena J, Pumar JM, Lizasoain I, Castillo J, et al. Plasma metalloproteinase-9 concentration predicts hemorrhagic transformation in acute ischemic stroke. Stroke: A Journal of Cerebral Circulation. 2003;34(1):40–6.

43. Hernandez-Guillamon M, Garcia-Bonilla L, Solé M, Sosti V, Parés M, Campos M, et al. Plasma VAP-1/SSAO activity predicts intracranial hemorrhages and adverse neurological outcome after tissue plasminogen activator treatment in stroke. Stroke: A Journal of Cerebral Circulation. 2010;41(7):1528–35.

44. Leira R, Sobrino T, Blanco M, Campos F, Rodríguez-Yáñez M, Castellanos M, et al. A higher body temperature is associated with haemorrhagic transformation in patients with acute stroke untreated with recombinant tissue-type plasminogen activator (rtPA). Clinical science (London, England : 1979). 2012 Feb;122(3):113–9.

45. Rodríguez-Yáñez M, Castellanos M, Blanco M, García MM, Nombela F, Serena J, et al. New-onset hypertension and inflammatory response/poor outcome in acute ischemic stroke. Neurology. 2006;67(11):1973–8.

46. Sienkiewicz-Jarosz H, Gałecka-Wolska M, Bidziński A, Turzyńska D, Sobolewska A, Lipska B, Płaźnik A RD. Predictive value of selected biochemical markers of brain damage for functional outcome in ischaemic stroke patients. Neurol Neurochir Pol. 2009;43(2):126–33.

47. Sobrino T, Hurtado O, Moro MA, Rodríguez-Yáñez M, Castellanos M, Brea D, et al. The increase of circulating endothelial progenitor cells after acute ischemic stroke is associated with good outcome. Stroke: A Journal of Cerebral Circulation. 2007;38(10):2759–64.

48. Whiteley W, Jackson C, Lewis S, Lowe G, Rumley A, Sandercock P, et al. Inflammatory Markers and Poor Outcome after Stroke: A Prospective Cohort Study and Systematic Review of Interleukin-6. Turnbull FM, editor. PLoS Medicine. Public Library of Science; 2009;6(9):e1000145.

49. Campos F, Rodríguez-Yáñez M, Castellanos M, Arias S, Pérez-Mato M, Sobrino T, et al. Blood levels of glutamate oxaloacetate transaminase are more strongly associated with good outcome in acute ischaemic stroke than glutamate pyruvate transaminase levels. Clinical science (London, England : 1979). 2011 Jul;121(1):11–7.

50. Welsh P, Barber M, Langhorne P, Rumley A, Lowe GDO, Stott DJ. Associations of inflammatory and haemostatic biomarkers with poor outcome in acute ischaemic stroke. Cerebrovascular diseases (Basel, Switzerland). 2009 Jan;27(3):247–53.

51. Castellanos M, Sobrino T, Pedraza S, Moldes O, Pumar JM, Silva Y, et al. High plasma glutamate concentrations are associated with infarct growth in acute ischemic stroke. Neurology. 2008 Dec;71(23):1862–8.

52. Oto J, Suzue A, Inui D, Fukuta Y, Hosotsubo K, Torii M, et al. Plasma proinflammatory and anti-inflammatory cytokine and catecholamine concentrations as predictors of neurological outcome in acute stroke patients. Journal of anesthesia. 2008 Jan;22(3):207–12.

53. Castillo J, Dávalos A, Noya M. Progression of ischaemic stroke and excitotoxic aminoacids. Lancet. 1997;349(9045):79–83.

54. Serena J, Leira R, Castillo J, Pumar JM, Castellanos M, Dávalos A. Neurological deterioration in acute lacunar infarctions: the role of excitatory and inhibitory neurotransmitters. Stroke: A Journal of Cerebral Circulation. 2001;32(5):1154–61.

55. Vila N, Castillo J, Dávalos A, Chamorro A. Proinflammatory cytokines and early neurological worsening in ischemic stroke. Stroke: A Journal of Cerebral Circulation. 2000;31(10):2325–9.

56. Dávalos A, Castillo J, Marrugat J, Fernandez-Real JM, Armengou A, Cacabelos P, et al. Body iron stores and early neurologic deterioration in acute cerebral infarction. Neurology. 2000;54(8):1568–74.

57. Gentile NT, Seftchick MW, Huynh T, Kruus LK, Gaughan J. Decreased mortality by normalizing blood glucose after acute ischemic stroke. Academic emergency medicine official journal of the Society for Academic Emergency Medicine. 2006;13(2):174–80.

58. Shibazaki K, Kimura K, Okada Y, Iguchi Y, Uemura J, Terasawa Y, et al. Plasma brain natriuretic peptide as an independent predictor of in-hospital mortality after acute ischemic stroke. Internal medicine Tokyo Japan. 2009;48(18):1601–6.

59. Hajdinjak E, Klemen P, Grmec S. Prognostic Value of a Single Prehospital Measurement of N-Terminal Pro-Brain Natriuretic Peptide and Troponin T after Acute Ischaemic Stroke. Journal of International Medical Research. 2012 Apr 1;40(2):768–76.

60. Rallidis LS, Vikelis M, Panagiotakos DB, Liakos GK, Krania E, Kremastinos DT. Usefulness of inflammatory and haemostatic markers to predict short-term risk for death in middle-aged ischaemic stroke patients. Acta neurologica Scandinavica. 2008 Jun;117(6):415–20.

61. Nayak AR, Kashyap RS, Kabra D, Deoras P, Purohit HJ, Taori GM, et al. Evaluation of routinely performed hematological and biochemical parameters for the prognosis of acute ischemic stroke patients. Neurological sciences : official journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology. 2011 Oct;32(5):855–60.

62. Sairanen T, Strbian D, Soinne L, Silvennoinen H, Salonen O, Artto V, et al. Intravenous thrombolysis of basilar artery occlusion: predictors of recanalization and outcome. Stroke; a journal of cerebral circulation. 2011 Aug;42(8):2175–9.

63. Strbian D, Atula S, Meretoja A, Kaste M, Tatlisumak T. Outcome of ischemic stroke patients with serious post-thrombolysis neurological deficits. Acta neurologica Scandinavica. 2013 Apr;127(4):221–6.

64. Millan M, Sobrino T, Castellanos M, Nombela F, Arenillas JF, Riva E, et al. Increased body iron stores are associated with poor outcome after thrombolytic treatment in acute stroke. Stroke: A Journal of Cerebral Circulation. 2007;38(1):90–5.

