University of Glasgow

Alghamdi, Jahad (2016) Epidemiology and genetic architecture of blood pressure: a family based study of generation Scotland. PhD thesis

http://theses.gla.ac.uk/7540/

Copyright and moral rights for this thesis are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the Author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the Author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

Glasgow Theses Service
http://theses.gla.ac.uk/
theses@gla.ac.uk

Epidemiology and Genetic Architecture of Blood Pressure – A Family Based Study of Generation Scotland

Jahad Ahmed Alghamdi

BSc., MSc.

Submitted in fulfilment of requirements for the degree of Doctor of Philosophy (Ph.D.) in the Faculty of Medicine, University of Glasgow

August 2016

BHF Glasgow Cardiovascular Research Centre
Institute of Cardiovascular and Medical Sciences
College of Medical, Veterinary, and Life Sciences
University of Glasgow

©J.A. Alghamdi 2016

Declaration

I declare that this thesis has been written entirely by myself and is a record of research performed myself, with the exception of DNA extractions, which were performed by Generation Scotland staff at the Wellcome Trust Clinical Research Facility, University of Edinburgh. Data linkage was performed with contribution from Mr Archie Campbell from Generation Scotland. This work has not been submitted previously for a higher degree and was carried out under the supervision of Professor Sandosh Padmanabhan.

Jahad A. Alghamdi

Acknowledgments

Firstly, I would like to express my sincere gratitude to my supervisor Professor Sandosh Padmanabhan for his continual support and guidance throughout my PhD study. His suggestions and comments have proved invaluable to the work involved in this thesis. In addition, I would like to extensively thank him for the opportunities and experiences afforded to me during my PhD study by assisting me to attend conferences and advanced courses organized by respected institutions such as the Wellcome Trust and Cold Spring Harbor Laboratory. I would like to extend my thanks to Dr Wai Kwong Lee for the time spent teaching me the required genotyping and laboratory skills. Many thanks to Generation Scotland for providing the resources and support to complete this study, and for all the people who participated in the study.

Secondly, my deepest gratitude goes to my parents (Ahmed & Refa'h) for their immeasurable support, prayers, and patience during my time abroad. Special thanks to my wife (Bashayer) and my daughter (Aroob) for all their support and sacrifices during the last four years. Thanks also goes to my brothers (Abdullah & Zoheer) and sisters (Zohoor, Jawaher & Abeer) for great support and patience in my absence. A profound thanks and appreciations goes to my friends and colleagues in Glasgow, with whom I had a pleasant, enjoyable and fruitful company. To all of these wonderful people in my life, I would not have achieved this without you all along the way.

Finally, an extended thanks to King Abdullah International Medical Research Centre (KAIMRC) and Saudi Arabia Cultural Bureau (SACB) in London for their financial and academic support. Thanks also to University of Glasgow and Institutes of Cardiovascular and Medical Sciences for supporting this research and providing me with necessary infrastructure and excellent research environment.

Table of Contents

Declaration 2
Acknowledgments 3
Table of Contents 4
List of Tables 7
List of Figures 8
List of abbreviations, Acronyms & symbols 10
Publications and conference abstracts 16
Abstract 17
1 Introduction 20
1.1 Cardiovascular disease 21
1.2 Blood pressure 22
1.2.1 BP physiology and its regulation 22
1.2.2 BP components 24
1.2.3 BP measurements 25
1.3 Hypertension 28
1.3.1 Pathophysiology of hypertension 31
1.3.2 Hypertension risk factors 33
1.3.3 Distribution of BP 35
1.3.4 Prevalence of hypertension 38
1.3.5 Prevalence of hypertension in Scotland 38
1.3.6 Global management of hypertension 39
1.3.7 The public health importance of blood pressure 42
1.4 Principles of basic genetics 44
1.5 Fundamental of complex trait genetics and association studies 47
1.5.1 Definition of phenotype 50
1.5.2 Family- or population-based sample 52
1.5.3 Descriptive genetic epidemiology 55
1.5.4 Linkage analysis 58
1.5.5 Association studies 59
1.5.6 Genome-wide association study (GWAS) 63
1.5.7 Power considerations 67
1.6 Genetics of BP and hypertension 70
1.6.1 Monogenic forms of hypertension 70
1.6.2 Overview of GWAS for BP and hypertension 77
1.6.3 Challenges of GWAS for BP and hypertension 87
1.7 Study aims 92
2 Materials and Methods 93
2.1 Introduction to this chapter 94
2.2 Generation Scotland 94
2.2.1 Participants 96
2.3 Phenotypic data 98
2.3.1 Blood pressure phenotypes 98
2.3.2 Drug exposure 99
2.3.3 Anthropometric measures 103
2.3.4 Social and demographic data 103
2.3.5 Family health history 103
2.3.6 Phenotype Quality Control 104
2.4 Genotypic data 106

2.4.1 Extraction and storage of DNA ... 106
2.4.2 Genotyping procedure ... 107
2.5 SNP Selection ... 116
2.6 Ethical approval ... 117

3 Quality control of GS:SFHS ... 119
3.1 Introduction about this chapter ... 120
3.2 Method ... 120
3.3 Results ... 121
3.3.1 Excluding individuals based on genotype call rate ... 121
3.3.2 Excluding individuals based on missing phenotypes ... 121
3.3.3 Excluding individuals based on medication history ... 121
3.3.4 QC for SNP ... 122
3.3.5 Family data review ... 122
3.3.6 BP values data check ... 126
3.3.7 Identifying participants taking BP-lowering medications based on EPRs ... 126
3.3.8 Identifying participants taking BP-lowering medications based on SRMs ... 126
3.4 Discussion ... 130

4 Study of BP phenotypes ... 133
4.1 Introduction ... 134
4.1.1 Introduction to this chapter ... 134
4.1.2 Pharmacoepidemiological analysis ... 134
4.1.3 Prevalence of hypertension-related health outcomes in GS:SFHS ... 135
4.2 Aims of this chapter ... 136
4.3 Methods ... 136
4.3.1 Overview of the method ... 136
4.3.2 Assessing the validity of the SRMs ... 136
4.3.3 Definitions of hypertension related indicators ... 141
4.3.4 Estimating of hypertension-related indicators prevalence ... 141
4.3.5 Assess the effect of socio-economic factors in hypertension-related indicators ... 142
4.4 Results ... 143
4.4.1 Participants baseline characteristics ... 143
4.4.2 Pharmacoepidemiological analysis ... 143
4.4.3 Validity of SRMs ... 147
4.4.4 Distribution of BP in GS:SFHS ... 150
4.4.5 Prevalence of hypertension ... 153
4.4.6 Prevalence of hypertension awareness ... 153
4.4.7 Prevalence of treated hypertension ... 154
4.4.8 Prevalence of controlled hypertension ... 154
4.4.9 Determinants of hypertension related indicators ... 158
4.5 Discussion ... 160
4.5.1 Quality of SRMs history ... 160
4.5.2 Prevalence of hypertension-related health indicators ... 163
4.5.3 Conclusion ... 166

5 Heritability and familial aggregation of BP and hypertension ... 168
5.1 Introduction to this chapter ... 169
5.1.1 Familial correlation and heritability of BP ... 169
5.1.2 Sibling recurrence risk ratio (λ_S) ... 170
5.2 Chapter aims ... 171
5.3 Statistical analysis ... 171
5.3.1 Estimate of familial correlation of BP traits ... 172
5.3.2 Calculating the univariate heritability of BP traits ... 173
5.3.3 Calculating the bivariate genetic correlation for BP traits ... 174
5.3.4 Estimate of sibling recurrence risk of hypertension and treated hypertension ... 174

5.4 Results **176**
5.4.1 Familial correlation of BP traits 176
5.4.2 Univariate heritability of BP traits 180
5.4.3 Bivariate genetic correlation 180
5.4.4 Sibling Recurrence Risk ratio (λ_S) 181
5.5 Discussion **184**
5.5.1 Conclusion 188

6 SNP association analysis **190**
6.1 Introduction **191**
6.1.1 Aims 191
6.2 Methods **191**
6.2.1 DNA extraction, genotyping, and SNP selection 191
6.2.2 BP phenotypes 192
6.2.3 Statistical analyses 193
6.3 Results **195**
6.3.1 SNPs association test results for SBP 195
6.3.2 SNPs association test results for DBP 201
6.3.3 SNPs association test results for MAP 207
6.3.4 SNPs association test for PP 213
6.3.5 Association of genetic risk scores with BP traits 219
6.4 Discussion **225**
6.4.1 SNP association with BP traits 225
6.4.2 Influence of phenotype measurement errors on association signal 232
6.4.3 SNPs with potential pharmacogenetic effects 234
6.4.4 Impacts of genetic risk scores on BP traits 235

7 General discussion and conclusion **238**

Appendix **249**

References **256**

List of Tables

Table 1-1 Factors affecting the accuracy of BP measures 27
Table 1-2 Definition and classification of hypertension based on ESH2013 and JNC7 30
Table 1-3 Prevalence of hypertension, awareness, treatment, and control in PURE Study 41
Table 1-4 Monogenic forms of hypertension 73
Table 1-5 Loci reported in GWAS for association with hypertension or BP traits 82
Table 2-1 Comparison between GS:SFHS cohort and the Scottish population 97
Table 2-2 List of selected SNPs for genotyping 118
Table 3-1 Summary statistics of SNPs QC 124
Table 3-2 Age- and gender-specific ranges of GS:SFHS and HSE 129
Table 4-1 Definitions and formulas of the concordance analysis measures 140
Table 4-2 Baseline characteristics of study participants 145
Table 4-3 Frequency of BP-lowering medications classes based on EPRs 146
Table 4-4 Concordance analyses of the two medication history sources in participants with BP-lowering medication history sources 148
Table 4-5 Age-standardized prevalence of BP classes in men and women 150
Table 4-6 Crude prevalence of hypertension, awareness, treatment, and control by baseline characteristics 155
Table 4-7 Age- and sex-adjusted prevalence of hypertension, awareness, treatment, and controlled hypertension 156
Table 4-8 Multivariate association of characteristics with hypertension, awareness, treatment, and control 159
Table 5-1 Familial correlation coefficients for different relative pairs 179
Table 5-2 Univariate heritability of BP traits 182
Table 5-3 Phenotypic, genetic, and environmental correlations for BMI and BP traits ... 182
Table 5-4 Number of siblings with hypertension per sibship 183
Table 5-5 Number of siblings with treated hypertension per sibship. 183
Table 6-1 SNPs association results for SBP 196
Table 6-2 SNPs association results for DBP 202
Table 6-3 SNPs association results for MAP 208
Table 6-4 SNPs association results for PP 214
Table 6-5 Association of the GRS with BP traits 220

List of Figures

Figure 1-1 Illustration of the Platt-Pickering debate.. 29
Figure 1-2 The Paige mosaic model of BP regulation and the updated mosaic 2014. 32
Figure 1-3 Distribution of unadjusted mean of the four BP components with age for men and women. 36
Figure 1-4 Distribution of mean SBP across the world for men and women in 2008. 37
Figure 1-5 Trends in hypertension-related ratio between 2003 and 2013 in Scotland. 40
Figure 1-6 Flowchart of a systematic approach to dissect the genetic basis of complex disease. 49
Figure 1-7 Linkage and linkage disequilibrium. 61
Figure 1-8 Direct and indirect association. 62
Figure 1-9 Traits examined by GWAS and functional distribution of GWAS variants. 65
Figure 1-10 Statistical association methods for different type of samples. 69
Figure 2-1 The overall strategy of the study. 95
Figure 2-2 Flowchart of the QC procedure. 105
Figure 2-3 Dynamics of allelic discrimination in the TaqMan® genotyping assay. 111
Figure 2-4 Layout of the OpenArray 384-Well Sample Plate, and the OpenArray Genotyping Plate. 112
Figure 2-5 Genotyping experiment workflow. 113
Figure 3-1 Flowchart of QC results and classification of individuals based on the sources of medication history. 123
Figure 3-2 Family size and number of generations per pedigree in GS:SFHS................ 125
Figure 3-3 Histogram for BP traits with normal distribution curve on the plot. 127
Figure 3-4 Box-Whisker plot with jitter-dots for BP traits. 128
Figure 4-1 Venn diagram of the status of high-risk population in GS for BP level, treatment based on EPRs and SRMs, and self-reported hypertension. 149
Figure 4-2 Distribution of the mean of SBP and DBP by sex and age. 151
Figure 4-3 Box plots of SBP and DBP levels by sex and hypertensive status. 152
Figure 4-4 Age- and sex stratified prevalence of hypertension, awareness, treatment, and control by age groups 157
Figure 5-1 Mean values of BP traits per family. 177
Figure 5-2 Familial correlation coefficients for all the four BP traits. 178
Figure 6-1 Effect size of the coded allele for SBP in GS:SFHS. 198
Figure 6-2 SNPs association with SBP for the different adjustment models. 199

Figure 6-3 Coded allele effect size for SBP in the different adjustment models............. 200
Figure 6-4 Effect size of the coded allele for DBP in GS:SFHS..................................... 204
Figure 6-5 SNPs association with DBP for the different adjustment models.................. 205
Figure 6-6 Coded allele effect size for DBP in the different adjustment models............. 206
Figure 6-7 Effect size of the coded allele for MAP in GS:SFHS.................................... 210
Figure 6-8 SNPs association results with MAP for the different adjustment models 211
Figure 6-9 Coded allele effect size for MAP in the different adjustment models............ 212
Figure 6-10 Effect size of the coded allele for PP in GS:SFHS 216
Figure 6-11 SNPs association results with PP for the different adjustment models......... 217
Figure 6-12 Coded allele effect size for PP in the different adjustment models. 218
Figure 6-13 GRS for SBP and DBP.. 221
Figure 6-14 GRS for MAP and PP 222
Figure 6-15 Prevalence of hypertension by GRS Quartiles... 223
Figure 6-16 Hypertension treatment by GRS Quartiles... 223
Figure 6-17 Hypertension controlled by GRS Quartiles.. 224
Figure 6-18 Hypertension awareness by GRS Quartiles 224

List of abbreviations, Acronyms & symbols

AA	African American
AA1Blockers	Alpha 1-Adrenoceptor Blockers
ABPM	Ambulatory BP Monitoring
ACE	Angiotensin-converting enzyme
ACEIs	Angiotensin Converting Enzyme Inhibitors
ACTH	Adrenocorticotropic hormone
AD	Autosomal dominant
ADAMTS8	A disintegrin-like and metalloproteinase with thrombospondin type 1 motif, 8
ADM	Adrenomedullin
ADRB1	Beta-1-adrengeric receptor
AGEN-BP	Asian Genetic Epidemiology Network Blood Pressure
AGT	Angiotensinogen
ALDH2	Aldehyde Dehydrogenase 2
AME	Apparent mineralocorticoid excess
ANOVA	Analysis of Variance
ANP	Atrial natriuretic peptide
APA	Aldosterone-producing adenoma
AR	Autosome recessive
ARBs	Angiotensin receptors blockers
ASPs	Affected sib pairs
ATG7	Autophagy related 7
ATP1A1	ATPase, Na+/K+ transporting, alpha 1 polypeptide
ATP2B1	ATPase, Ca++ transporting, plasma membrane 1
ATP2B3	ATPase, Ca++ transporting, plasma membrane 3
BAG1	BCL2-associated athanogene 1
BBs	Beta-Blockers
BI	Bias index
BLK	B-lymphocyte specific, tyrosine kinase
BMI	Body Mass Index
BNF	British National Formulary
BP	Blood Pressure
C10orf107	Chromosome 10 open reading frame 107
C20orf174	Chromosome 20 open reading frame 174
C5orf23	Chromosome 5 open reading frame 23
Ca^{2+}	Calcium ion
CACNA1D	Calcium channel, voltage-dependent, L type, alpha 1D subunit
CACNB2	Calcium channel, voltage-dependent, beta 2 subunit
CAPZA1	Capping protein (actin filament) muscle Z-line, alpha 1
CASR	Calcium-sensing receptor
CASZ1	Castor zinc finger 1
CCBs	Calcium-Channel Blockers
CCT	Cortical collecting tubule
CDK6	Cyclin-dependent kinase 6
CHARGE	Cohorts for Heart and Aging Research in Genomic Epidemiology
CHD	Chronic heart disease
CHIC2	Cysteine rich hydrophobic domain 2

CI	Confidence intervals
CLCN6	Chloride channel, voltage-sensitive 6
CLCNKB	Chloride channel, voltage-sensitive Kb
CNVs	Copy number variants
CO	Cardiac output
COGENT	The Continental Origins and Genetic Epidemiology Network study
COX5A	Cytochrome C oxidase subunit Va
CRIP3	Cysteine-rich protein 3
CUL3	Cullin 3
CVD	Cardiovascular diseases
CYP1A1	Cytochrome P450 family 1 subfamily A member 1
CYP17A1	Cytochrome P450 family 17 subfamily A member 1
CYP21A2	Cytochrome P450 family 21 subfamily A member 2
CYP11B1	Cytochrome P450 family 11 subfamily B member 1
CYP11B2	Cytochrome P450 family 11 subfamily B member 2
DALY	Disability-adjusted life years
DARC	Duffy antigen gene
DBP	Diastolic blood pressure
DCT	Distal convoluted tubule
DNA	Deoxyribonucleic acid
EBF1	Early B-cell factor 1
EDN3	Endothelin 3
EDTA	Ethylenediaminetetraacetic Acid
EHBP1L1	EH domain binding protein 1-like 1
ENaC	Epithelial NA+ Channel
ENPEP	Glutamyl aminopeptidase
EPF1	Epidermal patterning factor 1
EPRs	Electronic-prescription records
ESH	European Society of Hypertension
EVX1	Even-skipped homeobox 1
FBAT	Family-based association test
FBN1	Fibrillin 1
FCOR	Family correlation function
FER1L5	Fer-1-like family member 5
FES	Feline sarcoma viral oncogene
FGF5	Fibroblast growth factor 5
FH	Familial Hyperaldosteronism
FIGN	Fidgetin
FLJ32810	Rho GTPase activating protein 42
GATA4	GATA binding protein 4
GBD	Global burden of disease
GCTA	Genome-Wide Complex Trait Analysis
Global BPgen	Global Blood Pressure Genetic consortium
GNAS	GNAS complex locus
GOSR2	Golgi SNAP receptor complex member 2
GRS	Genetic risk score
GS	Generation Scotland
GS:21CGH	Generation Scotland: Genetic Health In The 21st Century
GS:3D	Generation Scotland: Donor DNA Databank
GS:SFHS	Generation Scotland: Scottish Family Health Study
GUCY1A3	Guanylate cyclase 1, soluble, alpha 3

GUCY1B3	Guanylate cyclase 1, soluble, beta 3
GWAS	Genome-Wide Association Studies
H_0	Null hypothesis
H_1	Alternative hypothesis
h^2	Narrow sense heritability
H^2	Broad sense heritability
h^2_{SNP}	Phenotypic Variance Explained By Common SNPs
HBPM	Home BP Monitoring
HFE	Hemochromatosis
HIC	High-income countries
HLA-DQB1	Major histocompatibility complex, class II, DQ beta 1
HOXA	Homeobox A
HOXC4	Homeobox C4
hr	Hour
HRH1	Histamine receptor H1
HSD11B2	Hydroxysteroid (11-beta) dehydrogenase 2
HSD3B2	Hydroxy-delta-5-steroid dehydrogenase, 3 beta- and steroid delta-isomerase 2
HSE	Health Survey For England
HWE	Hardy-Weinberg Equilibrium
IBD	Identical by descent
ICBP	International Consortium of Blood Pressure consortium
ICC	Intra-class correlation coefficient
IGFBP3	Insulin like growth factor binding protein 3
JAG1	Jagged 1
JNC7	Seventh Report of the Joint National Committee
K^+	Potassium ion
KCN	Potassium channel genes
KCNJ1	Potassium channel, inwardly rectifying subfamily J, member 1
KCNJ5	Potassium channel, inwardly rectifying subfamily J, member 5
KCNK3	Potassium channel, two pore domain subfamily K, member 3
kg	Kilogram
KLHL3	Kelch-like family member 3
LD	Linkage disequilibrium
LIC	Low-income countries
LMIC	Low-middle income countries
LMMs	Linear-mixed models
LOD	Logarithm of the odds
LSP1	Lymphocyte-specific protein 1
LTA	Long-term averaged
m^2	Meter square
MAF	Minor allele frequency
MAP	Mean arterial pressure
MAP4	Microtubule associated protein 4
MBG	Minor Groove Binder
MDM4	Mouse double minute 4 homolog
MECOM	MDS1 and EVI1 complex locus
MED13L	Mediator complex subunit 13-like
Mg^{2+}	Magnesium ion
MIC	Middle-income countries
MIR1263	microRNA 1263
ml	Millilitre

mM	Millimole
mmHg	Millimetre of mercury
MOV10	Mov10 RISC complex RNA helicase
MR	Mineralocorticoid Receptor
mRNA	Messenger RNA
MTHFR	Methylenetetrahydrofolate reductase
n	Sample size
Na^+	Sodium ions
NCC	Thiazide-sensitive Na^+/Cl^- co-transporter
NCR3LG1	Natural killer cell cytotoxicity receptor 3 ligand 1
NFAT5	Nuclear factor of activated T-cells 5, tonicity-responsive
ng	Nanogram
NHGRI	National Human Genome Research Institute
NHS	National Health Service
NKCC2	$Na^+/K^+/2Cl^-$ co-transporter
NLGN1	Neuroligin 1
NO	Nitric oxide
NOS3	Nitric oxide synthase 3
NOV	Nephroblastoma overexpressed
NPPA	Natriuretic peptide A
NPPB	Natriuretic peptide B
NPR3	Natriuretic peptide receptor 3
NPT	Negative Predictive Value
NR3C2	Nuclear receptor subfamily 3 group C member 2
NT5C2	5'-Nucleotidase, cytosolic II
NUCB2	Nucleobindin 2
OR	Odds ratio
P	Age-standardized prevalence
PABAK	Prevalence-Adjusted Bias-Adjusted Kappa
PCQ	Pre-Clinic Questionnaire
PCQ-1	Pre-Clinic Questionnaire Phase 1
PCQ-2	Pre-Clinic Questionnaire Phase 2
PCR	Polymerase chain reaction
PCT	Proximal convoluted tubule
PDE1A	Phosphodiesterase-1A
p_e	Expected agreement by chance
PGL	Paragangliomas
PHA	Pseudohypoaldosteronism
PHA1A	Pseudohypoaldosteronism Type 1
PHA2B	Pseudohypoaldosteronism Type 2B, "Gordon Syndrome"
PHA2D	Pseudohypoaldosteronism Type 2D
PHA2E	Pseudohypoaldosteronism Type 2E
PI	Prevalence index
PIK3C2A	Phosphatidylinositol-4-phosphate 3-kinase catalytic subunit type 2 alpha
PIK3CG	Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit gamma
PITX2	Paired-like homeodomain 2
PLCD3	Phospholipase C delta 3
PLCE1	Phospholipase C epsilon 1
PLEKHA7	Pleckstrin homology domain containing A7
PLEKHG1	Pleckstrin homology and RhoGEF domain containing G1

P_{neg}	Observed negative agreement
p_o	Observed agreement
PP	Pulse pressure
P_{pos}	Observed positive agreement
PPT	Positive predictive value
PURE	Prospective Urban Rural Epidemiology study
PVR	Peripheral vascular resistance
QC	Quality control
QOF	Quality and Outcome Framework
QTDT	Quantitative transmission disequilibrium test
RAAS	Renin-angiotensin-aldosterone system
RELA	V-rel avian reticuloendotheliosis viral oncogene homolog A
RET	Ret proto-oncogene
RNA	Ribonucleic acid
ROMK	Renal outer medullary potassium channel
RPL6	Ribosomal protein L6
r^2	familial correlation coefficients
rpm	Revolutions per minute
RSPO3	R-spondin 3
S.A.G.E	Statistical Application for Genetic Epidemiology
SBP	Systolic Blood Pressure
SCNN1B	Sodium channel, non voltage gated 1 beta subunit
SCNN1G	Sodium channel, non voltage gated 1 gamma subunit
SD	Standard deviation
SDH	Succinate Dehydrogenase Enzyme Complex
SDHA	Succinate dehydrogenase complex subunit A
SDHAF2	Succinate dehydrogenase complex assembly factor 2
SDHB	Succinate dehydrogenase complex subunit B
SDHC	Succinate dehydrogenase complex subunit C
SDHD	Succinate dehydrogenase complex subunit D
se	Standard error
SH2B3	SH2B adaptor protein 3
SHeS	Scottish Health Surveys
SIMD	Scottish-index for multiple deprivation
SLC12A1	Solute carrier family 12 (Na/K/Cl transporter), member 1
SLC12A3	Solute carrier family 12 (Na/K/Cl transporter),member 3
SLC16A1	Solute carrier family 16 (monocarboxylate transporter), member 1
SLC16A9	Solute carrier family 16 member 9
SLC39A8	Solute carrier family 39 (metal ion transporter), member 8
SLC4A7	Solute carrier family 4, sodium bicarbonate co-transporter, member 7
SNPs	Single nucleotide polymorphisms
SOLAR	Sequential Oligogenic Linkage Analysis Routines
SOX6	Sex Determining Region Y-box 6
SRMs	Self-reported medications
ST7L	Suppression of tumorigenicity 7 like
STK39	Serine threonine kinase 39
TAL	Thick ascending limb of Henle
TBX3	T-Box 3
TBX5	T-box 5
TDT	Transmission disequilibrium test

Tm	Melting temperature
TMEM133	Transmembrane protein 133
TNNT3	Troponin T type 3
ULK3	Unc-51 Like Kinase 3
ULK4	Unc-51 like kinase 4
UMOD	Uromodulin gene
VCL	Vinculin
VHL	von Hippel-Lindau tumour suppressor
WHO	World Health Organization
WNK1	Lysin deficient protein kinase 1
WNK4	Serine-threonine protein kinase WNK4
WTCCC	Wellcome Trust Case Control Consortium
YLD	Years lived with disability
YLS	Years of life lost
ZNF652	Zinc finger protein 652
α	Statistical significance level
β	Effect size
κ	Cohen’s Kappa Coefficients Statistics
λ_R	Relatives recurrence risk ratio
λ_S	Siblings recurrence risk ratio
µl	Microlitre
σ^2_A	Additive genetic component
σ^2_E	Environmental component
σ^2_G	Total genetic component
σ^2_p	Total phenotypic variance
ρ_E	Environmental correlation
ρ_G	Additive genetic correlation
ρ_P	Phenotypic correlation

Publications and conference abstracts

Publications

1. Quinn T. J., **Alghamdi J.**, Padmanabhan S., Porteous D. J., Smith B. H., Hocking L., et al. Association between cognition and gene polymorphisms involved in thrombosis and haemostasis. Age. Springer International Publishing; 2015 Aug;37(4):9820–9.

2. **Alghamdi J.,** Padmanabhan S., Fundamentals of Complex Trait Genetics and Association Studies. In: Handbook of Pharmacogenomics and Stratified Medicine. Elsevier; 2014. pp. 235–57.

Conference Abstracts

1- **Alghamdi J.,** Generation Scotland, Padmanabhan S., Estimating univariate and bivariate heritabilities of blood pressure traits in Generation Scotland. Presented at the 8th Saudi Student Conference in UK, December 28, 2014 in London, UK.

2- Hastie C., **Alghamdi J.,** Hocking L., Luciano M., Porteous D., Morris A. D., Smith B. H., Dominiczak A. F., Padmanabhan S., Family-based study of prevalence, awareness, treatment, and control of hypertension in Generation Scotland. Presented at the 23rd European Meeting on Hypertension and cardiovascular protection, June 15, 2013, Milan, Italy.

3- Brown C., Hastie C. E., **Alghamdi J.,** Schulz C., Hocking L. J., Luciano M., Porteous D., Morris A., Smith B. H., Generation Scotland, Dominiczak A. F., Tobias E. S., Delles C., Padmanabhan S., Predictors of QTC and QTC prolongation in the Generation Scotland family study. Presented at the 23rd European Meeting on Hypertension and cardiovascular protection, June 15, 2013, Milan, Italy.

4- Schulz C., Hastie C., Brown C., **Alghamdi J.,** Hocking L., Luciano M., Porteous D., Morris A., Smith B., Generation Scotland, Dominiczak A. F., Padmanabhan S., Bivariate heritability analysis of cardiac conduction and repolarisation measures demonstrates a paucity of shared genetic factors. Presented at the 23rd European Meeting on Hypertension and cardiovascular protection, June 15, 2013, Milan, Italy.

Abstract

Hypertension is a major risk factor for cardiovascular disease and mortality, and a growing global public health concern, with up to one-third of the world’s population affected. Despite the vast amount of evidence for the benefits of blood pressure (BP) lowering accumulated to date, elevated BP is still the leading risk factor for disease and disability worldwide. It is well established that hypertension and BP are common complex traits, where multiple genetic and environmental factors contribute to BP variation. Furthermore, family and twin studies confirmed the genetic component of BP, with a heritability estimate in the range of 30-50%. Contemporary genomic tools enabling the genotyping of millions of genetic variants across the human genome in an efficient, reliable, and cost-effective manner, has transformed hypertension genetics research. This is accompanied by the presence of international consortia that have offered unprecedentedly large sample sizes for genome-wide association studies (GWASs). While GWAS for hypertension and BP have identified more than 60 loci, variants in these loci are associated with modest effects on BP and in aggregate can explain less than 3% of the variance in BP.

The aims of this thesis are to study the genetic and environmental factors that influence BP and hypertension traits in the Scottish population, by performing several genetic epidemiological analyses. In the first part of this thesis, it aims to study the burden of hypertension in the Scottish population, along with assessing the familial aggregation and heritialbity of BP and hypertension traits. In the second part, it aims to validate the association of common SNPs reported in the large GWAS and to estimate the variance explained by these variants.

In this thesis, comprehensive genetic epidemiology analyses were performed on Generation Scotland: Scottish Family Health Study (GS:SFHS), one of the largest population-based family design studies. The availability of clinical, biological samples, self-reported information, and medical records for study participants has allowed several assessments to be performed to evaluate factors that influence BP variation in the Scottish population. Of the 20,753 subjects genotyped in the study, a total of 18,470 individuals (grouped into 7,025 extended families) passed the stringent quality control (QC) criteria and were

available for all subsequent analysis. Based on the BP-lowering treatment exposure sources, subjects were further classified into two groups. First, subjects with both a self-reported medications (SRMs) history and electronic-prescription records (EPRs; n =12,347); second, all the subjects with at least one medication history source (n =18,470). In the first group, the analysis showed a good concordance between SRMs and EPRs (kappa =71%), indicating that SRMs can be used as a surrogate to assess the exposure to BP-lowering medication in GS:SFHS participants. Although both sources suffer from some limitations, SRMs can be considered the best available source to estimate the drug exposure history in those without EPRs. The prevalence of hypertension was 40.8% with higher prevalence in men (46.3%) compared to women (35.8%). The prevalence of awareness, treatment and controlled hypertension as defined by the study definition were 25.3%, 31.2%, and 54.3%, respectively. These findings are lower than similar reported studies in other populations, with the exception of controlled hypertension prevalence, which can be considered better than other populations. Odds of hypertension were higher in men, obese or overweight individuals, people with a parental history of hypertension, and those living in the most deprived area of Scotland. On the other hand, deprivation was associated with higher odds of treatment, awareness and controlled hypertension, suggesting that people living in the most deprived area may have been receiving better quality of care, or have higher comorbidity levels requiring greater engagement with doctors. These findings highlight the need for further work to improve hypertension management in Scotland.

The family design of GS:SFHS has allowed family-based analysis to be performed to assess the familial aggregation and heritability of BP and hypertension traits. The familial correlation of BP traits ranged from 0.07 to 0.20, and from 0.18 to 0.34 for parent-offspring pairs and sibling pairs, respectively. A higher correlation of BP traits was observed among first-degree relatives than other types of relative pairs. A variance-component model that was adjusted for sex, body mass index (BMI), age, and age-squared was used to estimate heritability of BP traits, which ranged from 24% to 32% with pulse pressure (PP) having the lowest estimates. The genetic correlation between BP traits showed a high correlation between systolic (SBP), diastolic (DBP) and mean arterial pressure (MAP) (ρ_G: 81% to 94%), but lower correlations with PP (ρ_G: 22% to 78%). The

sibling recurrence risk ratio (λ_S) for hypertension and treatment were calculated as 1.60 and 2.04 respectively. These findings confirm the genetic components of BP traits in GS:SFHS, and justify further work to investigate genetic determinants of BP.

Genetic variants reported in the recent large GWAS of BP traits were selected for genotyping in GS:SFHS using a custom designed TaqMan® OpenArray®. The genotyping plate included 44 single nucleotide polymorphisms (SNPs) that have been previously reported to be associated with BP or hypertension at genome-wide significance level. A linear mixed model that is adjusted for age, age-squared, sex, and BMI was used to test for the association between the genetic variants and BP traits. Of the 43 variants that passed the QC, 11 variants showed statistically significant association with at least one BP trait. The phenotypic variance explained by these variant for the four BP traits were 1.4%, 1.5%, 1.6%, and 0.8% for SBP, DBP, MAP, and PP, respectively. The association of genetic risk score (GRS) that were constructed from selected variants has showed a positive association with BP level and hypertension prevalence, with an average effect of one mmHg increase with each 0.80 unit increases in the GRS across the different BP traits.

The impact of BP-lowering medication on the genetic association study for BP traits has been established, with typical practice of adding a fixed value (i.e. 15/10 mmHg) to the measured BP values to adjust for BP treatment. Using the subset of participants with the two treatment exposure sources (i.e. SRMs and EPRs), the influence of using either source to justify the addition of fixed values in SNP association signal was analysed. BP phenotypes derived from EPRs were considered the true phenotypes, and those derived from SRMs were considered less accurate, with some phenotypic noise. Comparing SNPs association signals between the four BP traits in the two model derived from the different adjustments showed that MAP was the least impacted by the phenotypic noise. This was suggested by identifying the same overlapped significant SNPs for the two models in the case of MAP, while other BP traits had some discrepancy between the two sources

1 Introduction

1.1 Cardiovascular disease

The global burden of disease (GBD) has substantially shifted from communicable diseases in children to non-communicable diseases in adults. This was highlighted in the GBD Study 2010, which clearly stated in the executive summary:

> "Infectious diseases, maternal and child illness, and malnutrition now cause fewer deaths and less illness than they did twenty years ago. As a result, fewer children are dying every year, but more young and middle-aged adults are dying and suffering from disease and injury, as non-communicable diseases, such as cancer and heart disease, become the dominant causes of death and disability worldwide. Since 1970, men and women worldwide have gained slightly more than ten years of life expectancy overall, but they spend more years living with injury and illness".[1]

The GBD 2010 study collated and analysed global data to estimate the deaths and disability-adjusted life years (DALYs; sum of years lived with disability [YLDs] and years of life lost [YLLs]). The above quotation reflects the finding that high blood pressure (BP) has shifted from being the fourth-highest cause of DALYs in 1990 to be the leading risk factor for disease in 2010, an estimated 7.0% [95% CI: 6·2–7.0] of global DALYs.[1] For instance, the number of people around the world with hypertension was estimated to be nearly one billion in 2000, resulting in more than seven million premature deaths and ninety-two million DALYs worldwide.[2,3] This shift is due to the aging population and the lower mortality rate in children below five years of age, as well as other factors, such as changes in cause-of-death composition and risk factor exposure. Yet, the extent of this shift varies greatly across the world, as it is not applied in much of sub-Saharan Africa, where the leading risk factors are still related to infectious diseases.[1]

The mortality rate from non-communicable diseases has increased from 60% in 2000, to 68% of global death in 2012. According to the World Health Organization (WHO) 2012 statistical report, cardiovascular disease (CVD) was the highest cause of death with an estimated 17.5 million deaths a year (that is 3 in every 10 deaths); of these, complications of hypertension are responsible for 9.4 million deaths, and hypertension accounts for at least 45% of deaths due to heart disease and 51% of deaths due to stroke.[4] The risk of stroke for middle-

aged individuals (40 to 60 years) is doubled for each increment of 20/10 mmHg blood pressure.[5] Furthermore, even high-normal values of BP were associated with an increased risk of CVD.[6] Consequently, evaluation of the current practice of hypertension management with rigorous research to expand our knowledge of the pathogenesis and risk factors of hypertension are fundamental for future treatment and prevention of the disease.