65. Mendioroz M, Fernández-Cadenas I, Alvarez-Sabín J, Rosell A, Quiroga D, Cuadrado E, et al. Endogenous activated protein C predicts hemorrhagic transformation and mortality after tissue plasminogen activator treatment in stroke patients. Cerebrovascular diseases (Basel, Switzerland). 2009 Jan;28(2):143–50.

66. Cocho D, Borrell M, Martí-Fàbregas J, Montaner J, Castellanos M, Bravo Y, et al. Pretreatment hemostatic markers of symptomatic intracerebral hemorrhage in patients treated with tissue plasminogen activator. Stroke: A Journal of Cerebral Circulation. 2006;37(4):996–9.

67. Ribo M, Montaner J, Molina C a, Arenillas JF, Santamarina E, Quintana M, et al. Admission fibrinolytic profile is associated with symptomatic hemorrhagic transformation in stroke patients treated with tissue plasminogen activator. Stroke; a journal of cerebral circulation. 2004 Sep;35(9):2123–7.

68. Topakian R, Strasak AM, Nussbaumer K, Haring H-P, Aichner FT. Prognostic value of admission C-reactive protein in stroke patients undergoing iv thrombolysis. Journal of Neurology. 2008;255(8):1190–6.

69. Montaner J, Fernandez-Cadenas I, Molina C, Ribó M, Huertas R, Rosell A, et al. Poststroke C-reactive protein is a powerful prognostic tool among candidates for thrombolysis. Stroke; a journal of cerebral circulation. 2006 May;37(5):1205–10.

APPENDIX 1

Search Strategy for Systematic Literature Review (Chapter 2)

Diagnostic and Prognostic Biomarkers for Ischaemic stroke

Modified from Whiteley *et al* 2008.

1. cerebrovascular disorders OR basal ganglia cerebrovascular disease OR brain ischemia [MeSH Term] OR carotid artery diseases OR carotid artery thrombosis OR carotid stenosis OR cerebrovascular accident OR brain infarction [MeSH Term] OR intracranial arterial diseases [MeSH Term] OR "intracranial embolism and thrombosis" [MeSH Term]

2. (brain [Title/Abstract] OR cerebr* [Title/Abstract] OR cerebell* [Title/Abstract] OR vertebrobasil* [Title/Abstract] OR hemispher* [Title/Abstract] OR intracran* [Title/Abstract] OR intracerebral [Title/Abstract] OR infratentorial [Title/Abstract] OR intraventricular [Title/Abstract] OR supratentorial [Title/Abstract] OR parenchymal [Title/Abstract] OR middle cerebr* [Title/Abstract] OR mca* [Title/Abstract] OR anterior circulation [Title/Abstract]) AND (ischemi* [Title/Abstract] OR ischaemi* [Title/Abstract] OR infarct* [Title/Abstract] OR thrombo* [Title/Abstract] OR emboli* [Title/Abstract] OR occlus* [Title/Abstract] OR hypoxi* [Title/Abstract])

3. (ischemi* [Title/Abstract] OR ischaemi* [Title/Abstract]) AND (stroke* [Title/Abstract] OR apoplex* [Title/Abstract] OR cerebral vasc* [Title/Abstract] OR cerebrovasc* [Title/Abstract] OR cva [Title/Abstract] OR attack* [Title/Abstract])

4. 1 OR 2 OR 3

5. "biological markers" [MeSH Term]

6. biomarker* [Title/Abstract]

7. (biochemical [Title/Abstract] OR clinical [Title/Abstract] OR immun* [Title/Abstract] OR laboratory [Title/Abstract] OR biological* [Title/Abstract] OR [Title/Abstract] serum [Title/Abstract] OR [Title/Abstract] surrogate [Title/Abstract] OR viral [Title/Abstract]) AND (marker* [Title/Abstract])

8. (blood [Title/Abstract] OR plasma [Title/Abstract]) AND (marker [Title/Abstract])

9. Fructose-Bisphosphate Aldolase OR activins OR inhibin-beta subunits OR Inhibins OR Adiponectin OR Antiplasmin OR Macroglobulins OR alpha 1-antichymotrypsin OR alpha 1-antitrypsin OR Orosomucoid OR Peptidyl-Dipeptidase A OR Fibroblast Growth Factor 2 OR angiotensins OR angiotensin i OR angiotensin ii OR angiotensin iii OR Antithrombin III OR apolipoproteins OR apolipoproteins a OR apolipoprotein a-i OR apolipoprotein a-ii OR apolipoproteins b OR apolipoprotein

b-48 OR apolipoprotein b-100 OR apolipoproteins c OR apolipoprotein c-i OR apolipoprotein c-ii OR apolipoproteins d OR apolipoproteins e OR apolipoprotein e2 OR apolipoprotein e3 OR apolipoprotein e4 OR beta 2-Glycoprotein I OR Natriuretic Peptide, Brain OR Brain-Derived Neurotrophic Factor OR caspases, effector OR caspase 3 OR caspase 6 OR caspase 7 OR caspase 14 OR Cathepsin B OR antigens, cd40 OR cd40 ligand OR Ceruloplasmin OR Chitinase OR Cholesterol Ester Transfer Proteins OR Chromogranin A OR Clusterin OR Fibronectins OR Chimerin Proteins OR Chimerin 1 OR complement system proteins OR anaphylatoxins OR complement activating enzymes OR complement c1 OR complement c2 OR complement c3 OR complement c4 OR complement c5 OR complement c6 OR complement c7 OR complement c8 OR complement c9 OR complement factor b OR complement inactivator proteins OR complement membrane attack complex OR properdin OR C-Reactive Protein OR Fibrin Fibrinogen Degradation Products OR phosphopyruvate hydratase OR crystallins OR cell adhesion molecules OR antigens, cd22 OR antigens, cd24 OR antigens, cd31 OR antigens, cd146 OR antigens, cd164 OR cadherins OR carcinoembryonic antigen OR cd4 immunoadhesins OR cell adhesion molecules, neuronal OR integrin alpha beta2 OR intercellular adhesion molecule-1 OR receptors, lymphocyte homing OR selectins OR vascular cell adhesion molecule-1 OR endothelins OR endothelin-1 OR endothelin-2 OR endothelin-3 OR Erythropoietin OR E-Selectin OR Factor XI OR Factor IX OR Factor XII OR Factor V OR Factor VII OR Factor VIII OR Factor X OR Factor XIIa OR Interleukins [MeSH Term] OR Fibrinogen [MeSH Term] OR Antigens, CD95 OR Ferritins [MeSH Term] OR fibrinopeptide a OR fibrinopeptide b OR Fibronectins [MeSH Term] OR Follistatin-Related Proteins [MeSH Term] OR Follistatin [MeSH Term] OR Fatty Acids [MeSH Term], Nonesterified OR Glial Fibrillary Acidic Protein [MeSH Term]OR Glutathione Transferase [MeSH Term] OR Granulocyte-Macrophage Colony-Stimulating Factor OR Selectins [MeSH Term] OR Platelet Glycoprotein GPIIb-IIIa Complex OR growth hormone OR human growth hormone OR Haptoglobins [MeSH Term] OR Hemopexin OR Heparin Cofactor II OR Intercellular Adhesion Molecule-1 [MeSH Term] OR Immunoglobulin G [MeSH Term] OR Laminin OR Leptin OR Macrophage Colony-Stimulating Factor OR Malondialdehyde OR matrix metalloproteinases [MeSH Term], secreted OR Monocyte Chemoattractant Proteins [MeSH Term] OR Myelin Basic Proteins OR Peroxidase OR S100 Proteins [MeSH Term] OR Neurotrophin 3 OR 9 Nitric Oxide OR Nucleoside-Diphosphate Kinase OR Aryldialkylphosphatase OR Phosphoglycerate Mutase OR Pregnancy-Associated Plasma Protein-A OR Plasminogen Activator Inhibitor 1 OR Plasminogen OR Plasminogen Activator Inhibitor 2 OR Platelet Activating Factor OR Antigens, CD31 OR Platelet-Derived Growth Factor OR Platelet Factor 4 OR Protein C OR Protein S OR Prothrombin OR Resistin OR Plasminogen Inactivators OR Platelet Activation OR tau Proteins OR Thrombin OR Thrombomodulin OR Thromboplastin OR Tumour Necrosis Factor-Alpha OR Transforming Growth Factor beta OR Vascular Endothelial Growth Factor A OR Vitronectin OR von Willebrand Factor