1.2 Blood pressure

1.2.1 BP physiology and its regulation

BP is the pressure exerted by circulatory blood on the walls of blood vessels. The principal function of circulation is to deliver nutrients to and remove wastes from tissue; this is achieved through blood flow effected by the difference in pressure that occurs during the pumping action of the heart. The relationship between pressure and blood flow is analogous to Ohm’s law, which can be clinically represented as BP being directly proportionate to the product of the blood flow [i.e. cardiac output (CO)] and total peripheral vascular resistance (PVR) [mean arterial pressure (MAP) =CO x PVR]. One way to control BP is by regulating CO and PVR in three anatomic sites: arterioles, postcapillary venules (capacitance vessels), and heart. The kidney is the fourth anatomic site that controls BP by regulating the volume of intravascular fluids. These four anatomic sites are under the control of baroreflexes, which are mediated by autonomic nerves to accommodates acute change in BP, and humoral mechanisms such as the renin-angiotensin-aldosterone system (RAAS) that affects volume homeostasis and vascular tone, and the adrenergic receptor system that affects heart rate, vascular tone, and cardiac contraction, and kinin-kallikrein system that influences vascular tone and renal salt handling.[7] In addition, vascular endothelium may also release local vasoactive substances to regulate the vascular resistance, such as the vasodilator nitric oxide, or a vasoconstrictor endothelin-1. All of these multiple physiological systems act in an integrated complex manner to ensure homeostasis of BP in all metabolizing tissue.

RAAS has a central role in regulation of BP, and is supported by various pharmacological agents that lower BP by blocking its activity, such as renin inhibitors, angiotensin-converting enzyme (ACE) inhibitors, and angiotensin

receptors blockers (ARBs). RAAS begins with the biosynthesis of renin in the form of prorenin, the inactive precursor of renin, by the juxtaglomerular cells in the kidney. Renin converts angiotensinogen, which is mainly released from the liver, to form the biologically inert angiotensin I that is rapidly hydrolysed by ACE enzyme to angiotensin II. Angiotensin II is a potent vasoconstrictor that increases BP by several mechanisms: it binds to angiotensin II-type 1 receptor to stimulate several tyrosine kinases, which in turn phosphorylate the tyrosine residues in several proteins, causing vasoconstriction, cell growth, and cell proliferation.[8] In addition, it raises BP by increasing blood volume directly by inducing water and sodium reabsorption, and indirectly by stimulating the release of aldosterone which further raises BP.[9,10] The degree of RAAS activity is variable between individuals and it depends on several factors such as sex, ethnicity, salt intake, genetic components, and the uses of medication.[11] Particularly, BP-lowering medication that acts on RAAS are less effective in people with low level of RAAS, for example, Blacks and elderly.

The second system that has an important role in regulating BP is the sympathetic nervous system, which can cause both arteriolar constriction and arteriolar dilation. Higher activity of this system can increase BP by acting on the heart, peripheral vasculature, and kidneys, leading to increased cardiac output, vascular resistance, and fluid retention.[8] The roles of this system in regulating BP are complex and include alteration in both baroreflex and chemoreflex at central and peripheral level, that mainly affect the short-term changes in BP in response to stimuli such as physical exercise or stress. The potent BP lowering effect of pharmacological drugs such as alpha- and beta-sympathetic blockers highlights its role in controlling BP. Importantly, it was found that stimulation of beta-adrenergic receptors leads to reduced kidney expression of the serine-threonine protein kinase WNK4, which is a regulator of the sodium-chloride cotransporter, a target of thiazide diuretics.[12,13]

Vascular endothelial cells have also been implicated in BP regulation and CVD, by releasing potent local vasoactive substances, including the vasodilator nitric oxide (NO) and the vasoconstrictor peptide endothelin. Endothelial dysfunction is a phenotypical alteration of endovascular lining of blood vessels that is characterized by a pro-thrombotic, pro-inflammatory, and pro-constrictive

phenotype.[14] It occurs over time from aging and hypertension, leading to increased arterial stiffness and structural abnormality that may be irreversible once established.[8,9] Though, deficiency of NO can be restored by antihypertensive therapy such as nitrates, which increases arterial compliance and dispensability and lowering systolic blood pressure (SBP) but not diastolic blood pressure (DBP).[8,9] Other vasoactive substances that may act on different systems are also important in regulation of BP, an example is bradykinin that is a potent vasodilator which is inactivated by ACE enzyme, meaning ACE inhibitors (ACEIs) may also lower BP by blocking bradykinin inactivation.[9,10] Atrial natriuretic peptide (ANP) is a cardiac hormone that is secreted from the atria of the heart in response to increased blood volume; ANP has a natriuretic, diuretic and vasodilatory properties, where deficiency of this hormone may cause fluid retention and hypertension.

1.2.2 BP components

The two components of blood pressure are SBP and DBP. SBP is the maximum pressure exerted against arteries and vessels during contraction of the left ventricle, thus SBP mainly depends on the CO. DBP is the minimum pressure on the walls of arteries during the time that the left ventricle is relaxing and refilling, and just before the ventricle ejects blood into the aorta. The mean of SBP and DBP during the cardiac cycle is the time-weighted average arterial pressure, and is called MAP, which represents the steady components of BP that reflects CO, vascular resistance, elasticity averaged over time, and heart rate.[15-17] MAP can be determined directly by catheterization or can be calculated by the formula (DBP + ⅓ x [SBP-DBP]). This is because ventricles spend approximately one-third (⅓) of their time in systole, and two-thirds (⅔) in diastole in each cardiac cycle. The pulsatile components of BP are represented by pulse pressure (PP), which occurs as a result of the outward and inward movements of the arterial walls during the SBP and DBP. PP is the difference between SBP and DBP, and reflects the large artery stiffness, ventricular ejection, and the timing of wave reflection. It is influenced by other haemodynamic mechanisms such as the change in ventricular ejection, large artery compliance, and timing of reflected waves.[15-17].

The relative importance of these four components have established a continuing debate in the field of hypertension in regard to which one should be used to estimate CVD risk. DBP has historically been considered the most important components of BP, as it was determined by pooled data from 420,000 individuals that showed a log-linear association between DBP and risk of stroke and myocardial infarction.[18] The importance of SBP over DBP has been later recognized with the publication of a series of epidemiologic studies showing that SBP is the best predictor of risk in the elderly, and that low DBP is also associated with higher risk of CVD. The Framingham Heart Study was the first study to show that there is a decline in importance of DBP and a corresponding increase of importance of SBP with age.[19] Since then, several studies have shown the superiority of either SBP or PP in the elderly, whereas DBP is a better predictor in subjects younger than 50 years old.[20] PP has also been reported to have a possible additional prognostic role, as it was found that patients with high SBP and low or normal DBP (i.e. isolated systolic hypertension) have a higher risk of CVD.[21] Generally, age plays an important factor in determining which BP component is a better predictor of CVD, with findings from the Framingham Heart Study suggesting that with increasing age there is a gradual shift from DBP to SBP and then to PP as predictor of CVD risk, and that combining two components such as DBP and SBP, or MAP and PP provides a better prediction than any single BP component.[20,22]

1.2.3 BP measurements

BP is traditionally measured by non-invasive methods using the auscultatory technique (Korotkoff sounds) with a mercury sphygmomanometer that depends on the transient occlusion of the brachial artery by an appropriately sized cuff inflated over the upper arm. Although this may be considered the “gold standard”, it is no longer applied in most of the European countries due to concerns regarding mercury toxicity. The advancement of technologies has allowed automated devices, which should be carefully used after validation and calibration to ensure accurate measurements.[23] Currently, two types of BP measurements are used: office-BP measurement and out-of-office measurement. While the office measurement is always performed by a specialized person in the clinic and involves single or multiple readings, out-of-

office provides larger number of observations and it performed in less stressful environments and may represent a more reliable estimate of the actual BP.

The out-of-office BP measurement can be performed by either ambulatory BP monitoring (ABPM) or home BP monitoring (HBPM). ABPM is performed by using a portable device that measures BP at frequent intervals for 24 hours, where the cuff is attached to a small electric recording device. The HBPM can be considered as a cost-effective alternative to the ABPM, which has the advantage of not interfering with a patient's daily activity. The process of measuring BP with HBPM is exactly the same as office BP measurement. Although ABPM provides more complete information of BP than HBPM, both of them are superior to office measurements in regard to predicting CVD risk.[24,25]

1.2.3.1 BP measurements errors

The presence of these different BP monitoring methods highlights the importance of taking accurate BP measurements. The challenge of obtaining a correct measurement of BP is due to the considerable variability of BP because of issues involving the observer (measurement variation), or biological factors within the patient that may occur from moment to moment with respiration, emotion, smoking, temperature, pain, meals, and bladder distension.[26] Factors that are related to variation in the measurement include white-coat effect, suboptimal measurement procedures such as inappropriate cuff size, or rounding bias (Table 1-1).[27] Obtaining an accurate measurement of BP is of great clinical importance as inaccuracy could expose normotensive patients to unnecessary treatments, or it may deprive hypertensive patients of a useful treatment. It has been estimated that a systematic error of underestimating of true BP by 5 mmHg would mislabel 20 million persons as having normal BP, disqualifying them from receiving treatment. On the other hand, a systematic error of overestimated BP by 5 mmHg could misclassify 27 million people as being hypertensive.[28]

Table 1-1 Factors affecting the accuracy of BP measures

Factor	**Magnitude of SBP/DBP discrepancy (mmHg)**
Talking or active listening	10/10
Distended bladder	15/10
Cuff over clothing	5–50/
Cuff too small	10/2–8
Smoking within 30 minutes of measurement	6–20/
Paralyzed arm	2–5/
Back unsupported	6–10/
Arm unsupported, sitting	1–7/5–11
Arm unsupported, standing	6-8/
Table is reproduced from [29]	

1.3 Hypertension

The current agreement of the quantitative nature of hypertension has resulted from a large number of epidemiological studies, which showed that BP is normally distributed in the population. However, this agreement was not always accepted especially during the 1950s, with competing views expressed by Sir Robert Platt and Sir George Pickering, known later as the “Platt versus Pickering debate”.[30-32] Platt examined BP in normotensive and hypertensive probands and their relatives, and concluded that hypertension is a qualitative disease that follows a bimodal distribution, and that the hypertensive subpopulation can be distinguished from the normotensive majority. On the other hand, Pickering believed that BP is normally distributed in the population with a unimodal distribution, in which there is no discernible separation between normotensive and hypertensive subjects. Hence, those people in the extreme top of the distribution can be classified as having hypertension, but it does not exist as a separate entity. Pickering view was supported with the observation that BP is normally distributed in the first-degree relatives of normotensive and hypertensive probands; he concluded that BP was inherited as a “graded character” with a polygenic nature. At the beginning, Platt’s qualitative model was much preferred, but it was only changed with mounting evidence from epidemiological studies showing the benefit of reducing BP. Despite the rejection of the Platt model, evidence from the rare monogenic forms of hypertension do indicate that there exists a small sub-population of people with hypertension where the cause of hypertension is due to a single gene defect and thus existing as a separate entity.

The consensus that hypertension is a quantitative disease implies that any blood level used to define hypertension is arbitrary. Basically, hypertension is defined based on a cut-off at the upper extreme of the BP distribution at which investigation and treatment do more good than harm. In practical terms, a conventional cut-off of 140/90 mmHg are used to simplify the diagnostic approach and to facilitate the decision about starting treatment.[23,33,34] The threshold for hypertension diagnosis and classification is slightly different between different guidelines; the most common classifications are based on the Seventh Report of the Joint National Committee (JNC7), and the European Society of Hypertension (ESH) guidelines 2013 (Table 1-2).[23,35] It must be noted

that "JNC8" was published during the writing of this thesis, and it differs from the earlier version in that it no longer addresses the old definition of hypertension but it only states the threshold for treatment intervention.[36]

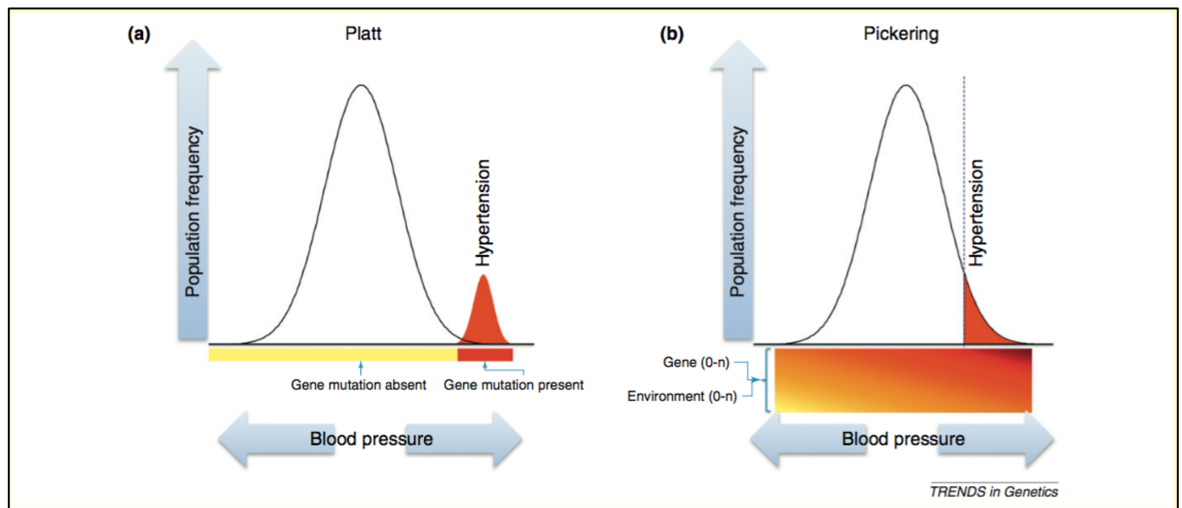


Figure 1-1 Illustration of the Platt-Pickering debate.
(a) Platt argued that hypertensive population represents a discrete subpopulation, appeared due to a single, heritable genetic mutation. (b) Pickering debated that hypertensive population represents the extreme top of the normally distributed population, with no clear separation line between hypertensive and normotensive, suggesting that hypertension is a polygenic disease. Figure is reproduced with permission from [37].

Table 1-2 Definition and classification of hypertension based on ESH2013 and JNC7

1. Definition of hypertension based on ESH2013 and JNC7			
Category	SBP		DBP
Office BP	≥ 140	and/or	≥ 90
ABPM (awake)	≥ 135	and/or	≥ 85
ABPM (asleep)	≥ 120	and/or	≥ 70, (≥ 75 in JNC7)
ABPM (24-hr)	≥ 130	and/or	≥ 80
HBPM	≥ 135	and/or	≥ 85
2. Classification of hypertension			
A. Based on ESH2013			
Optimal	<120	and	<80
Normal	120–129	and/or	80–84
High normal	130–139	and/or	85–89
Grade 1 hypertension	140–159	and/or	90–99
Grade 2 hypertension	160–179	and/or	100–109
Grade 3 hypertension	≥ 180	and/or	≥ 110
Isolated systolic hypertension	≥ 140	and	<90
B. Based on JNC7			
Normal	<120	and	<80
Prehypertension	120–139	or	80–89
Stage 1 hypertension	140–159	or	90–99
Stage 2 hypertension	≥ 160	or	≥ 100
Table is produced based on information from ESH2013[23] and JNC7[35]			

1.3.1 Pathophysiology of hypertension

The pathophysiology of hypertension involves complex interacting systems and pathways that maintain BP homeostasis (some of these system were described in Section 1.2.1 -p22). These interactions may even involve unknown pathways yet to be discovered. Primary or essential hypertension is the most common type of hypertension, estimated to cause 90% of the cases, where the patient has high BP without an obvious secondary cause. The remaining cases are classified as secondary hypertension (<10%), which is an elevation in BP that is due to specific causes such as renovascular disease, hyperaldosteronism, phaeochromocytoma, monogenic disorders, or medication-induced. The enormous advances made in the genetic studies of hypertension and BP regulation affirms the multifactorial polygenic nature of hypertension as represented by the Mosaic Theory of hypertension proposed by Paige in 1960, who indicated that hypertension is due to an interaction of genetic, environmental, adaptive, neural, mechanical, and hormonal perturbations (Figure 1-2).[38]

There is clear evidence that the kidney plays an important role in the pathogenesis of hypertension, especially with its capacity to rapidly reduce BP by increasing urinary sodium execration, a phenomenon known as “pressure natriuresis”; this mechanism plays an important role to balance BP, regardless of the initiating cause.[13] The concept of pressure natriuresis was suggested by Arthur Guyton and his colleagues in the 1970’s, who proposed that the kidney controls the level of BP through regulating extracellular fluid volume, by matching urinary excretion of salt and water with dietary intake.[39] In addition, the presence of different BP-lowering agents that target pathways in the kidney, such as RAAS; or that inhibit sodium reabsorption such as thiazide diuretics affirm the role of kidney in hypertension pathogenesis. Lastly, the majority of the identified Mendelian forms of hypertension are due to mutations that alter renal hemodynamics or tubular reabsorption.[38,40] This hypothesis of the central role of the kidney is supported by experimental kidney cross-transplantation studies, where long-term BP level followed the kidney donor.[13] These studies showed that long-term BP level is elevated in the genetically normotensive recipient controls, after transplanting a kidney from genetically hypertensive

donor; whereas, BP level is reduced when a genetically hypertensive recipient receives a kidney from a genetically normotensive donor.

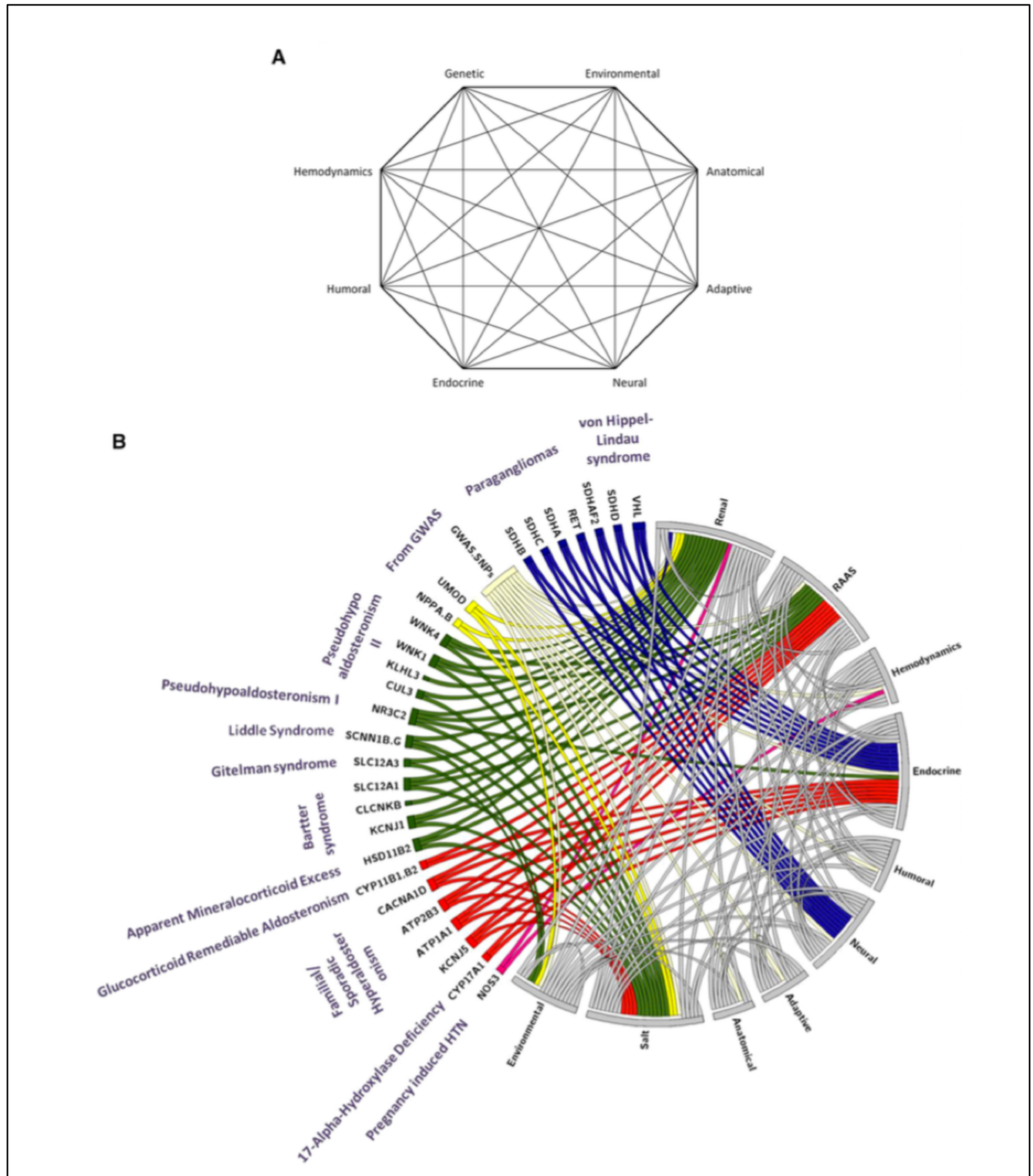


Figure 1-2 The Paige mosaic model of BP regulation and the updated mosaic 2014.
A. Paige mosaic model of BP regulation in 1960, B. The updated BP mosaic of BP regulation which shows the several types of monogenic hypertension. Figure is reproduced with permission from [38].

1.3.2 Hypertension risk factors

Although the aetiology of essential hypertension remains unclear, some of the main risk factors that elevate BP are known, including obesity and overweight,[41] family history,[42] race (more common in black people until the age of 75 years),[43] excess sodium intake (e.g. >3000 mg/day),[44] insulin resistance, aging,[23] stress, sedentary lifestyle, low potassium intake,[44] and low calcium intake.[45] The influence of these risk factors on BP and hypertension risk will be briefly explained in this section, except for the monogenic forms of hypertension, which will be explained in a separate section.

Advancing age is associated with hypertension, particularly an increase in SBP in both genders. Women tend to have lower SBP than men before the menopause by around 6/3 mmHg, but slightly higher than men after the menopause. Yet, women tend to have greater pulsatile load with increasing age.[46] The differences in hypertension prevalence between genders may be explained by this differences in BP level, in addition to differences in other factors related to awareness, treatment, and control of hypertension (see also Section 1.3.3).[47]

Ancestry has a major impact on blood pressure. BP level tends to be higher in people of African descent, who have a more severe and higher prevalence of hypertension and an earlier onset of hypertension. This difference is particularly observed between Blacks and Whites in the 45 to 74 year age-range, but not after age 75 years.[43] Effects of stressors were found to be greater in Blacks, especially during childhood, leading to greater long-term variability in BP in Blacks.[48] A study that used a 24-hour ABPM showed a significant ethnic difference, even after adjusting for height, body mass index (BMI), socioeconomic status, and stress-related coping styles. This study demonstrated that Blacks had higher BP levels during both night and day, but the differences were significantly greater at night than during the day.[49] That is, Blacks had higher nocturnal BP and smaller difference between daytime BP, a terminology that is called “nondipping” and it is associated with higher risk of vascular disease.[50] The difference between Whites and Asians is controversial with most studies reporting similar prevalence of hypertension and BP level.[51]

The impact of obesity and overweight on BP and hypertension is well established for both children and adults. For instance, the Framingham Heart Study showed that obesity and overweight accounted for about 26% and 28% of hypertension cases in men and women, respectively; and the risk of new onset of hypertension was 1.5 for overweight men (BMI, 25–29.9 kg/m^2) and 2.2 for obese men (BMI, >30 kg/m^2).[41] It was also estimated that there is a 20–30% higher risk of developing hypertension for every 5% increment in weight gain.[52] Obesity increases the risk of hypertension through the interaction of several factors including dietary, genetic, epigenetic, and environmental factors. The elevation of BP in obese subjects is initially associated with raising cardiac output, which would also be the same in normotensive obese people. However, the systemic vascular resistance tends to be relatively normal in hypertensive obese patients, and lower than normal in normotensive obese people.[53] Vascular and systemic insulin resistance is usually a result of adipocyte dysfunction in the obese patients, which is accompanied with dysfunction in sympathetic nervous system and RAAS. Also, structural and functional changes in the kidney such as activation of intrarenal angiotensin II are important factors in pathogenesis of obesity induced hypertension.[52] Possible other mechanisms that link obesity to hypertension include diet, sodium retention, metabolic factors, neuroendocrine imbalances, proteinuria, endothelial and vascular dysfunction, glomerular hyperfiltration, and maladaptive immune and inflammatory responses.[52] Importantly, the combination of these two conditions has two implications; first, that affected people are at a higher risk of morbidity and mortality from CVD. Second, they are at higher risk of treatment-resistance and may require multiple medications.[54]

Dietary factors have received much attention for their contribution in hypertension risk. The dietary factors can also be influenced by genetic factors, as the association between salt intake and BP is mediated by the subject’s salt sensitivity, which is partly genetically determined. A diet that is high in sodium salt intake has been shown to increase BP; the Dietary Approaches to Stop Hypertension study showed that reduction in sodium intake resulted in a stepwise decrease in BP, in subjects who were allocated to receive food with high, intermediate, and low level of sodium for 30 consecutive days.[55] A meta-analysis of long-term trials (at least four weeks duration) showed that reducing

sodium intake by 3 gram per day decreased BP by an average 4/2.5 mmHg in hypertensive, and 2/1 mmHg in normotensive individuals.[56] The diet of isolated populations tends to be rich in potassium with a low level of sodium, and their hypertension prevalence is less than 1%; on the other hand, people of industrialized countries on average consume more processed food, which is sodium-rich and potassium-poor, and approximately one-third are affected by hypertension.[44]

Excessive alcohol consumption is associated with the development of hypertension and higher levels of BP and BMI, and reduction of alcohol consumption was associated with a significant reduction in BP by an average of 3.3/2.04 mmHg.[57,58] The risk factors listed above can themselves be multifactorial, affected by genetic and environments factors, even though some of them would be considered predominantly environmental factors that can be minimized by lifestyle modification. For instance, alcohol intake would be determined by consumption; yet, some individuals may inherit a genetic variant in the aldehyde dehydrogenase 2 gene *(ALDH2)*, which influences the rate of alcohol degradation and metabolism. Individuals homozygous for this variant experience severe adverse symptoms of alcohol consumption and are less likely to drink. Through the concept of “Mendelian randomization” it was found that men homozygous for the slow metabolism variant are less likely to develop hypertension, and their SBP is lower by 4.24 mmHg and 7.44 mmHg than those heterozygous or homozygous for the other variant, respectively.[59] Furthermore, factors such as alcoholism and obesity tend to be influenced by both genetic and environmental factors, leading to more confounding factors in studying the proportion of BP variability that is caused by inheritance, and that they may vary in different populations.

1.3.3 Distribution of BP

Studies of the BP distribution with age showed a trend of steadily increasing SBP with age in both sexes, and an increase of DBP with age until the fifth decade and then progressively reduction in both sexes. MAP increases steadily with age and reach a relative plateau in the seventh decade, whereas PP remained relatively constant until the fifth decade, after which it steeply increases in men and women (Figure 1-3).[46] The age-related increase of SBP was observed in a

sub-group of a German study after including only participants with no antihypertensive usage and no CVD risk factors (n =22,550), suggesting that this increase is part of the normal ageing process that occurs independently of known risk factors.[60] The distribution of BP between genders shows that women tend to have lower SBP than men before the menopause by around 6/3 mmHg, but slightly higher than men after the menopause.[46] Average SBP differs widely between world regions with a difference up to 16.8 mmHg and 19.4 mmHg between the lowest and highest regions for men and women, respectively (Figure 1-4).[61] During the last 30 years, the global age-standardized SBP has declined by 1 mmHg per decade from 1980 to 2008, but trends varied significantly across regions and countries. SBP is currently highest in low-income and middle-income countries.[61]

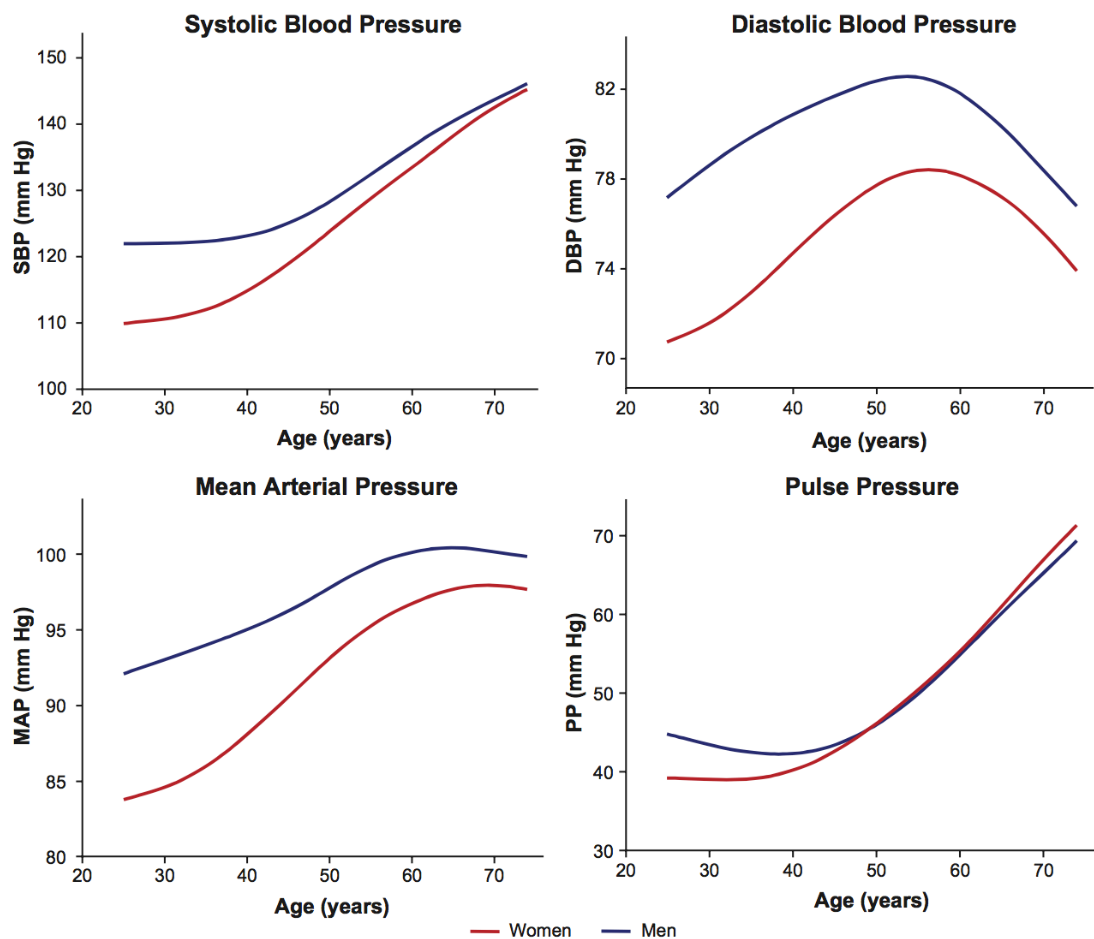


Figure 1-3 Distribution of unadjusted mean of the four BP components with age for men and women.
Figure is reproduced with permission from [46].

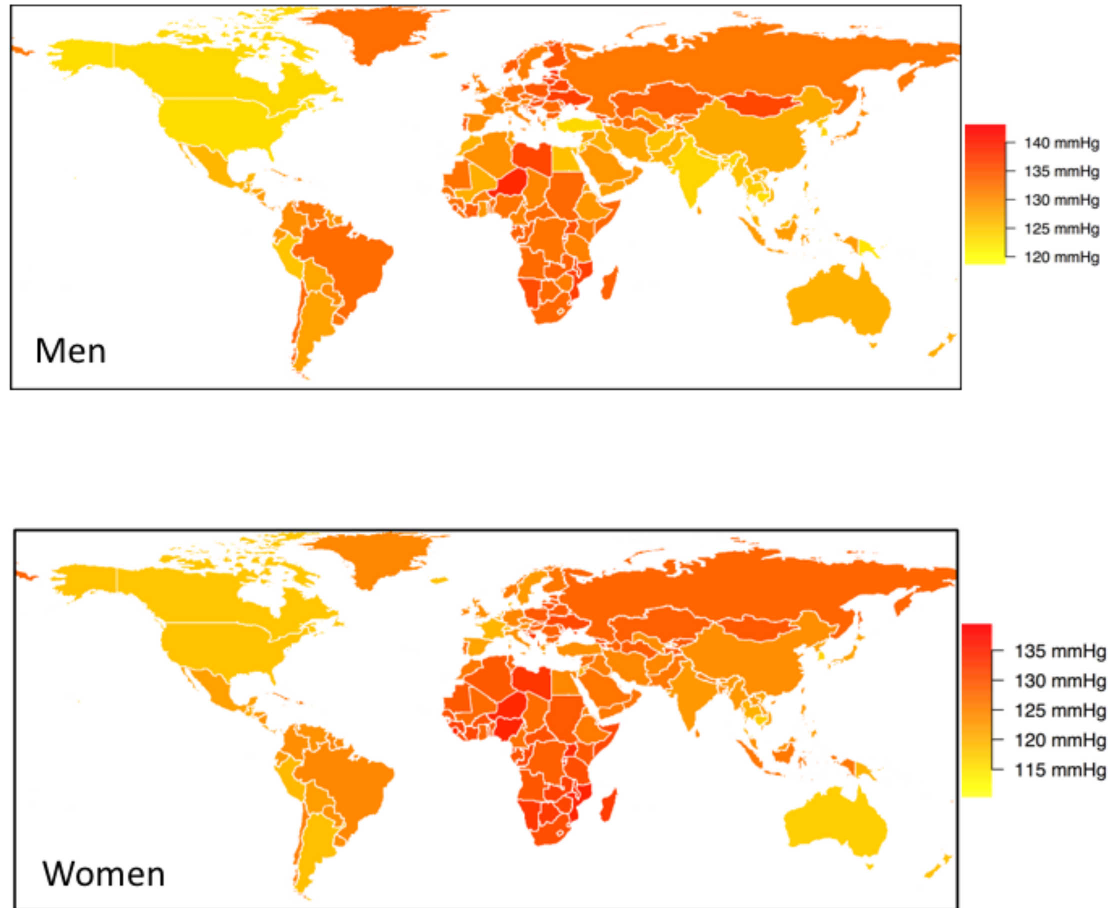


Figure 1-4 Distribution of mean SBP across the world for men and women in 2008. ***Figure is reproduced with modification from [61].***

1.3.4 Prevalence of hypertension

Estimation of the global prevalence of hypertension in the year 2000 demonstrated that more than a quarter of the world's adult population had hypertension, nearly one billion cases, and it was expected that this would increase to 29%, or 1.56 billion, by 2025.[3] The prevalence of hypertension varies widely across populations and regions of the world. For instance, mean prevalence of hypertension among men in developing countries was statistically lower by 6.5% than those in developed countries.[62] Moreover, a study that pooled data from different regions of the world showed that regions with the highest estimated prevalence of hypertension had roughly twice the rate of regions with the lowest estimated prevalence.[3] Moreover, an international comparison showed that hypertension is substantially higher in European countries than North American countries; for instance, England had higher prevalence of hypertension in 2006 (30%) than both Canada (19.5%) and USA (29%).[47,63] Variation in the prevalence of hypertension within the country was also documented. For instance, the prevalence of hypertension was significantly higher among blacks compared to whites in USA,[64] and among Afro-Caribbeans than Caucasians in England.[51] Furthermore, rural populations had significantly higher prevalence of hypertension than urban populations in high-income countries (HIC), middle-income countries (MIC), and low-middle income countries (LMIC), with low-income countries (LIC) the exception, where rural population had significantly lower rate of hypertension than urban populations.[65]

1.3.5 Prevalence of hypertension in Scotland

Prevalence of hypertension in Scotland was 28.4% in 2012/2013, as reported by the Scottish Health Survey (SHeS) 2013 for adults aged 16 years and over.[66] Men had marginally higher prevalence of hypertension than women (28.6% vs. 28.3%). The trend in prevalence of hypertension over the ten-year period between 2003 and 2013 showed a stable prevalence between 2003 and 2010/2011, before a statistically significant decline to 28.4% in 2012/2013 (Figure 1-5); this observation was similar for both men and women. In 2012/2013, 25% of hypertensive participants were receiving treatment for hypertension but still having high BP, and only 20% of the hypertensive participants were controlled by the treatment and maintained BP levels below 140/90 mmHg (i.e. controlled

hypertension). Moreover, more than half of the hypertensive interviewees were untreated; particularly, no participants under the age of 35 years were being treated for hypertension. Importantly, identifying people with BP >140/90 mmHg who are not on treatment may provide an estimate of the prevalence of potentially undiagnosed hypertensive. Yet, it must be considered carefully as a clinical diagnosis of hypertension is defined as “sustained raised of BP” that is not based on a single measurement.