10. Aldolase A [Title/Abstract] OR aldolase b [Title/Abstract] OR aldolase c [Title/Abstract] OR fructose bisphosphate aldolase [Title/Abstract] OR activin* [Title/Abstract] OR inhibin* [Title/Abstract] OR adiponectin [Title/Abstract] OR adipocyte specific secretory protein [Title/Abstract] OR gelatin binding protein [Title/Abstract] OR adipocyte complement related protein [Title/Abstract] OR antiplasmin [Title/Abstract] OR Alpha-2-AP [Title/Abstract] OR Alpha-2-PI [Title/Abstract] OR plasmin inhibitor [Title/Abstract] OR pigment epithelium derived factor [Title/Abstract] OR plasmin inhibitor [Title/Abstract] OR macroglobulin* [Title/Abstract] OR alpha 2 [Title/Abstract] OR antichymotrypsin [Title/Abstract] OR alpha 1-antichymotrypsin [Title/Abstract] OR alpha 1-antitrypsin [Title/Abstract] OR Seromucoid [Title/Abstract] OR sialomucin [Title/Abstract] OR alpha 1-acid glycoprotein [Title/Abstract] OR acid seromucoid [Title/Abstract] OR acid seromucoid [Title/Abstract] OR acid alpha 1-glycoprotein [Title/Abstract] OR alpha 1 -acid glycoprotein [Title/Abstract] OR alpha 1-acid glycoprotein [Title/Abstract] OR alpha 1-glycoprotein acid [Title/Abstract] OR angiotensin converting enzyme [Title/Abstract] OR cd143 [Title/Abstract] OR cd143 [Title/Abstract] OR kininase ii [Title/Abstract] OR angiotensin i-converting enzyme [Title/Abstract] OR carboxycathepsin [Title/Abstract] OR dipeptidyl peptidase [Title/Abstract] OR kininase a [Title/Abstract] OR ACE [Title/Abstract] OR kininase 2 [Title/Abstract] OR Dipeptidyl carboxypeptidase I [Title/Abstract] OR basic fibroblast growth factor [Title/Abstract] OR fibroblast growth factor [Title/Abstract] OR hbgf-2 [Title/Abstract] OR cartilage-derived growth factor [Title/Abstract] OR heparin-binding growth factor [Title/Abstract] OR fgf-2 [Title/Abstract] OR fgf2