1.3.6 Global management of hypertension

The global importance of hypertension treatment and control is now recognized, leading to frequent epidemiological studies and community surveillance to evaluate the effort of treating and controlling hypertension. These surveys mainly report the prevalence of hypertension along with three important indices- ratios of treatment, control, and awareness of hypertension. In the case of comparable measurements between surveys, an international comparison between countries and regions can help devise potential ways to improve public health strategies. Although the benefits of lowering blood pressure are robustly evident and there are several classes of antihypertensive medication available, the proportion of awareness, treatment, and control of hypertension are still low. For instance, more than half of the individuals defined as hypertensive in the Prospective Urban Rural Epidemiology study (PURE) were unaware of their conditions, and only 41% of them received antihypertensive therapy (Table 1-3).[65] The proportion of treated and controlled hypertension with target BP level of <140/90 mmHg was only 33%. Similar to hypertension prevalence, the proportion of hypertension management indices varied widely between countries and within sub-populations of the same country. For example, the proportion of awareness, treatment, and control of hypertensive individuals in England were 65.3%, 51.3%, and 53% for the year 2006, compared to 83.4%, 79.9%, and 82%, respectively, in Canada for the year 2007/2009.[47] In addition, Table 1-3 shows the variation in hypertension management indices between the different country categories from the PURE study. Similarly, the Study on Global Ageing and Adult Health examined these hypertension indices in six countries and reported lower proportion of control and awareness in LICs.[67] Consequently, there is a substantial shortfall in the effective management of hypertension worldwide, with a wide gap between different countries. Nevertheless, the proportion of

hypertension management indices have shown improvement. A cross-sectional study of five Health Survey for England (HSE), conducted between 1994 and 2011, showed an improvement of the mean BP level in treated hypertensive from 150/80 mmHg in 1994 to 135.4/73.5 mmHg in 2011.[68] There was also significant improvement in the proportion of awareness, treatment and control, the proportion of control among treated hypertensive almost doubling from 33% in 1994 to 63% in 2011.

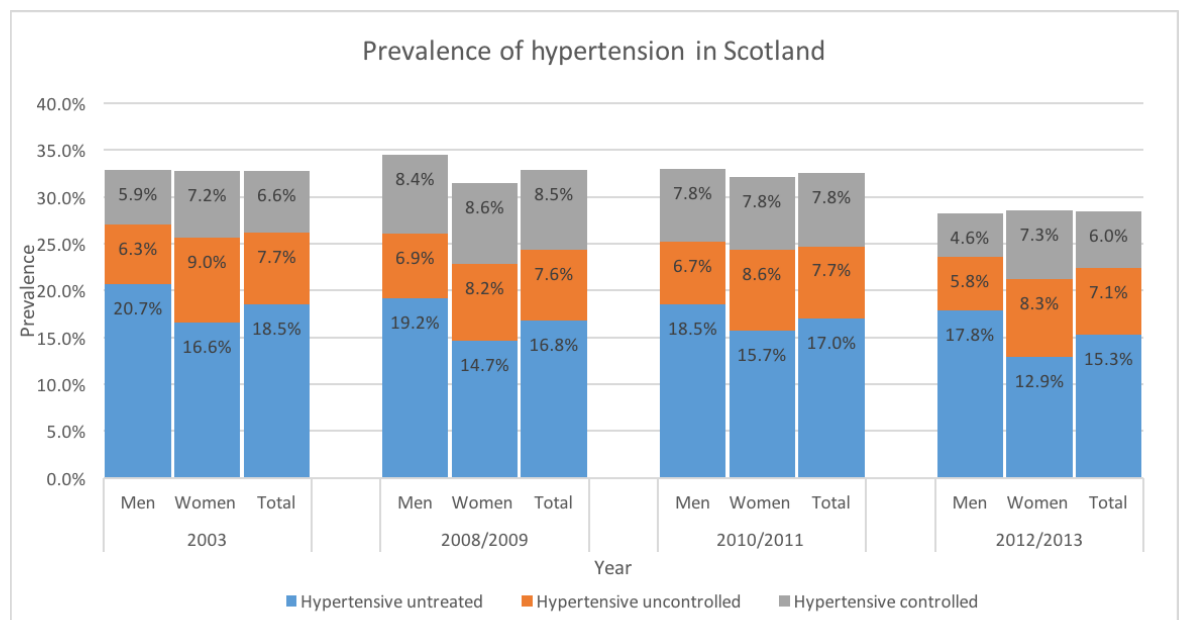


Figure 1-5 Trends in hypertension-related ratio between 2003 and 2013 in Scotland. Figure is produced based on data from the Scottish Health Survey 2013 for adults aged 16 years and over. In the 2012/2013 survey, BP measurements were taking by the interviewer, replacing the previous method of nurse-collected BP measurements. However, the prevalence stated in this graph is the equivalent to the nurse-collected BP measurements, which was reported in the survey as "nurse / nurse equivalent". ***Data available in http://www.gov.scot/Publications/2014/12/9982/20#t8.6.***

Table 1-3 Prevalence of hypertension, awareness, treatment, and control in PURE Study

Variables	Total	No. (%) of Participants			
		Prevalence of hypertension[a]	Awareness[b] among hypertension	Treated[c] among hypertension	Controlled[d] among treated
Country by income level					
HIC	15,418	6,263 (40.7)	3,070 (49.0)	2,924 (46.7)	1189 (40.7)
UMIC	36,463	18,123 (49.7)	9,516 (52.5)	8,761 (48.3)	2833 (32.3)
LMIC	58,476	23,269 (39.9)	10,134 (43.6)	8,595 (36.9)	2314 (26.9)
LIC	31,685	10,185 (32.2)	4,157(40.8)	3,230 (31.7)	1,298 (40.2)
Sex					
Women	82,607	32,649 (39.5)	16,440 (50.4)	14,491(44.4)	4,891 (33.8)
Men	59,435	25,191(42.4)	10,437 (41.4)	9,019 (35.8)	2,743 (30.4)
Overall					
All	142,042	57,840 (40.8)	26,877 (46.5)	23,510 (40.6)	7,634 (32.5)

a. hypertension is defined as self-reported hypertension, receiving treatment, or BP ≥140/90 mmHg
b. awareness is defined as answering "yes" to the question whether they had a medical diagnosis of hypertension
c. treatment is defined as answering "yes" to the question whether they receiving BP-lowering medications.
Abbreviations:
HIC, high-income country; includes (Canada, Sweden, and United Arab Emirates)
LIC, low-income country; includes (Argentina, Brazil, Chile, Malaysia, Poland, South Africa, and Turkey)
LMIC, low–middle-income country; includes (China, Colombia, and Iran)
UMIC, upper–middle-income country; includes (Bangladesh, India, Pakistan, and Zimbabwe)
Table is derived from data in Chow et al.[65]

1.3.7 The public health importance of blood pressure

High BP ranked as the leading single risk factor for global disease burden in 2010, above tobacco smoking and household air pollution from solid fuels.[1] Moreover, high BP was one of the five leading risk factors for GBD in all world regions, except the Western sub-Saharan Africa where it was ranked as number six.[1] It was estimated that hypertension affected nearly one billion people in the year of 2000, leading to more than seven million premature deaths and ninety-two million disability-adjusted life years in worldwide.[2,3] Generally, more than 80% of the attributable burden of high BP (SBP >115 mmHg) occurred in LIC and MIC regions, and over 50% occurred in people aged 45-69 years, and just over half occurred in people with mean SBP less than 145 mmHg.[2] This suggests that the relative risk of CVD events start to rise at this level of BP (>115 mmHg), and restricting prevention and treatment strategies to only people identified as hypertension by the current guidelines may miss much of BP-related disease.

High BP has been found to be positively and progressively related to the risk of stroke and coronary heart disease. For instance, the risk of CVD for people aged 40-60 years is doubled for each incremental increase of 20/10 mmHg of BP above the usual BP level of 115/75 mmHg.[5] Moreover, a stepwise increase in the rate of CVD events was observed in persons with higher BP categories.[6] Consequently, adequate management of hypertension can effectively protect against high BP-related complications, and reduction in mean BP was associated with reduced chronic heart disease (CHD) and stroke events, even for individuals without a history of cardiovascular diseases.[69,70] In addition, healthcare costs attributed to sub-optimal BP was estimated to be at least US$370 billion, which represents nearly 10% of the world healthcare expenditures, highlighting the fact that both lifestyle and pharmacological interventions are more cost-effective options.[71]

Several epidemiological studies have consistently demonstrated the benefit of BP lowering through population-wide and individual intervention (i.e. behavioural or pharmacological). A meta-analysis of 147 randomized trials of BP lowering and 464,000 participants showed that a 10 mmHg reduction in SBP is associated with 22% reduction in CHD and 41% reduction in stroke.[70] Moreover, each 10 mmHg reduction in mean population SBP was associated with 46% and

41% reduction in cardiometabolic mortality for men and women, respectively.[72] Consequently, developing a strategy to lower BP in everyone above a certain age is a superior method than measuring it in everyone and treating it in some. This can be achieved by implementing effective primary prevention strategies to reduce exposure to behavioural risk factors. Remarkably, the degree of exposure to such risk factors are influenced by the lifestyle in the society, which necessitate developing healthy public health policies and environmental interventions that make healthy lifestyle options much easier for people. An example of primary prevention program for hypertension includes two strategies that can be applied at a population-level and high-risk individual level, to maintain a BMI between 18.5–24.5 kg/m2, reduce dietary sodium intake, regular aerobic exercise, limit alcohol consumption, and adequate diet that is rich of potassium, fruit, and vegetables.[73]

Despite the overwhelming benefits of lowering BP, the rates of awareness and controlled-hypertension remain low (see Section 1.3.6 –p39). This highlights the importance of more intensive and cost-effective strategies that are applicable even to population of LIC, which may have major challenges to adopt the international guidelines due to a poor healthcare infrastructure or insufficiency of antihypertensive medications. Nevertheless, medication availability, affordability, adherence, and acceptability are important barriers that need to be considered carefully in each setting to improve the control rate of hypertension. The difficulty in attaining a high rate of controlled hypertension can be extending to inadequate response to monotherapy, and difficulties in adopting healthy lifestyle. Improving our understanding of the biological mechanisms of BP regulation with identifying new pathways for antihypertensive medications, along with early identification of people at higher risk of BP-related complication may offer hope that some of these problems will be surmounted in the future.

1.4 Principles of basic genetics

The human genome consists of deoxyribonucleic acid (DNA), which is made up from a long sequence of four nucleotide bases; adenine (A), cytosine (C), guanine (G), and thymine (T). Two strands of DNA are held together by hydrogen bonds to form a double helix, in which each two of the complementary base pairs are linked together based on Watson-Crick rules; (A) pairs with (T), and (C) pairs with (G). Each nucleotide base is linked to its neighbours in the same strand by phosphodiester bonds, that links carbon atoms number 3' and number 5' of successive sugar residues. Hence, the two single strands can be identified by their ends, which are the 3' and 5', pronounced as 3 prime and 5 prime, respectively. During replication, the DNA double strand is first unwound by the action of the helicase enzyme and each strand serves as a template for synthesizing a new complementary DNA strand by the action of different types of DNA polymerases. A single strand of DNA can also serve as a complementary strand for ribonucleic acid (RNA) during the transcription process when certain segments of coding DNA (exons) that are spaced at irregular intervals by non-protein coding regions (introns) are transcribed to make initial primary RNA, which then undergoes post-transcriptional processing to cut out the introns and splices the exons to make a mature RNA (mRNA), or messenger RNA that will in turn serve as a template to make a polypeptide in the translation process.

The double helix DNA in the human genome is divided into 46 chromosomes; 22 homologous pairs of autosomes and one pair of sex chromosomes (XY in males, and XX in females). One copy of the homologous chromosome is maternally inherited and the other is paternally inherited. Although the two homologues share the same sequence of genes in the same positions, they usually can be distinguished by the presence of sequence variations at several loci. The term "locus or loci for plural" refers to a unique chromosomal location that defines the position of an individual gene or DNA sequence; if this locus has a different version of DNA sequence in the population, then each version is called an "allele". The alleles present at a specific locus is called the genotype; for instance, for three loci, A, B, and C that lie in one chromosome and take alleles A1-B1-C1 along one homologous chromosomes and A2-B2-C1 along the other, then the genotype of this subject for the three loci is, respectively, A1A2, B1B2, and C1C1. This subject would be called as homozygous at locus "C", as it

consists of the same allele, and heterozygous at loci “A” and “B” as they contain two different alleles at the same locus. The haplotype refers to the allelic configuration along the same single DNA strand, thus in the above example it would be A1-B1-C1 and A2-B2-C1. This haplotype can be rearranged by a recombination process during meiosis; in which the pairs of the two non-sister chromatids of chromosome homologues are aligned at meiosis I, resulting in recombined chromosomes that have a portion of DNA from both mother and father. This one source of genetic variability during meiosis is then further expanded by the independent assortment of chromosomes and random fertilization.

The concept of recombination is very important for gene mapping, as the probability of an odd number of crossovers (recombination) between two genes (or loci) is highly dependent on the physical distance between them; that is, if the two segments of DNA are close together, the chance of recombination is very small, and the probability of recombination increases as the distance between them increases, up to a maximum of 50%. The physical distance in chromosome is expressed in “centimorgans”, which refers to the region of DNA within which a crossover is expected once in every 100 meiosis events. The size of this region in base pair depends on the gender and place in the genome, but a rule of thumb is that one centimorgan corresponds to about one million bases.[74] When a variant is introduced into a population by mutation, it is perfectly correlated with nearby variants and is said to be in linkage, which is the tendency of genes at specific loci to be inherited together as a consequence of their physical proximity on a single chromosome. If two particular alleles at two different loci are found together more often than expected by chance, they are said to be in linkage disequilibrium (LD), which is defined as the non-random association of alleles in adjacent loci. The new variant perfect LD with its neighbouring loci tends to break up (“LD decay”) over successive generations due to several reasons such as mutation rates, admixture, gene conversion, genetic drift, population growth, population structure, and natural selection.[75]

The LD between two markers within the same genomic region is commonly measured by the absolute value of *D’* and r^2. The concept of these measures is to calculate the difference between the observed allele frequencies, i.e. a two

marker haplotype, and the frequencies if the two alleles are independent. The higher the value of *D’*, the lower the possibility that a recombination event occurred between these two loci (*D’* =1 means that the two markers have not been separated by a recombination event). The absolute value of r^2 is more commonly used to quantify and compare LD in the context of mapping. When r^2 =1, the two markers have not been separated by recombination and have the same allele frequency. In this case of perfect LD, the two markers are completely linked and observation at one marker provide complete prediction about the other. If the alleles D and A are in complete LD, the power to detect the association between allele A and the disease requires no change in sample size. However, if the r^2 value is less than one, the sample size must be increased by $1/r^2$ of the sample size required for detecting the direct association with the disease’s allele.

The human haploid genome is very large as it consists of more than 3 billion base-pairs; of which, about 3% contains coding sequences with an estimated 21,000 distinct protein-coding genes.[74,76] While the vast majority of the genome is identical between any two unrelated individuals, there are DNA sequence variants that contribute to genetic differences within and between populations, which are responsible for human diversity in expressing traits such as eye or hair colour, and even disease susceptibility. These genetic variants vary between populations and have arisen over time as a result of several factors, such as positive natural selection, which is the increase in prevalence of a specific genetic variant that improves survival and fertility. An example is the mutation that protect against malaria by disrupting expression of the Duffy antigen gene (*DARC)*, which encodes the receptor used by the parasite to enter blood cells.[77]

The types of genetic variation within the human genome are diverse, ranging from a single base pair alteration such as in single nucleotide polymorphisms (SNP) to structural variants, which are any change that involves more than a single nucleotide such as insertion-deletion (indels), block substitutions, inversions of DNA sequences, and copy number variants (CNVs). These variants can be also classified by the frequency of the minor allele to rare and common variants; common variants or polymorphisms are defined as genetic variants with a minor allele frequency (MAF) of at least 1% in the population, and rare variants

have a MAF of less than 1%.[78] It is estimated that the human genome contains at least 11 millions SNPs, with about 7 million SNPs that have MAF over 5%, and the remaining have MAF between 1 and 5%.[78] Despite the huge number of these variants, the vast majority of SNPs with MAF of at least 5% tend to be in LD with each other in different regions (known as LD bins). Hence, a limited set of about 500,000 to 1,000,000 "tagging" SNPs from each LD bin could capture more than 90% of genetic variation in the population.[76] Most of these SNPs are catalogued in the public database SNPs and designated by a reference number.

There are four main factors that make SNPs more preferable than other type of genetic variation for dissecting the genetic basis of common diseases. First, they are the most common type of variant and can be found throughout the genome, in exons, introns, promoters, enhancers, and intragenic regions. Thus, some of these polymorphisms might themselves be functional. Second, adjacent SNPs can be used for gene mapping by examining the correlation pattern between them, which may underline recombination hot-spots. Third, SNP frequency differences between different populations can be used in population-based trans-ethnic genetic studies. Fourth, SNPs are more stable than other types of genetic variation, which allow more consistent estimates from association studies.