[Title/Abstract] OR fibroblast growth factor-2 [Title/Abstract] OR heparin-binding growth factor [Title/Abstract] OR prostate epithelial cell growth factor [Title/Abstract] OR prostatropin [Title/Abstract] OR Fibroblast Growth Factor 2 [Title/Abstract] OR heparin-binding growth factor 2 [Title/Abstract] OR angiotensin* [Title/Abstract] OR antithrombin ii [Title/Abstract] OR heparin cofactor i [Title/Abstract] OR at iii [Title/Abstract] OR antithrombin iii [Title/Abstract] OR antithrombin iii-alpha [Title/Abstract] OR atenativ [Title/Abstract] OR antithrombin [Title/Abstract] OR factor xa inhibitor [Title/Abstract] OR heparin cofactor i [Title/Abstract] OR thrombate iii [Title/Abstract] OR antithrombin 3 [Title/Abstract] OR antithrombin-3 [Title/Abstract] OR antithrombin iii [Title/Abstract] OR apolipoprotein* [Title/Abstract] OR beta 2 glycoprotein* [Title/Abstract] OR beta 2-Glycoprotein I [Title/Abstract] OR brain natriuretic peptide [Title/Abstract] OR nesiritide [Title/Abstract] OR b-type natriuretic peptide [Title/Abstract] OR bnp [Title/Abstract] OR bnp-32 [Title/Abstract] OR brain natriuretic peptide-32 [Title/Abstract] OR natrecor [Title/Abstract] OR natriuretic factor [Title/Abstract] OR natriuretic peptide type-b [Title/Abstract] OR type-b natriuretic peptide [Title/Abstract] OR ventricular natriuretic peptide [Title/Abstract] OR Brain-Derived Neurotrophic Factor [Title/Abstract] OR casp3 [Title/Abstract] OR apopain [Title/Abstract] OR caspase-3 [Title/Abstract] OR procaspase-3 [Title/Abstract] OR procaspase-3 [Title/Abstract] OR caspase 3 [Title/Abstract] OR cathepsin b-like activity [Title/Abstract] OR cathepsin b-like proteinase [Title/Abstract] OR cathepsin b1 [Title/Abstract] OR cathepsin b [Title/Abstract] OR amyloid precursor protein secretase [Title/Abstract] OR endoglin* [Title/Abstract] OR CD105 [Title/Abstract] OR cd40 [Title/Abstract] OR Bp50 [Title/Abstract] OR caeruloplasmin [Title/Abstract] OR caeruloplasmin [Title/Abstract] OR ferroxidase [Title/Abstract] OR ceruloplasmin ferroxidase [Title/Abstract] OR ceruloplasmin oxidase [Title/Abstract] OR ferroxidase i [Title/Abstract] OR alpha 2 -ceruloplasmin [Title/Abstract] OR endochitinase [Title/Abstract] OR chitinase* [Title/Abstract] OR chitotriosidase [Title/Abstract] OR cholesterol ester transport protein [Title/Abstract] OR cetp [Title/Abstract] OR cholesteryl ester exchange protein [Title/Abstract] OR cholesteryl ester transfer protein [Title/Abstract] OR parathyroid secretory protein [Title/Abstract] OR secretory protein i [Title/Abstract] OR Chromogranin A [Title/Abstract] OR pancreastatin [Title/Abstract] OR parastatin [Title/Abstract] OR Pituitary secretory protein I [Title/Abstract] OR vasostatin [Title/Abstract] OR apoj protein [Title/Abstract] OR apolipoprotein j [Title/Abstract] OR complement lysis inhibitor [Title/Abstract] OR complement-associated protein sp-40,40 [Title/Abstract] OR mac393 antigen [Title/Abstract] OR sgp-2 protein [Title/Abstract] OR sp 40,40 protein [Title/Abstract] OR sulfated glycoprotein 2 [Title/Abstract] OR sulfated glycoprotein-2 [Title/Abstract] OR trpm-2 protein [Title/Abstract] OR xip8 protein [Title/Abstract] OR Complement cytolysis inhibitor [Title/Abstract] OR clusterin [Title/Abstract] OR cold-insoluble globulins [Title/Abstract] OR lets proteins [Title/Abstract] OR fibronectin [Title/Abstract] OR opsonic glycoprotein [Title/Abstract] OR opsonic alpha 2 sb glycoprotein [Title/Abstract] OR alpha 2-surface binding glycoprotein [Title/Abstract] OR c-fibronectin [Title/Abstract] OR c fibronectin [Title/Abstract] OR cellular fibronectin [Title/Abstract] OR -1 chimerin [Title/Abstract] OR -2 chimerin [Title/Abstract] OR -chimerin [Title/Abstract] OR n-chimerin [Title/Abstract] OR chimaerin [Title/Abstract] OR -1 chimaerin [Title/Abstract] OR -2 chimaerin [Title/Abstract] OR -chimaerin [Title/Abstract] OR alpha1-chimaerin [Title/Abstract] OR n-chimaerin [Title/Abstract] OR chimerin [Title/Abstract] OR chimerin* [Title/Abstract] OR complement [Title/Abstract] OR c reactive protein [Title/Abstract] OR c-reactive protein [Title/Abstract] OR CRP [Title/Abstract] OR antithrombin vi [Title/Abstract] OR fibrin degradation products [Title/Abstract] OR fibrin fibrinogen split products [Title/Abstract] OR Fibrin Fibrinogen Degradation Products [Title/Abstract] OR D-dimer [Title/Abstract] OR D dimer [Title/Abstract] OR enolase [Title/Abstract] OR neuron-specific enolase [Title/Abstract] OR 2-phospho-d-glycerate hydrolase [Title/Abstract] OR nervous system-specific enolase [Title/Abstract] OR non-neuronal enolase [Title/Abstract] OR Phosphopyruvate Hydratase [Title/Abstract] OR Neuron specific enolase [Title/Abstract] OR Neurone specific enolase [Title/Abstract] OR Neurone-specific enolase [Title/Abstract] OR endothelial protein c receptor [Title/Abstract] OR endothelial cell protein c receptor [Title/Abstract] OR protein c receptor [Title/Abstract] OR CD201 [Title/Abstract] OR cd106 [Title/Abstract] OR cd106 antigen [Title/Abstract] OR vcam-1 [Title/Abstract] OR cd106 [Title/Abstract] OR incam-110 [Title/Abstract] OR inducible cell adhesion molecule 110 [Title/Abstract] OR vascular cell adhesion molecule [Title/Abstract] OR big endothelin [Title/Abstract] OR big endothelin-1 [Title/Abstract] OR et-1