1.5 Fundamental of complex trait genetics and association studies

Dissecting the genetics architecture of a trait (i.e. all the genetic and environmental factors that contribute to the trait, as well as their magnitude and their interactions) requires knowledge of the epidemiology and genetics, which are combined together in the field of genetic epidemiology. Traditionally, these two fields were independent disciplines with minimal interaction. Yet, advancements in molecular genomic applications technology have enabled the integration of epidemiological methods in human genetics. While epidemiology can be defined as the study of the distribution and determinants of health related states and events in specified populations, the term genetic epidemiology is less clearly defined and has been used to describe different aspects such as familial aggregation, inherited disease in populations, genetic structure of populations, and gene-environment interaction.[79] A comprehensive

definition for genetic epidemiology is given by Morton as " a science which deals with the aetiology, distribution and control of disease in groups of relatives and with inherited causes of disease in populations".[79] These two definitions of epidemiology and genetic epidemiology emphasize that both disciplines aim to draw inferences at the level of population rather than at the level of an individual. For instance, genetic epidemiology aims to pool information across families or subset of individuals to draw inferences about potentially weak effects at the level of a population. Importantly, the definition of genetic epidemiology used the term "inherited" but not "genetic" to also considers non-genetic factors of traits clustering within families, which may includes cultural and environmental factors.[74] Hence, genetic epidemiology incorporates several analytical approaches to discover the action and transmission of genes, and the potential mechanisms of non-genetic factor contributions in familial aggregation.

An illustrative example of approaches used by genetic epidemiologists to dissect the genetic basis of a trait was provided by Burton et al., and depicted in Figure 1-6. These methods can be used to establish if the trait of interest is likely to be influenced by one or more genes. However, important considerations need to be applied to this flowchart. First, it was not meant to be a guideline about how to conduct the study but rather just an overview of the broad range of process and its complexity, Second, a detailed descriptive analysis about the population should be performed first to prioritise the subsequent methods, which may be expensive and requires collaboration with other research groups. Finally, it is not necessarily to follow these steps in sequence, for instance, the presence of supporting evidence of a specific locus can justify doing an expensive genotyping experiment without performing all the previous steps. This is especially true with the advance of genomic technology and reducing cost of genotyping. Moreover, obtaining a sample of related individuals (i.e. family) with proper phenotyping data is one of the most expensive parts of many contemporary studies, which in many cases tempt scientists to jump directly to the association analysis without investing time and resources to examine the familial distribution of the phenotype. Nevertheless, such procedures are still important to refine the genetic model and risk estimates.

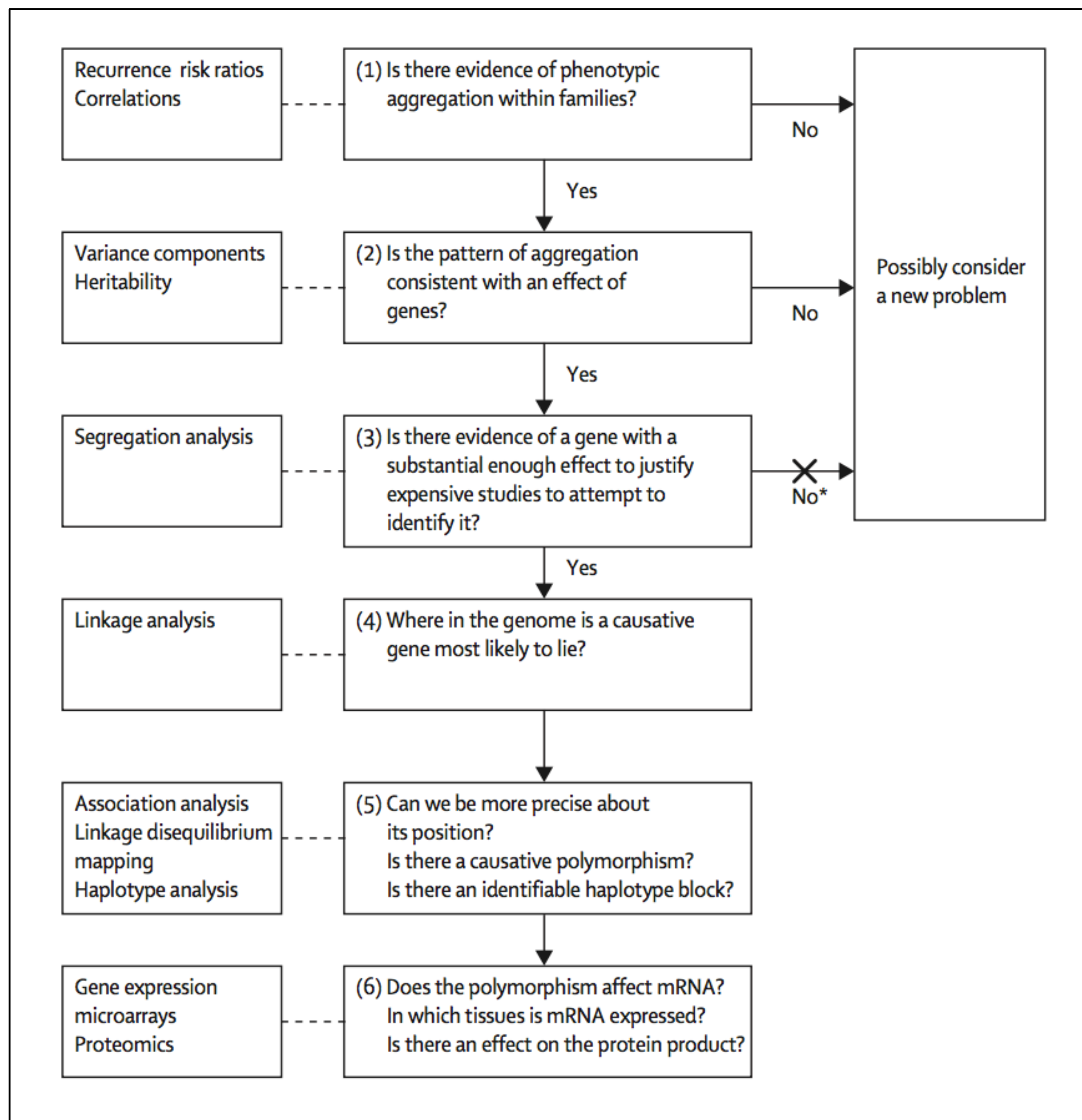


Figure 1-6 Flowchart of a systematic approach to dissect the genetic basis of complex disease.
*** Absence of significant results from segregation analysis does not exclude the possibility of genetic disease. Figure is reproduced with permission from [74].**

1.5.1 Definition of phenotype

The term "phenotype" refers to a measurable and observable characteristics of individuals;[74] this term will be used interchangeably with the term "trait" throughout this thesis. This definition generally covers two types of traits – discrete or binary, and continuous. Discrete or simple binary traits are those where only two phenotypes are observed, such as hypertension or not (yes/no). Continuous or quantitative traits are those where the phenotype has a range of continuous values such as SBP. Choosing to study the phenotype quantitatively or qualitatively is usually motivated by the balance of maximizing the statistical power and reducing heterogeneity, i.e. showing similar phenotypic characteristics that are due to a different genetic background. Examining complex phenotypes quantitatively can greatly enhance the statistical power, help to uncover biological pathways from gene to disorder, explore the genetic overlap among quantitative traits that contribute to the complex disease, and chart the relative contributions of particular genetic variants to build up aetiological profiles that may be influenced by different factors such as age or environment.[80] For instance, defining someone as hypertensive based on arbitrary BP values is of clinical importance but it is generally unclear if this definition is genetically relevant. Essentially, what is considered as a distinct disease on the clinical scale may be a heterogeneous group of potentially overlapping diseases on a genetic aetiology scale, leading to both challenges and opportunities in discovering genes for such heterogeneous disorders. For instance, phenotypic variation due to locus heterogeneity (i.e. different genes contribute to one phenotype) can affects all the methods of gene localizations; nonetheless, allelic heterogeneity (i.e. different alleles in the same locus cause a similar phenotype) can also affect the association methods. In both cases, statistical power is reduced by phenotypic complexity due to heterogeneity.[81]

The complexity of the clinical definition may be minimized by using intermediate phenotypes or "endophenotypes", which are phenotypes correlated with disease that might be closer to gene action than the overall disease definition. The term "endophenotype" was coined in 1966 to distinguish between exophenotype (external) and endophenotype (internal). In general, an endophenotype must meet the following criteria: (1) it is linked to an illness in a population; (2) it is heritable; (3) it manifests itself whether or not the illness is

active; and (4) it co-segregates with an illness in families.[75] Separate analysis of the underlying quantitative traits (i.e. SBP, DBP, MAP, and PP) instead of an overall clinical phenotype (i.e. hypertension) can be a more powerful approach. The qualitative approach may be relevant in certain instances, especially to reduce the heterogeneity by studying extreme phenotype. For example, Padmanabhan et al. reported a novel locus in the promoter region of the Uromodulin gene (*UMOD)* by using cases and controls drawn from the extremes of the BP distribution.[82] This approach can increase the statistical power of the study by achieving the maximum possible phenotypic separation between the cases and controls.

1.5.1.1 Phenotype error

The accuracy of phenotype is important for the outcome of genetic studies; that is, "even the most precise molecular-genetic data cannot be useful if the phenotypes are not well defined".[83] In general, reduction of phenotype accuracy can lead to lower statistical power and increases the likelihood of a type II error, or failure to identify true genetic signals, resulting in findings that are less likely to be replicated in future studies. However, it is possible to compensate for the attenuation of estimates towards the null from random measurement error by increasing the sample sizes using standard phenotypes. It is important recognise that large sample sizes are obtained usually by combining data from different studies. Thus adequate phenotyping performed similarly across large samples is a valid strategy as sample size factor mitigates the need for expensive hi-fidelity phenotyping strategies to resolve a phenotype that is reasonably captured by standard methods more economically.

For case-control GWAS, several studies have showed that an increase in diagnostic or classification errors are associated with a substantial reduction in statistical power, particularly for lower allele frequencies and genotype relative risks.[84-86] For quantitative traits, however, the impact of measurement errors on the statistical power is much smaller compared to misclassification errors in case-control studies.[85] This was shown in a study that examined two traits (nuclear cataract and BP) using different measurements methods.[85] In this study, two phenotypes were used for testing the effect of measurement errors for SBP and DBP in a Chinese cohort- the first phenotype derived from the average of

the two closest readings, which was assumed to be the “true” values, and the second phenotype derived from the third reading only. For DBP, the same significant signal was replicated in both phenotypes (i.e. true phenotype and its error counterpart); for SBP, however, three significant signals were reported for each phenotype but only two signals were in common. For the cataract trait, no overlapping signals were identified between the two measurement methods, suggesting that the impact of phenotypic errors on statistical power is much smaller for quantitative traits than binary traits (i.e. misclassification errors), and that small phenotypic measurement errors in BP phenotype is likely to have only a little impact on genetic association signals.[85]

The issue of phenotype accuracy can also be exacerbated in the presence of other confounders such as treatment effect, where other adjustment need to be performed to account for the medication effect.[87,88] The effect of treatment on the genetic association and epidemiological studies was demonstrated by Tobin et al. who showed by simulation and real sample data that inappropriate adjustment for BP-lowering medication can lead to substantial shrinkage in the estimated effect of the examined parameters and a marked reduction in statistical power.[87] A proposed method to offset this effect was by adding a fixed value to the observed BP values, such as 15 mmHg and 10 mmHg for the observed SBP and DBP, respectively. Later, an extended method was developed to apply a step-wise adjustment based on the class and number of medication along with the subject’s ethnicity.[88]

1.5.2 Family- or population-based sample

Family-based studies were the *sine que non* in the field of genetics. Following the Human Genome Project, family-based linkage studies supported by accurate genome map information, greatly expanded the discovery of genes for monogenic diseases. However, linkage analysis was less fortunate in the setting of complex diseases, such as hypertension. Advances in technology along with a reduction in the cost of genotyping has shifted the focus of gene mapping from family based linkage analysis to association studies using sample of unrelated individuals. This is currently performed on the scale of genome-wide association studies (GWAS) by genotyping a huge number of SNPs in large sample size of unrelated individuals. The association studies are believed to have greater

statistical power compared to linkage analysis, especially for disease variants with weak effect.[89] However, variants identified by GWAS explain little of the heritability of common diseases, which has led to a renewed interest in linkage and other family-based methods.[89-92] In particular, the current GWAS approaches are less effective in detecting rare risk variants that are in LD with common SNPs, but such variants can be detected by family-based studies (the comparison between linkage and association approaches are discussed in more details in the next section). Essentially, the availability of exome and whole-genome sequence data has reinforced the importance of linkage analysis in detecting rare variants involved in complex disease using family-based data.

Family-based design offers some advantages over population-based designs, and can be performed using several possible designs, which range from simple trios (two parents and one affected child) to a large extended pedigree with multiple generations. The main advantage of family-based design is its robustness to the effect of population stratification, a phenomenon that can lead to discovery of spurious genetic signals due to differences of genotype frequency between cases and controls that are only present because of differences in the ethnic background. The issue of population stratification is acknowledged in population-based studies and there are several methods to account for it, either during the study design stage or statistical analysis stage, for instance by computing genomic control (λ).[93] A family-based design can test within-family and between-family information; while the former is protected against the stratification issue, the later is still susceptible to stratification (more details about method to test for association in family design studies are provided in Section 1.5.7 – p67).[93] The family-based design can also offer the opportunity to test for the effect of imprinted or parent-of-origin effect, which occurs when the expressed phenotype is dependent on whether the risk allele was inherited from the father or mother. The importance of parent-of-origin effect is usually ignored because of the unavailability of this information. However, its importance as a potential contributor to complex disease is highlighted in one study from Iceland which reported five variants with parental-origin-specific association; one allele of the variant rs2334499 is protective against type 2 diabetes if it was inherited from the mother and confers risk when inherited from the father.[94] Furthermore, a survey of 97 complex traits measured in

outbred mice showed that 91 of 97 (93%) of these traits have measureable parent-of-origin effects [95] The family-based design also offers the possibility to examine whether a particular variant was inherited or has arisen as a *de novo* mutation. This has greater importance for examining CNVs where *de novo* mutation appear to occur with greater frequency in individuals with a neurodevelopmental disorders than in individuals without such disorders.[96,97]

Family-based designs can offer the ability to perform more quality control (QC) by detecting genotyping errors in the form of Mendelian inconsistencies, by checking that genotypes between relatives such as parent-offspring follow Mendelian rules. The proportion of genotypic error that can be detected in trios family ranges between 25% to 30%, which can be increased by adding more relatives that would highlight unlikely genotypes within the family.[98] Another advantage of family-based design is the possibility of imputing missing genotypes in relatives, and include the imputed genotypes scores and phenotypes in the total association test; this would increase the power of the study when genotyping resources are limited or when obtaining DNA sample for some relatives is not possible.[99]

The possibility to carry out a combined test for linkage and association is only applicable in a family-based design where linkage analysis tests the presence of linkage between a genomic locus and the disease, followed by fine-mapping this linked region through association test by using several genetic markers that span this genomic region. These markers will be evaluated if they account for the linkage signal to determine if they can explain all of the genetic contribution that can be detected from this region. The simultaneous evaluation of linkage and association by using data from pedigrees with different relationship structures is likely to be the most powerful and useful approach to detect genetic variants that cannot be detected using traditional GWAS.[89,100] However, it requires complex computation and statistical methods to account for family relationship and avoid false discovery, as the power to detect association is maximized with association test that incorporates data on related individuals.[89,101]

The family-based design does however suffer from some disadvantages, similar to any other study design. The recruitment of individuals (i.e. probands) and

their relatives is more difficult than recruitment of unrelated individuals, and requires more resources in terms of time and cost. Furthermore, the family-based design has less power per genotype than population-based study, especially in the methods that test the within-family effect, which has been attributed to the fact that these markers were used to test for the association and guard against population stratification.[99] This reduction in power can be gained by considering the total association effect rather than within-family effect only.[97] Furthermore, Visscher et al. have argued that the advantages of using related individuals in GWAS for quantitative traits may well outweigh the small disadvantage in terms of statistical power, by providing a more robust and flexible strategy for analysis.[100] The analysis of family-based association studies requires special software to account for relatedness and usually cannot be performed using classical statistical tools.

1.5.3 Descriptive genetic epidemiology

Familial aggregation of disease is an important factor that has its own clinical and epidemiological meaning, and these two meanings need to be carefully distinguished. Whilst the clinical meaning relies on the fact that extended families tend to have multiple individuals with the disease, the epidemiological meaning relies on the fact that relatives of affected individuals tend to have greater frequency of the disease than the general population. Furthermore, simple analyses of familial aggregation aims to find if the disease clusters within this cluster unit, and no attempt is made to attribute this aggregation to a specific causes (i.e. genes or environment). However, as most non-genetic factors associated with complex traits are weakly correlated in relatives and have only modest effects, the magnitude of familial aggregation can be considered to have a predominantly genetic basis.[75,102] Different metrics are used to quantify familial aggregation and heritability of the traits, which will be discussed separately in the next sections.

1.5.3.1 Familial aggregation

For a binary phenotype (disease/health), one of the simplest metrics of familial aggregation is the recurrence risk ratio in relatives of affected individuals.[103] This risk ratio is defined as the recurrence risk ratio in relatives (λ_R), which is

the prevalence of the disease in relatives, compared with disease prevalence in the general population. Λ_R can be estimated across different types of relatives, for example λ_S for siblings, which is the most common form. Measuring λ_R provides valuable information about the mode of the transmission of the disease, and is an important determinant of the power of affected relative pair studies to detect linkage. Three points need to be considered about the value of $\lambda_{S;}$ first, it only reflects aggregation between siblings without explaining the source (i.e. either genetic or environmental). Second, the value of λ_S can be distorted by the presence of systematic differences of ages between the study sample population (i.e. siblings) and the general population sample from which the disease prevalence is calculated, which is a particular concern for diseases of late onset. Third, the value of λ_S is specific to the studied population and it should be carefully interpreted in different populations, especially for diseases with prevalence that may differ across populations.[79]

Familial aggregation for continuous traits can be assessed using a covariance-based method, for instance the intra-family correlation (i.e. intraclass correlation coefficient [ICC]) is one way to describe the correlation of quantitative trait for units that are clustered on groups (families). The ICC statistic indicates the proportion of the total variability in the trait that can be attributed to real variability between families. For a predominantly genetic trait, the within-family variability is expected to be small, resulting in higher estimates of between-family variability and higher ICC. Hence, non-random ascertainment can seriously bias the ICC estimate leading to inflated estimates of familial aggregation, as the full range of the trait values in the total population is not well-represented across families.[104]

1.5.3.2 Heritability

The heritability of a continuous trait is defined as the proportion of its total phenotypic variance (σ^2_P) attributable to genetic factors in a specific population. Narrow sense heritability (h^2) is defined as σ^2_A/σ^2_P, and broad sense heritability (H^2) is defined as σ^2_G/σ^2_P, where σ^2_A includes the additive genetic components of variance and σ^2_G includes all genetic components of variance.[105] For binary traits, heritability is usually calculated by invoking a hypothetical construct, known as liability, and applying a version of variance components

modelling. Liability is an underlying, unobservable, normally distributed trait that is assumed to determine the probability that an individual will develop the disease of interest. The main point to note here is that the numerator has a simple genetic meaning, but the denominator σ^2_P captures the variance attributable to genes and shared environment as well as residual variance attributable to unshared and unmeasured determinants and measurement error.

For a given trait, heritability can vary quite substantially from study to study depending on the population being investigated as a result of varying environmental exposures between populations, the structure of the analytic model, and measurement error. If a disease process is entirely dependent on the presence of a particular allele of a particular gene, but all individuals in the population are homozygous for that allele, variation at that locus does not exist and hence plays no role in variation in disease phenotype, therefore making no contribution to heritability. On the other hand, the gene is clearly implicated in the causal architecture of the disease. Equivalently, a near ubiquitous environmental exposure makes little or no contribution to the denominator σ^2_P. The power of most studies for discovering genes is positively associated with the heritability of the trait of interest. Thus, all else being equal and if the option exists, analytic efficiency may be enhanced by selecting a study population in which heritability of the trait of interest is thought to be high. Whilst the proportion of variance explained by genetic factors can provide insight into the value of genetic studies, it is important to note that heritability estimates do not provide information on the direction of effect of these factors nor insight into their utility in prediction.

To estimate the magnitude of familial aggregation of a phenotype that is due to genes requires statistical modelling of how the phenotype is modulated by the effect of one or more genes. One of the most common methods is the additive genetic effects model, which adds or subtracts a constant value from the expected value of a trait for each copy of an allele at a locus.[80] The amount added or subtracted varies in an unknown way from allele to allele and from locus to locus. The additive model assumes that the effect of any one allele is independent of other alleles, and uses probabilities of alleles that are shared identical by descent (IBD) as a measure of allele sharing among different classes

of relatives. Under these assumptions, the additive model captures much of the genetic effect on the phenotype, and this can be quantified by estimating the variance components in families or pedigree using mixed linear models.[105] Crucially, variance components analyses require no information about genotypes or measured environmental determinants. This approach can be extended to include the covariance or correlation patterns (or both) that would be expected for other more complex models of genetic determination (for example, genetic dominance) due to unmeasured environmental determinants that are shared by a whole family, those that are shared just by siblings, and those which wax and wane as individuals spend time living together or living apart. Finally, many environmental and lifestyle exposures are unique to an individual. These unshared determinants contribute nothing to the tendency for relatives to be more similar than non-relatives (i.e. they do not contribute to the covariance between relatives), but they do affect the total variability of a quantitative trait. Other genetic and non-genetic models might also be consistent with the data, so a good fit of any one model does not prove that that model is right.[80,105]

1.5.4 Linkage analysis

The purpose of linkage analysis is to search for alleles or chromosomal segments shared by affected relatives that are more than expected by random Mendelian segregation. These segments are passed entirely from the parents to the offspring without recombination at meiosis (Figure 1-7). The number of crossover sites in the human genome is thought to follow a Poisson distribution (averaging around 35 crossover points), and their locations are generally random and independent of each other. Co-segregation should therefore be detectable for marker loci quite far away from the disease-causing variant. Because linkage operates over long genetic distances, a positional mapping approach based on linkage can cover the entire genome by using a relatively small number of highly polymorphic markers. Standard marker sets for whole-genome linkage scans, based on 200–800 microsatellite polymorphisms, which became available in the 1990s, enabled the successful mapping of hundreds of rare single-gene disorders. Linkage analysis is carried out only in families with affected relatives and involves genotyping of several markers that spread over the entire genome. Markers that flank the disease gene or mutation tend to highly segregate with disease status in families. Identifying markers within such a segment that

consistently accompany the disease may indicate the presence of susceptibility genetic factors near them. The power to detect small effect size increases with decreasing relatedness of subjects (for example by recruiting extended family), however, higher numbers of markers are required as the likelihood of recombination also increases (i.e. shorter segments are shared).

Linkage studies were the predominant method for gene mapping and were successful in uncovering genes for monogenic Mendelian disease. However, linkage analysis has achieved only limited success for most of the common complex diseases. This has made the field of common complex disease to shift to apply genetic association methods that analyse common variants, which have a modest effect. For such variants, association methods are more powerful than linkage analysis.

1.5.5 Association studies

The goal of population association studies is to identify patterns of polymorphisms that vary systematically between individuals with different disease states and can therefore represent the effects of risk-enhancing or protective alleles. This implies that traits are still linked to the surrounding genetic sequence of the original evolutionary ancestor through linkage disequilibrium, or that they are found more often in a given haplotype than outside of it. Association mapping is based on the idea that genetic variants underlying complex traits occur with a relatively high frequency (>1%), have undergone little or no selection in earlier populations, and are likely to date back >100,000 years (the common disease/common variant hypothesis). Association analysis potentially has far greater power than linkage analysis for detecting variants with a modest effect on disease risk, provided that the genetic marker is close enough to exhibit strong LD with the functional variant (Figure 1-8).

A direct association study tests the association between a known functional variant and disease. Indirect association studies are more commonly performed and rely on the principle of LD, by testing the association between the disease and a marker locus that lies close enough for the disease locus to be in LD with it; they may be enhanced by examining multiple markers simultaneously using

haplotypes. The association study can be performed by two main approaches; the first is the candidate-gene approach where it examines the association between a trait of interest and selected markers, based on an *a priori* hypothesis about their relationship to the trait. The second is a genome-wide approach, where a survey of most of the genome is performed to find the associated genetic variants without previous hypothesis or assumptions about the location of the causal variant.

Finding an association between a marker and a trait does not necessarily indicate a genetic causation. Other factors related to the studied population can also result in an association signal. In general, an association between a genetic marker and a trait might be a result of (a) direct causation, where this marker does affect the trait status, or this marker is in LD with an ancestral chromosomal segment that carries a functional variant, or (b) due to an epistatic effect, where the trait status is more likely to be affected in the presence of specific genotypes of this marker along with other markers, or (c) due to chance or other factors related to the study design and the studied population such as population stratification or type 1 error. Confirming that an association is a direct association is a challenging process that may include re-sequencing the target region, dense genotyping of all the available markers, or functional studies to confirm the role of a putative mutation in the disease pathophysiology.

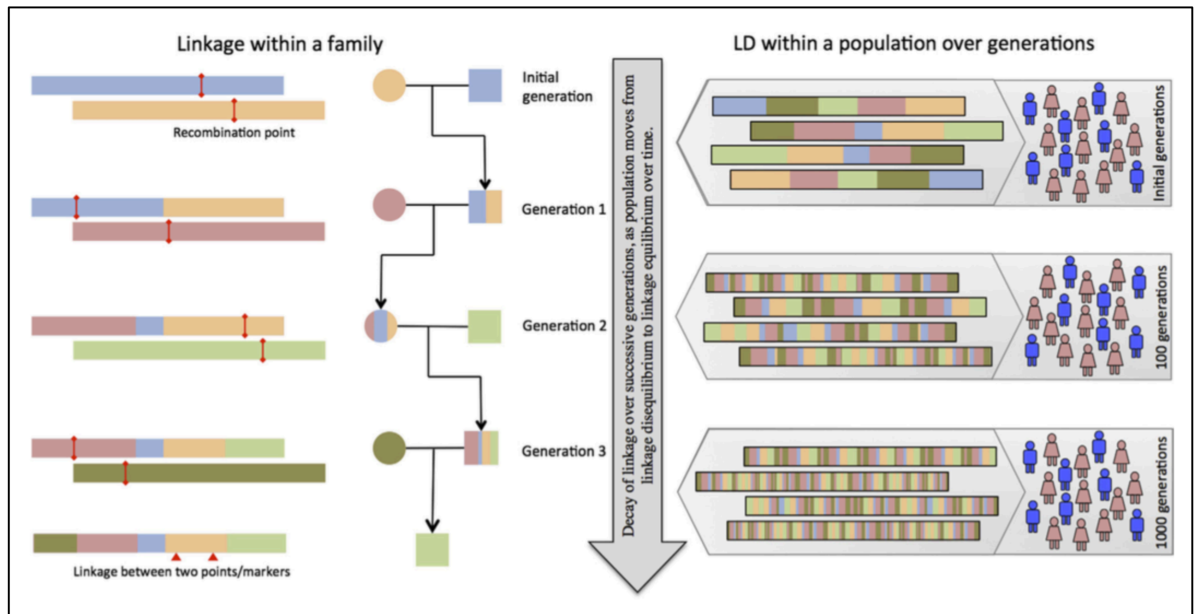
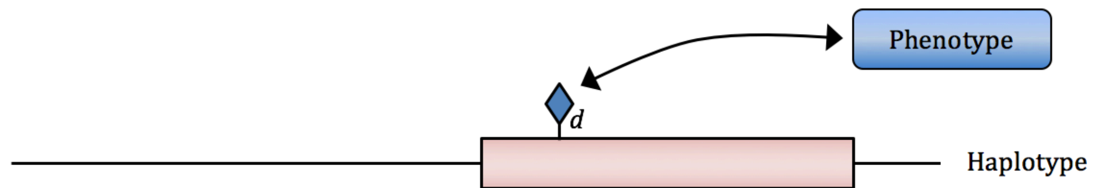


Figure 1-7 Linkage and linkage disequilibrium.
Within a family, a recombination occurs during meiosis in the germ cell between the paternal chromosome (blue) and the maternal chromosome (yellow). The offspring carry a recombinant chromosome, which undergoes further recombination in the next generation. In generation 3, the two markers (red triangles) in the ancestral (yellow) region are said to be in linkage. Within a population, the stretches from the ancestral chromosomes are decayed by recombination events over time. At some point, a recombination event occurs between every possible point in the chromosome, which leads the population to move from LD to linkage equilibrium. *Reproduced with permission from* [75]*.*

A. Direct test the association of the functional variant



B. Indirect test the association with a dense SNP map in LD with the functional variant

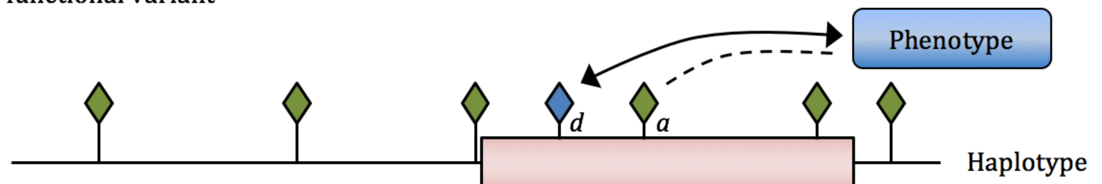


Figure 1-8 Direct and indirect association.
LD is the key factor in an association study because it induces correlation in short regions of the genome and underlines the susceptibility factors that have been inherited from ancient common ancestors. The *d* allele is associated with the disease. The *a* allele, on the other hand, is always associated with the *d* allele due to close physical proximity. Thus, an association with the disease can be directly found by genotyping the disease allele *d* (A), or indirectly found by genotyping the *a* allele that is in LD with the disease allele (B). Finding the direct association always results in higher association power, but indirect association should result in a significant association when the r^2 of alleles *a* and *d* is high and an adequate sample size has been used. Reproduced with permission from [75].

1.5.6 Genome-wide association study (GWAS)

GWAS offer a hypothesis-free approach that systematically tests hundreds of thousands or more variants in the genome without prior knowledge of the location of the causal variants. Essentially, GWAS are dependent on three critical factors; (1) sufficiently large sample size drawn from a population of appropriate genetic background, (2) efficient genotyping panel that adequately covers the whole genome, and (3) powerful statistical methods that can reveal genuine association signals. The rapid developments in these three areas over the last ten years has led to more than 2000 GWAS, reporting more than 15,000 SNPs as of February 20, 2015. Regarding the first factor, several consortia and collaborations were established combining samples from multiple cohorts to conduct GWAS meta-analyses, such as the International Consortium for Blood Pressure (ICBP).[107] These consortia have resulted in discovery and replication cohorts of large size that are sufficient to detect variants of moderate to small effect sizes; for instance, a reasonable statistical power to detect a variant with effect size typically observed for complex traits such as BP can only be achieved by recruiting a large sample size of at least 10,000 or preferably 100,000 or more.[108] As expected, a common finding from all GWAS is the very modest effects on risk of the trait, typically with odds ratio (OR) ranging from 1.1 to 1.3 and thus explaining a very small amount of the examined trait's heritability. For instance, all the reported SNPs for association with BP have a very modest effect on BP, and the collective effect of all loci identified through GWAS explain only 2% of BP heritability,[38,107] compared to an estimated heritability from family studies that varies from 30% to 50%.[109-111] Regarding the second factor, cataloguing SNPs in projects such as the international HapMap Project and more recently the 1000 Genome Project have identified and mapped a substantial number of SNPs that are easily and publically accessed. These were used by companies to develop several types of commercial genotyping "chips" that can assay up to 5 million SNPs with coverage down to 1% MAF. For the third factor, the large number of statistical tests performed in GWAS increases the chance of type I error, thus the high number of false-positive results is addressed by QC procedures, stringent multiple testing correction and seeking evidence from multiple replication and validation studies of the top signals. For instance, a genome-wide significant p-value for GWAS that tested one million markers

would be set at $5x10^{-8}$ or lower (Bonferroni-corrected alpha of 0.05). Furthermore, rigorous QC steps are always performed before the association test to minimize the possibility of errors.

The important insight from GWAS of complex traits can be summarised in two points: (1) complex traits are more likely influenced by very large number of genetic variants, at least in the hundreds if not thousands. For example, the ICBP GWAS used a multi-stage design of almost 200,000 individuals, and discovered 16 novel loci for association with BP, and estimated that there are potentially 116 (95% confidence interval 57–174) independent BP variants with similar effect size to be discovered.[107] (2) The statistical power of GWAS to detect a genetic variants depends on its effect size and MAF. Thus, it is possible that many variants may have been missed because of either a small effect size (OR <1.2), or low MAF (<5%). It is now possible to have a genotyping array that simultaneously genotypes up to 5 million SNPs including variants with MAF down to 1% such as the HumanOmni5Exom array by Illumina. (3) Other structural variants such as CNVs may play a role in disease aetiology[112], but accurate calling of such structural variants remains problematic in standard GWAS.[113] (4) Several variants were reported in association with multiple traits, highlighting the presence of pleiotropy in complex traits.[114] For instance, a non-synonymous SNP rs3184594 in *SH2B3*, which introduces the amino acid substitution W262R in a plekstrin homology domain on exon 3, has showed significant association with chronic kidney disease, celiac disease, type 1 diabetes, coronary artery disease, cholesterol, haemoglobin, retinal vascular calibre, plasma eosinophil count and rheumatoid arthritis, SBP, and DBP.[37,115] (5) Most trait-associated SNPs lie within non-coding sequence, of which 43% were intergenic region and 45% were intronic regions, suggesting a greater than anticipated role for non-coding SNPs in complex traits.[116] Several studies have consistently reported patterns of enrichment among genic regions, where GWAS SNPs were enriched in function-rich regions and depleted in function-poor regions.[117-119]