endothelin-1 [Title/Abstract] OR endothelin type 1 [Title/Abstract] OR endothelin [Title/Abstract] OR preproendothelin [Title/Abstract] OR preproendothelin-1 [Title/Abstract] OR proendothelin [Title/Abstract] OR proendothelin-1 precursor [Title/Abstract] OR Erythropoietin [Title/Abstract] OR cd62e [Title/Abstract] OR cd62 [Title/Abstract] OR e selectin [Title/Abstract] OR elam-1 [Title/Abstract] OR endothelial leukocyte adhesion molecule-1 [Title/Abstract] OR lecam-2 [Title/Abstract] OR cd62e antigen [Title/Abstract] OR endothelial leukocyte adhesion molecule 1 [Title/Abstract] OR e-selectin [Title/Abstract] OR selectin e [Title/Abstract] OR autoprothrombin ii [Title/Abstract] OR christmas factor [Title/Abstract] OR coagulation factor ix [Title/Abstract] OR ptc [Title/Abstract] OR plasma thromboplastin component [Title/Abstract] OR blood coagulation factor ix [Title/Abstract] OR factor ix complex [Title/Abstract] OR factor ix [Title/Abstract] OR coagulation factor xi [Title/Abstract] OR plasma thromboplastin antecedent [Title/Abstract] OR blood coagulation factor xi [Title/Abstract] OR coagulation factor xii [Title/Abstract] OR hageman factor [Title/Abstract] OR blood coagulation factor xii [Title/Abstract] OR coagulation factor v [Title/Abstract] OR proaccelerin [Title/Abstract] OR ac globulin [Title/Abstract] OR blood coagulation factor v [Title/Abstract] OR factor pi [Title/Abstract] OR factor v [Title/Abstract] OR factor ix [Title/Abstract] OR factor xii [Title/Abstract] OR factor xi [Title/Abstract] OR coagulation factor vii [Title/Abstract] OR proconvertin [Title/Abstract] OR stable factor [Title/Abstract] OR blood coagulation factor vii [Title/Abstract] OR factor vii [title/abstract] or antihemophilic factor [title/abstract] or coagulation factor viii [title/abstract] or factor viii clotting antigen [title/abstract] or factor viii coagulant antigen [title/abstract] or factor viii procoagulant activity [title/abstract] or thromboplastinogen [title/abstract] or blood coagulation factor viii [title/abstract] or f viii-c [title/abstract] or factor viii-heavy chain [title/abstract] or factor viiic [title/abstract] or hemofil [title/abstract] or hemofil m [title/abstract] or hemophil [title/abstract] or humate-p [title/abstract] or monoclate [title/abstract] or factor viii [title/abstract] or autoprothrombin iii [title/abstract] or coagulation factor x [title/abstract] or stuart factor [title/abstract] or stuart-prower factor [title/abstract] or blood coagulation factor x [title/abstract] or stuart prower factor [title/abstract] or factor vii activating protease [title/abstract] or coagulation factor xiia [title/abstract] or factor xii [title/abstract] or activated factor xii [title/abstract] or blood coagulation factor xii [title/abstract] or hageman-factor fragments [title/abstract] or prekallikrein activator [title/abstract] or factor xiia [title/abstract] or interleukin [title/abstract] or fibrinogen [title/abstract] or coagulation factor i [title/abstract] or factor i [title/abstract] or blood coagulation factor [title/abstract] or -fibrinogen [title/abstract] or apo-1 antigen [title/abstract] or apoptosis antigen 1 [title/abstract] or cd95 antigens [title/abstract] or fas [title/abstract] or tumor necrosis factor receptor superfamily [title/abstract] or fas antigens [title/abstract] or fas receptors [title/abstract] or cd95 antigen [title/abstract] or tnfrsf6 receptor [title/abstract] or fas antigen [title/abstract] or fas receptor [title/abstract] or basic isoferritin [title/abstract] OR ferritin [Title/Abstract] OR isoferritin [Title/Abstract] OR fibrinopeptide [Title/Abstract] OR cold-insoluble globulins [Title/Abstract] OR lets proteins [Title/Abstract] OR fibronectin [Title/Abstract] OR opsonic glycoprotein [Title/Abstract] OR opsonic alpha 2 sb glycoprotein [Title/Abstract] OR alpha 2-surface binding glycoprotein [Title/Abstract] OR activin-binding protein [Title/Abstract] OR follistatin [Title/Abstract] OR fatty acids [Title/Abstract] OR free fatty acids [Title/Abstract] OR nefa [Title/Abstract] OR glial fibrillary acidic protein [Title/Abstract] OR GFAP [Title/Abstract] OR glial intermediate filament protein [Title/Abstract] OR astroprotein [Title/Abstract] OR gfa-protein [Title/Abstract] OR glial fibrillary acid protein [Title/Abstract] OR glutathione s-alkyltransferase [Title/Abstract] OR glutathione s-aryltransferase [Title/Abstract] OR glutathione s-epoxidetransferase [Title/Abstract] OR ligandins [Title/Abstract] OR glutathione organic nitrate ester reductase [Title/Abstract] OR glutathione s-transferase [Title/Abstract] OR glutathione s-transferase 3 [Title/Abstract] OR glutathione s-transferase a [Title/Abstract] OR glutathione s-transferase b [Title/Abstract] OR glutathione s-transferase c [Title/Abstract] OR glutathione s-transferase iii [Title/Abstract] OR glutathione s-transferase p [Title/Abstract] OR glutathione transferase e [Title/Abstract] OR glutathione transferase mu [Title/Abstract] OR glutathione transferases [Title/Abstract] OR heme transfer protein [Title/Abstract] OR ligandin [Title/Abstract] OR csf-gm [Title/Abstract] OR colony-stimulating factor [Title/Abstract] OR gm-csf [Title/Abstract] OR histamine-producing cell-stimulating factor [Title/Abstract] OR csf-2 [Title/Abstract] OR granulocyte macrophage colony stimulating factor [Title/Abstract] OR cd62p [Title/Abstract] OR cd62p antigens [Title/Abstract] OR gmp-140 [Title/Abstract] OR p selectin [Title/Abstract] OR

platelet -granule membrane protein [Title/Abstract] OR cd62p antigen [Title/Abstract] OR padgem [Title/Abstract] OR cd62l [Title/Abstract] OR cd62l antigens [Title/Abstract] OR lecam-1 [Title/Abstract] OR cd62l antigen [Title/Abstract] OR l selectin [Title/Abstract] OR lam-1 [Title/Abstract] OR leu-8 antigen [Title/Abstract] OR leukocyte adhesion molecule [Title/Abstract] OR mel-14 antigen [Title/Abstract] OR tq1 antigen [Title/Abstract] OR cd62e [Title/Abstract] OR cd62e antigen [Title/Abstract] OR e selectin [Title/Abstract] OR endothelial leukocyte adhesion molecule-1 [Title/Abstract] OR lecam-2 [Title/Abstract] OR cd62e antigen [Title/Abstract] OR endothelial leukocyte adhesion molecule 1 [Title/Abstract] OR gp130 [Title/Abstract] OR sgp130 [Title/Abstract] OR interleukin 11 receptor [Title/Abstract] OR gpiib-iiia receptors [Title/Abstract] OR integrin alphaiibbeta3 [Title/Abstract] OR glycoproteins iib-iiia [Title/Abstract] OR pituitary growth hormone [Title/Abstract] OR somatotropin [Title/Abstract] OR growth hormone [Title/Abstract] OR haptoglobin [Title/Abstract] OR hemopexin [Title/Abstract] OR hemopexin [Title/Abstract] OR heparin cofactor ii [Title/Abstract] OR cd54 [Title/Abstract] OR cd54 antigens [Title/Abstract] OR icam-1 [Title/Abstract] OR cd54 antigen [Title/Abstract] OR intercellular adhesion molecule 1 [Title/Abstract] OR gamma globulin 7s [Title/Abstract] OR igg [Title/Abstract] OR allerglobuline [Title/Abstract] OR igg t [Title/Abstract] OR igg1 [Title/Abstract] OR igg2 [Title/Abstract] OR igg2a [Title/Abstract] OR igg2b [Title/Abstract] OR igg3 [Title/Abstract] OR igg4 [Title/Abstract] OR immunoglobulin [Title/Abstract] OR polyglobin [Title/Abstract] OR immunoglobulin g [Title/Abstract] OR insulin [Title/Abstract] OR ischemia modified albumin [Title/Abstract] OR merosin [Title/Abstract] OR glycoprotein gp-2 [Title/Abstract] OR laminin m [Title/Abstract] OR laminin m chain [Title/Abstract] OR laminin [Title/Abstract] OR leptin [Title/Abstract] OR ob protein [Title/Abstract] OR obese protein [Title/Abstract] OR ob gene product [Title/Abstract] OR obese gene product [Title/Abstract] OR lipoprotein associated phospholipase [Title/Abstract] OR lipoprotein lipase [Title/Abstract] OR csf-1 [Title/Abstract] OR csf-m [Title/Abstract] OR colony-stimulating factor 1 [Title/Abstract] OR colony-stimulating factor [Title/Abstract] OR m-csf [Title/Abstract] OR macrophage colony stimulating factor [Title/Abstract] OR malonaldehyde [Title/Abstract] OR propanedial [Title/Abstract] OR malonylaldehyde [Title/Abstract] OR malonyldialdehyde [Title/Abstract] OR sodium malondialdehyde [Title/Abstract] OR Malondialdehyde [Title/Abstract] OR interstitial collagenase [Title/Abstract] OR mmp-1 [Title/Abstract] OR mmp1 metalloproteinase [Title/Abstract] OR matrix metalloproteinase-1 [Title/Abstract] OR promatrix metalloproteinase-1 [Title/Abstract] OR promatrixmetalloproteinase-1 [Title/Abstract] OR prommp-1 [Title/Abstract] OR matrix metalloproteinase 1 [Title/Abstract] OR gelatinase a [Title/Abstract] OR 72-kda gelatinase [Title/Abstract] OR 72-kda type iv collagenase [Title/Abstract] OR mmp2 [Title/Abstract] OR mmp2 metalloproteinase [Title/Abstract] OR matrix metalloproteinase-2 [Title/Abstract] OR matrix metalloproteinase 2 [Title/Abstract] OR stromelysin 1 [Title/Abstract] OR transin [Title/Abstract] OR mmp-3 metalloproteinase [Title/Abstract] OR stromelysin [Title/Abstract] OR matrix metalloproteinase 3 [Title/Abstract] OR gelatinase b [Title/Abstract] OR 92-kda gelatinase [Title/Abstract] OR 92-kda type iv collagenase [Title/Abstract] OR mmp-9 metalloproteinase [Title/Abstract] OR mmp9 metalloproteinase [Title/Abstract] OR matrix metalloproteinase-9 [Title/Abstract] OR matrix metalloproteinase 9 [Title/Abstract] OR metallopeptidase 9 [Title/Abstract] OR monocyte chemoattractant protein-1 [Title/Abstract] OR myelin basic protein [Title/Abstract] OR encephalitogenic basic proteins [Title/Abstract] OR encephalitogenic basic protein [Title/Abstract] OR neuritogenic protein [Title/Abstract] OR myeloperoxidase [Title/Abstract] OR hemi-myeloperoxidase [Title/Abstract] OR Peroxidase [Title/Abstract] OR antigen s 100 [Title/Abstract] OR nerve tissue protein s 100 [Title/Abstract] OR s-100 protein [Title/Abstract] OR s100 protein family [Title/Abstract] OR s 100 [Title/Abstract] OR ngf-2 [Title/Abstract] OR nerve growth factor 2 [Title/Abstract] OR neurotrophin 3 [Title/Abstract] OR neutrophil gelatinase associated lipocalin [Title/Abstract] OR neutrophil protease 4 [Title/Abstract] OR gdp kinase [Title/Abstract] OR nucleoside diphosphokinases [Title/Abstract] OR nucleoside-diphosphate kinases [Title/Abstract] OR oxidized ldl [Title/Abstract] OR osteoprotogerin [Title/Abstract] OR aryl-dialkyl phosphatase [Title/Abstract] OR homocysteine thiolactone hydrolase [Title/Abstract] OR opa anhydrase [Title/Abstract] OR oph enzyme [Title/Abstract] OR organophosphorus acid anhydrase [Title/Abstract] OR organophosphorus acid anhydrolase [Title/Abstract] OR organophosphorus acid hydrolase [Title/Abstract] OR organophosphorus hydrolase [Title/Abstract] OR paraoxonase [Title/Abstract] OR paraoxonase-1 [Title/Abstract] OR paraoxonase-2 [Title/Abstract] OR

phosphoglyceromutase [Title/Abstract] OR phosphoglycerate phosphomutase [Title/Abstract] OR phosphoglycerate mutase [Title/Abstract] OR papp-a [Title/Abstract] OR igfbp-4 metalloproteinase [Title/Abstract] OR igfbp-4 protease [Title/Abstract] OR insulin-like-growth factor binding protein-4 protease [Title/Abstract] OR pregnancy associated alpha plasma protein [Title/Abstract] OR pregnancy-associated -plasma protein [Title/Abstract] OR pregnancy associated plasma protein a [Title/Abstract] OR profibrinolysin [Title/Abstract] OR glu-plasminogen [Title/Abstract] OR glutamic acid 1-plasminogen [Title/Abstract] OR glutamyl plasminogen [Title/Abstract] OR agepc [Title/Abstract] OR acetyl glyceryl ether phosphorylcholine [Title/Abstract] OR paf-acether [Title/Abstract] OR platelet aggregating factor [Title/Abstract] OR platelet aggregation enhancing factor [Title/Abstract] OR platelet-activating substance [Title/Abstract] OR thrombocyte aggregating activity [Title/Abstract] OR platelet activating factor [Title/Abstract] OR cd31 antigens [Title/Abstract] OR pecam-1 [Title/Abstract] OR platelet endothelial cell adhesion molecule-1 [Title/Abstract] OR cd31 antigen [Title/Abstract] OR platelet derived microvesicles [Title/Abstract] OR platelet derived growth factor [Title/Abstract] OR antiheparin factor [Title/Abstract] OR pf 4 [Title/Abstract] OR heparin neutralizing protein [Title/Abstract] OR pf4 [Title/Abstract] OR -thromboglobulin [Title/Abstract] OR prorenin [Title/Abstract] OR Protein C [Title/Abstract] OR Protein S [Title/Abstract] OR Protein Z [Title/Abstract] OR coagulation factor ii [Title/Abstract] OR factor ii [Title/Abstract] OR blood coagulation factor ii [Title/Abstract] OR differentiation reversal factor [Title/Abstract] OR prothrombin [Title/Abstract] OR resistin [Title/Abstract] OR RAGE [Title/Abstract] OR secretagogin [Title/Abstract] OR plasminogen activator inhibitors [Title/Abstract] OR bae-pai [Title/Abstract] OR endothelial plasminogen activator inhibitor [Title/Abstract] OR plasminogen activator inhibitor [Title/Abstract] OR platelet activation [Title/Abstract] OR PARK [Title/Abstract] OR SCD40L [Title/Abstract] OR Tau [Title/Abstract] OR thrombase [Title/Abstract] OR thrombinar [Title/Abstract] OR thrombostat [Title/Abstract] OR -thrombin [Title/Abstract] OR thrombin [Title/Abstract] OR thrombin-antithrombin complex [Title/Abstract] OR Thrombomodulin [Title/Abstract] OR cd142 [Title/Abstract] OR coagulation factor iii[Title/Abstract] OR factor iii [Title/Abstract] OR tissue factor [Title/Abstract] OR tissue thromboplastin [Title/Abstract] OR blood coagulation factor [Title/Abstract] OR coagulin [Title/Abstract] OR glomerular procoagulant activity [Title/Abstract] OR prothrombinase [Title/Abstract] OR tissue factor procoagulant [Title/Abstract] OR urothromboplastin [Title/Abstract] OR thromboplastin [Title/Abstract] OR tissue factor pathway inhibitor [Title/Abstract] OR tissue inhibitor of metalloproteinase [Title/Abstract] OR cachectin [Title/Abstract] OR tnf-alpha [Title/Abstract] OR tumor necrosis factor ligand superfamily [Title/Abstract] OR cachectin-tumor necrosis factor [Title/Abstract] OR tnf superfamily [Title/Abstract] OR tumor necrosis factor [Title/Abstract] OR bone-derived transforming growth factor [Title/Abstract] OR platelet transforming growth factor [Title/Abstract] OR tgf-beta [Title/Abstract] OR milk growth factor [Title/Abstract] OR tgfbeta [Title/Abstract] OR Transforming Growth Factor beta [Title/Abstract] OR ubiquitin fusion degradation protein 1 [Title/Abstract] OR Vascular Endothelial Growth Factor A [Title/Abstract] OR vascular endothelial growth factor [Title/Abstract] OR vascular endothelial growth factor-a [Title/Abstract] OR gd-vegf [Title/Abstract] OR glioma-derived vascular endothelial cell growth factor [Title/Abstract] OR vegf [Title/Abstract] OR vegf-a [Title/Abstract] OR vascular permeability factor [Title/Abstract] OR vasculotropin [Title/Abstract] OR vitronectin [Title/Abstract] OR factor viii-related antigen [Title/Abstract] OR f viii-vwf [Title/Abstract] OR factor viiir-ag [Title/Abstract] OR factor viiir-rco [Title/Abstract] OR plasma factor viii complex [Title/Abstract] OR ristocetin cofactor [Title/Abstract] OR ristocetin-willebrand factor [Title/Abstract] OR vwf ag [Title/Abstract] OR von willebrand protein [Title/Abstract] OR von Willebrand Factor

11. DIAGNOSIS: Diagnos* [Title/Abstract] OR sensitivity [Title/Abstract] OR specificity [Title/Abstract] OR odds ratio [Title/Abstract] OR likelihood ratio [Title/Abstract] OR LR [Title/Abstract]

11. PROGNOSIS: Incidence [Title/Abstract] OR mortality [MeSH term] OR follow up studies OR mortality OR prognos* [Title/Abstract] OR predict* [Title/Abstract] OR course [Title/Abstract] OR rankin [Title/Abstract] OR Glasgow outcome scale [Title/Abstract] OR NIHSS [Title/Abstract]

12. #5 OR #6 OR #7 OR #8 OR #9 OR #10

13. #4 AND #11 AND #12

14. Limit #13 to humans

Biomarkers for Prognosis After Thrombolysis

1.thrombolytic therapy [Title/Abstract] OR thromboly* [Title/Abstract] OR fibrinoly* [Title/Abstract] OR fibrinolytic agents [Title/Abstract] OR plasmin [Title/Abstract] OR plasminogen [Title/Abstract] OR tPA [Title/Abstract] OR t-PA [Title/Abstract] OR rtPA [Title/Abstract] OR rt-PA [Title/Abstract] OR recanalisation [Title/Abstract] OR clot lysis [Title/Abstract]

2. plasminogen activators [MeSH Term]

3. intracranial embolism [MeSH Term] OR thrombosis [MeSH Term] OR Thromboembolism [MeSH Term]

4. 1 OR 2 OR 3

5. "biological markers" [MeSH Term]

6. biomarker* [Title/Abstract]

7. (biochemical [Title/Abstract] OR clinical [Title/Abstract] OR immun* [Title/Abstract] OR laboratory [Title/Abstract] OR biological* [Title/Abstract] OR [Title/Abstract] serum [Title/Abstract] OR [Title/Abstract] surrogate [Title/Abstract] OR viral [Title/Abstract]) AND (marker* [Title/Abstract])

8. (blood [Title/Abstract] OR plasma [Title/Abstract]) AND (marker [Title/Abstract])

9. incidence [Title/Abstract] OR mortality [MeSH term] OR follow up studies OR mortality OR prognos* [Title/Abstract] OR predict* [Title/Abstract] OR course [Title/Abstract] OR rankin [Title/Abstract] OR Glasgow outcome scale [Title/Abstract] OR NIHSS [Title/Abstract]

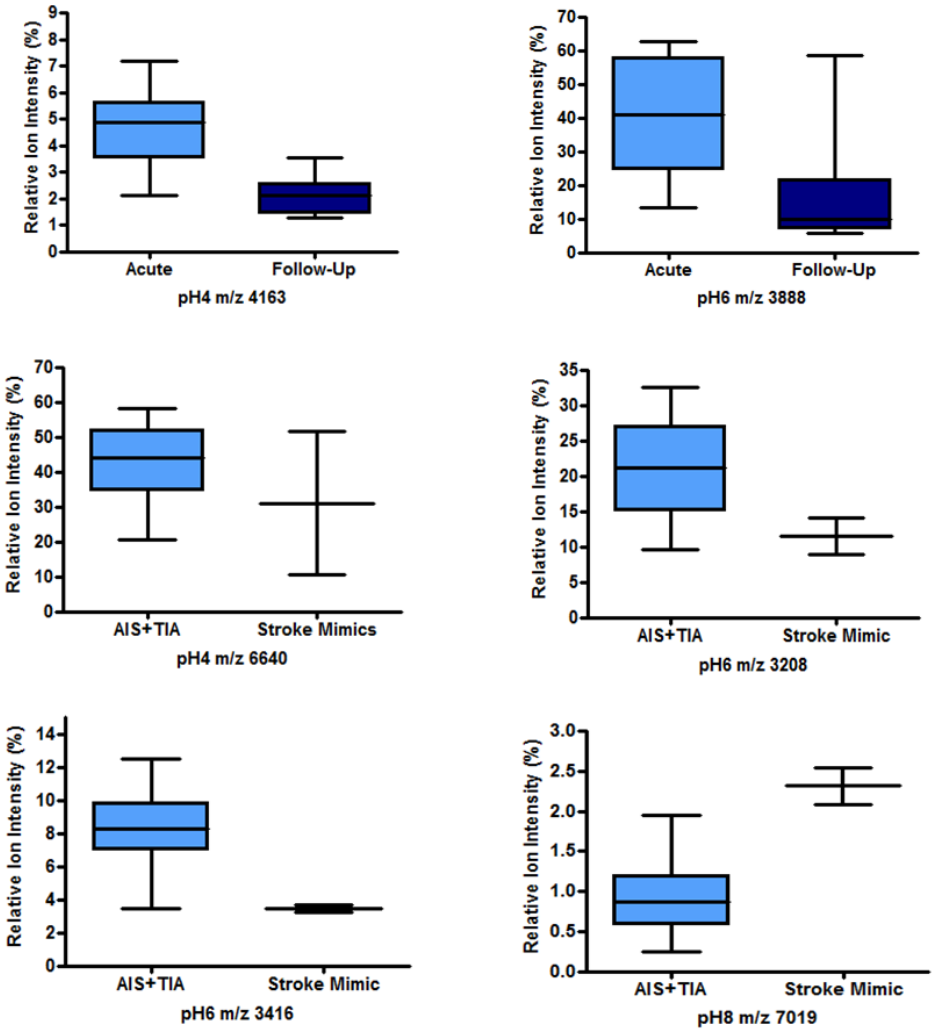
10. #5 OR #6 OR #7 OR #8

11. #4 AND #9 AND #10

12. Limit #11 to humans

APPENDIX 2

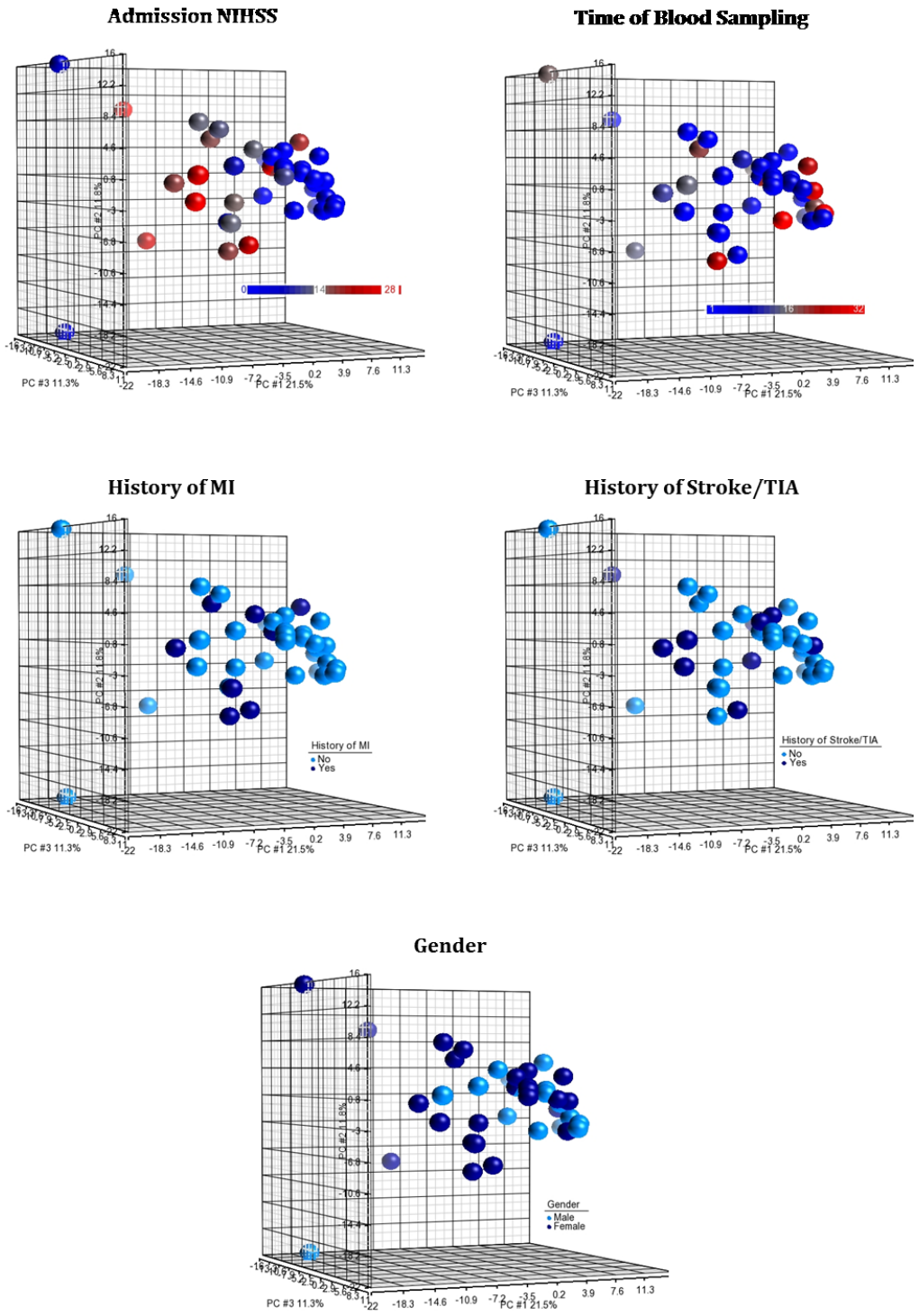
Putative Biomarkers Identified from the Pilot Study (Chapter 3)



Appendix 2: Box and whisker plots of putative diagnostic biomarkers identified from the pilot study. Six protein ions from the pilot study were selected as 'putative markers' from the pilot study. Protein ions were captured using binding conditions of pH4 (m/z 4163, m/z 6640), pH6 (m/z 3888, m/z 3208, m/z 3416) and pH8 (m/z 7019), on CM10 ProteinArray chips with SPA matrix. The putative markers were selected based upon visual inspection of their ability to distinguish between study groups, as well as degree of significance (Wilcoxon Signed Rank test p-value) and fold difference between study groups. Box and whisker plots depict mean relative ion intensity +/- two standard deviations. Associated data is detailed in Table 3.2.

APPENDIX 3

PCA Analysis of Training Set Data (Chapter 3)



Appendix 3: Unsupervised analysis of Training Set data by principal component analysis (PCA). Unsupervised analysis of all protein ions obtained at pH4 binding conditions from the Training Set data (comprising 20 AIS patients, 10 TIA patients and 10 stroke mimics) performed on a number of demographic variables (admission NIHSS, time of blood sampling,, history of MI, stroke or TIA and gender) did not demonstrate an ability for classification exhibited by diagnosis (Figure 3,5). This indicated that differential protein ion expression can be useful for patient diagnostic classification, independent of other confounding variables.