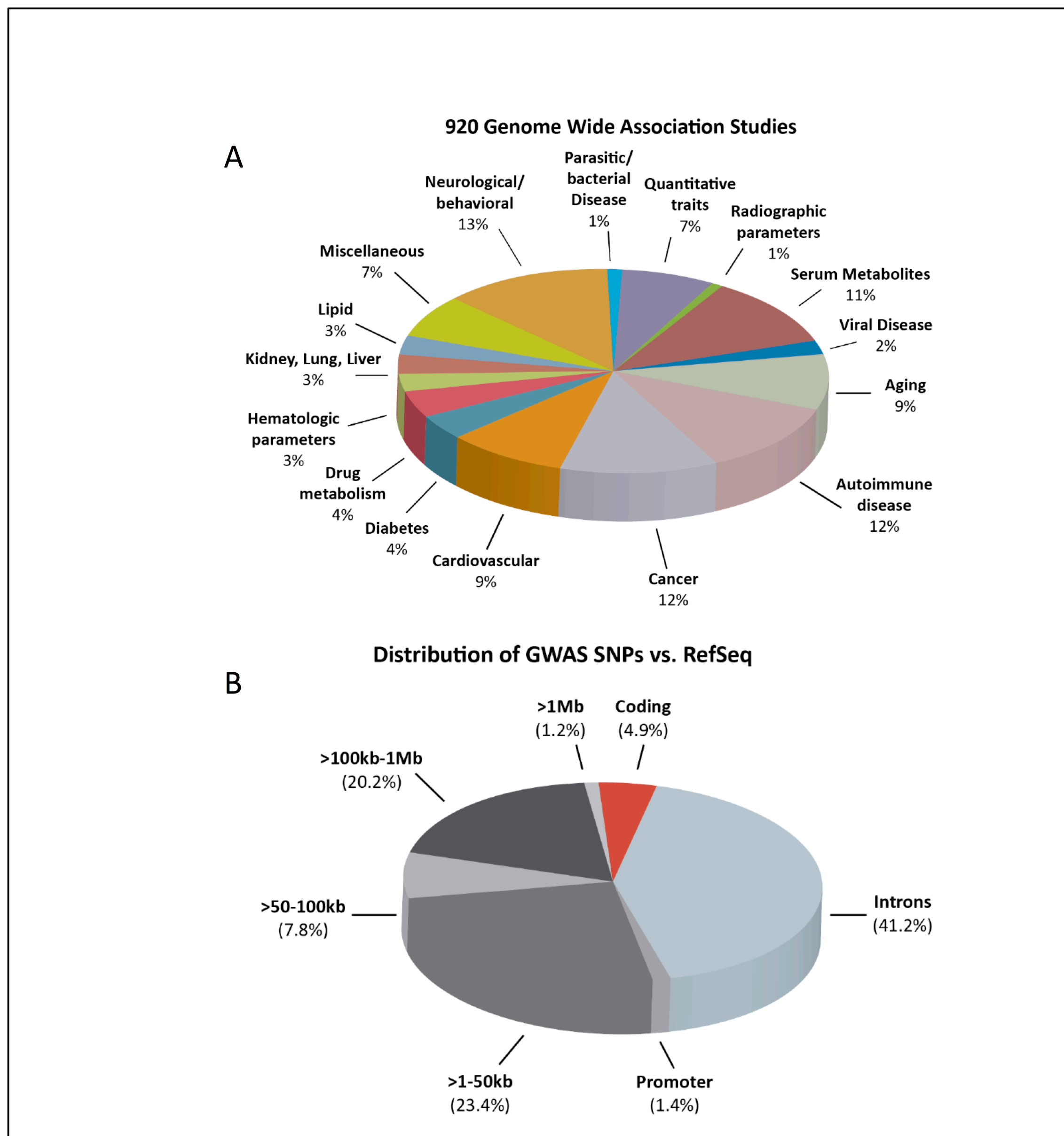


Figure 1-9 Traits examined by GWAS and functional distribution of GWAS variants. (A) A pie chart the shows the percentage of GWAS SNPs by disease/trait class, including a total of 6,011 trait-SNP associations (5,386 SNPs) from 920 different studies as of January 4, 2012. (B) Location of GWAS SNPs relative to genic features, showing that only 4.9% of GWAS SNPs lie in coding region sequence. ***Figure is reproduced with permission from*** **[119].**

1.5.6.1 GWAS quality control

QC is an essential procedure for GWAS to reduce the possibility of systematic bias. This is carried out by performing several QC procedures to exclude individuals or markers with particularly high error rates. Although it is possible to start by performing QC for markers to exclude ones with poor quality, it is generally advised to start by excluding individuals with poor quality data before conducting QC on markers.[120,121] This is because the association signal may be lost if it was tagged by one of the removed SNPs. Hence, performing individual QC first prevents markers from being erroneously removed because of poorly genotyped samples. Essentially, the question of which procedure to start with is of relative importance and based on study design, sample size, and SNP panel. Several procedures and protocol for performing GWAS QC are already available [120-123], such as the Nature protocol for case-control GWAS.[120] However, a brief listing of the QC method is explained in the next paragraph to emphasize the stringency of the procedure.

QC for individuals starts by (1) checking sex inconsistencies to identify issues that typically result from sample handling errors; (2) checking sample relatedness by comparing the estimated kinship coefficient based on the pedigree to the one calculated based on the markers, this would identify possible non-paternity, adoption, sample mix-up, or duplicate process of a single individual.[122] For a population-based GWAS, individuals with greater relatedness would be excluded to ensure that sample includes only unrelated individuals; (3) checking population stratification to ensure that study sample is drawn from a relatively homogenous population, for instance this can be done by principal component analysis; (4) checking the sample genotyping call rate that may be indicative of individuals with a poor quality DNA sample.

QC for markers starts by (1) checking the marker genotyping efficiency or call rate to exclude marker with poor quality, as marker assays that fail on a large number of individuals are poor assays, and are likely to result in spurious results; (2) checking Mendelian inconsistencies to identify genotyping error if pedigree information is correct; (3) filtering SNP based on MAF because statistical power is extremely low for rare SNPs; (4) checking for Hardy-Weinberg Equilibrium

(HWE) as departure from HWE can be indicated of potential genotyping errors, population stratification, or even actual association signal.[122]

1.5.7 Power considerations

Several factors influence the power of the study and many of them are beyond the control of the investigators, such as the degree of complexity of the trait or phenotype, the allele frequency and its effect size, and the genetic and historical characteristics of the study populations. However, other factors that can be addressed by investigators include selection of a homogeneous study sample, obtaining the maximum statistically powered sample size, using appropriate methods to measure the phenotype and genotypes, and applying correct methods to ensure high data QC, and performing the right statistical analyses to increase the statistical power within the constraints of available resources. For instance, the statistical power is influenced by the hypothesis for which the test is designed; for a variant with additive effects, a test that assumes additive effects would have greater power than a test that also allows dominance, and the opposite is also true as the statistical power is lost by carrying out an analysis that assumes additivity for a variant with a recessive effect.[124]

The power of a study is generally a function of the sample size (n), the effect size of the gene or locus, and the significance level threshold (α). Typically, genetic association studies are performed under the null hypothesis (H_0) that the effect size of the genetic marker is zero, and the alternative hypothesis (H_1) is that the genetic marker has non-zero effect size. The observed results are considered significant if the p-value is lower than the p-value of the significance threshold (α =0.05). However, due to the large size of the dataset and analysing very large number of markers it is necessary to choose a more realistic significance threshold. For instance, keeping the value of α at 0.05 in a study that test 1,000,000 markers would reveal about 50,000 significant markers (i.e. 5% of the tested markers) just by chance when the H_0 is in fact true. Hence, more stringent significance thresholds (α) are usually considered for GWAS to allow for multiple testing penalties. For GWAS of a European population, a genome-wide significance threshold of $5x10^{-08}$ has been widely adopted regardless of the actual SNP density of the study.[124] An alternative approach has

also been widely used to adjust for multiple testing in studies that use custom SNP arrays, or candidate-gene studies where the traditional Bonferroni correction has been used to define the significant threshold (α), by dividing 0.05 by the number of tests.[124]

Genetic studies can be performed using two approaches; either by recruiting related participants as in the family-based studies, or using unrelated individuals as in the population-based studies. These two approaches can be considered as complementary as each has its own strengths and limitations. Furthermore, the two approaches can be performed using several methods and generally the differences between them in terms of statistical power is small, when using the necessary sample size. For instance, a case-control study of 200 cases and 200 controls were slighlt powerful than a study of 200 trios (of an affected offspring plus parents) in the context of common disease with a prevalence of 14%.[90] However, family-based studies require more resources and are time consuming, leading to greater popularity of population-based study designs.

One of the best advantages of the family-based studies is their robustness to the population stratification issues, provided that the performed test was to assess within-family information. This can be applied using the transmission disequilibrium test (TDT), as both of the transmitted and untransmitted alleles came from the same ancestral source. These tests can be performed using family-bases association test (FBAT) or quantitative transmission disequilibrium test (QTDT). However, these kinds of tests ignore the between-family information, which can increase the power of gained information. Performing between-family association test in a family-based study make them similar to the population-based studies in terms of importance to test for population stratification issues. Hence, several methods were developed to incorporate the linear-mixed models (LMMs) in the genetic association studies that involve multiple sub-populations or family data (Figure 1-10).[93,125] Basically, the concept of LMM is to model the trait using a mixture of fixed effects and random-effects in which fixed effects include the genotypic markers and any covariates such as age or sex, and random-effects include the phenotypic covariance matrix which is based on kinship matrix. Originally, the kinship matrix was proposed to be constructed using the information from pedigree.[126] However, more advanced

methods were developed to be used with unrelated individuals using genome-wide genotypic data rather than being fixed at their known theoretical values.[127]

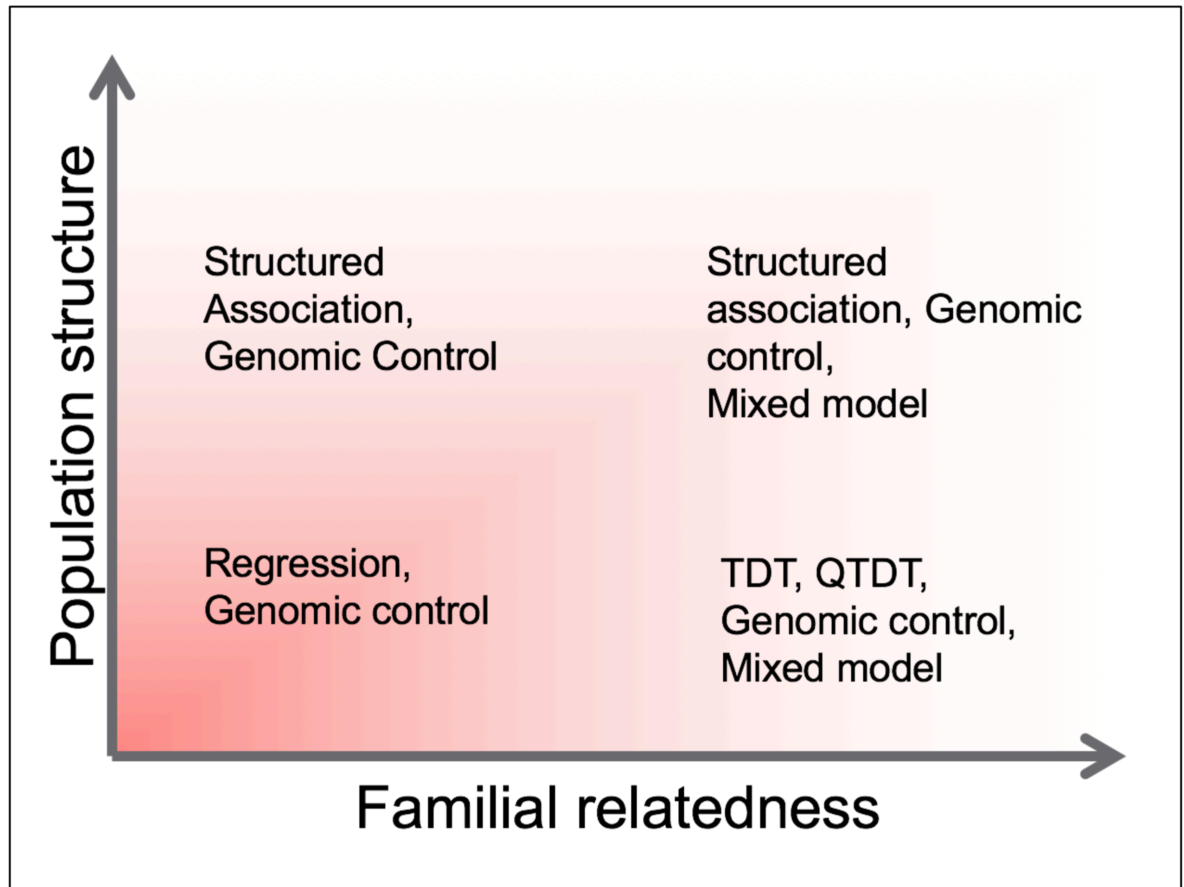


Figure 1-10 Statistical association methods for different type of samples.
The figure shows the suggested statistical methods to be used in genetic association studies to control for population structure and relatedness. The y-axis represents the population structures among the sample that may include major sub-populations. The x-axis represents the familial relatedness that may reflect a relationship among individuals from recent ancestry. In the first part (left-bottom), the sample contains minimal familial relatedness and population structures resulting in greatest statistical power, provided that the trait is well distributed. The second area (right-bottom) represents family-based sample that can be assessed using QTDT. Studies of larger sample size can include a mixture of these two samples to include population structure as in the right-top area, or include familial relationship as in the top-left area. Figure is reproduced with modification from Yu et al.[125]

1.6 Genetics of BP and hypertension

The importance of genetics factors in the aetiology of hypertension has been established in several ways. Firstly, the presence of monogenic forms of hypertension has proved that gene mutations that affect gene function can influence BP, and lead to either hypertension or hypotension. The success in identifying these mutations has expanded our understanding of the physiological pathways that contribute to BP regulation. The single gene variants that cause monogenic forms of hypertension are rare, have large effect sizes, and affect a single pathway, which primarily involves renal electrolyte balance; Table 1-4 shows a summary of the important monogenic forms of BP dysregulation. Secondly, family and twin studies have shown that BP is heritable trait, with heritability that ranging from 15% to 40% for the clinic SBP, and 15% to 30% for clinic DBP; a higher heritability was noted for the ABPM (sleep) around 69% and 51% for SBP and DBP, respectively.[109-111] Additionally, the risk of hypertension is about the double in subjects with one or two hypertensive parents, and BP levels correlate more in monozygotic twins than dizygotic twins.[42,128] The last section introduced the concept and some approaches that are used to dissect the genetic of complex disease. This section reports the findings of these approaches in identifying some types of monogenic forms of hypertension, or revealing common genetic variants that are associated with BP traits.

1.6.1 Monogenic forms of hypertension

Studying the genetic architecture of BP and hypertension started with identifying monogenic Mendelian forms of hypertension, by first using a linkage analysis approach and more recently exome sequencing methods. This has led to identification of several causal mutations in genes, which are primarily related to sodium homeostasis through renal and adrenal mechanisms. Discovering these specific causal pathways has expanded our understanding of the complex mechanisms underlying BP regulation and highlighted how this can be translated into targeted therapy, for instance Gordon’s syndrome patients can be treated by thiazides diuretics, and Familial Hyperaldosteronism (FH) type 1, or “glucocorticoid remediable aldosteronism” with steroid therapy. Hence, finding the genetic mutation that causes the monogenic hypertension is a classical example of directing the therapy based on the information of the culprit genetic

mutations, and the underlying perturbed pathway. The presence of strong family history of hypertension, especially early onset with Mendelian pattern of inheritance indicates monogenic forms. Though, this can sometimes be misleading, as essential hypertension is a highly prevalent disease. Other indications include abnormality in serum potassium level, and aldosterone, suppressed renin, and presence of metabolic acidosis or alkalosis.

Discovery of the monogenic Mendelian forms of hypertension has mainly been through positional cloning using large family pedigrees, with multiple members of the family showing a clear inheritance pattern. This is usually performed by a linkage analysis approach (discussed in Section 1.5.4 - p58), which is based on genotyping highly polymorphic markers across the genome. This approach can reveal the chromosome, and the position of the locus, or even the genes that are most likely to be involved. This is then followed by fine-mapping to identify the culprit variants. Patients with these types of disorders represent less than 1% of patients of essential hypertension and are considered to have secondary hypertension. The contribution of identifying these types of variants is great, as they represent the clearly definable genetic influence on BP.

Mutations causing monogenetic hypertension are characterized by being rare with a major defect that usually disrupts a single pathway. Given the complexity and the presence of several systems and physiological pathways that control BP, it is surprising that most of the identified monogenic hypertension are due to mutation in genes that play key roles in renal-sodium handling. These mutations can be categorized based on their mechanisms, though all of which lead to common pathways related to increased sodium reabsorption, volume expansion, and low plasma renin activity.[7] First, gain of function mutations that increase sodium and chloride reabsorption in the distal tubular, or mutations of mineralocorticoid receptors that resemble mineralocorticoid excess, leading to volume expansion and hypertension, such as in Liddle syndrome, Gordon’s syndrome and hypertension exacerbated by pregnancy. Second, mutations that reduce regulatory enzymes of adrenal steroid synthesis and deactivation, leading to accumulation of precursors with mineralocorticoid activity such as congenital adrenal hyperplasia and apparent mineralocorticoid excess. Third, mutations that causes excessive aldosterone synthesis that escape regulatory mechanisms,

leading to volume-dependent hypertension that suppresses renin release, such as in FH types I and II.[7] Table 1-4 summaries the different forms of monogenic hypertension based on the affected pathway.

It is important to differentiate between the impact of rare variants that cause monogenic forms of hypertension, and common variants that are assumed to play a role in essential hypertension. As discussed before, monogenic hypertension is caused by rare variants that often cluster within families, and have a large effect size. On the other hand, common variants are assumed to have very small effect with unequivocal association with BP, and the functional effects of these variants are obscure. Tobin et al. studied the common variants in genes associated with monogenic forms in the general population, with the aim to evaluate comprehensively the effect of common variants as potential contributors to BP variation in the general population.[129] The study examined the association of 298 SNPs across 11 loci, in 2018 individuals from 520 families unselected for BP, using mean 24-hr SBP and DBP as primary phenotypes. The key findings were for association of five variants in the *KCNJ1* gene (potassium inwardly-rectifying channel, subfamily J), which encodes the renal outer medullary potassium channel (ROMK), which is associated with Bartter syndrome type 2. Evidence of nominal associations were also found for variants in *CASR*, *NR3C2*, *SCNN1B*, and *SCNN1G*. This study showed that common variants in genes causing monogenic forms of hypertension may also play a role in regulating BP in general population.

Table 1-4 Monogenic forms of hypertension

<table>
<tr><th>No</th><th>Condition (mode of inheritance)</th><th>Phenotypes/treatment</th><th>Gene, locus</th><th>Features</th></tr>
<tr><td colspan="5">Renal ion channels
Mutations in this pathway affect the renal net salt homeostasis, by disruption of the normal regulatory mechanisms of ion channel transporters, such as the epithelial Na+ channel (ENaC), the thiazide-sensitive Na/Cl co-transporter (NCC), and the Na-K-2Cl co-transporter (NKCC2). Normally, 60% of sodium is reabsorbed in the proximal convoluted tubule (PCT) of the nephron by Na+/H+ exchange. Then, 30% is reabsorbed in the thick ascending limb of Henle (TAL) by the NKCC2 co-transporter. Then, 7% is reabsorbed in the distal convoluted tubule (DCT) by NCC co-transporter. Finally, 3% is reabsorbed in the cortical collecting tubule (CCT) by ENaC.[7]</td></tr>
<tr><td>1.</td><td>Liddle Syndrome
MIM 177200 (AD)</td><td>↑↑BP, ↓↓serum aldosterone, ↓↓K^+, ↓↓ plasma renin activity, metabolic alkalosis.
Treated with amiloride or triamterence.</td><td>SCNN1B-SCNN1G
16p12.2</td><td>Gain of function mutations in the genes coding β- or γ-subunits of ENaC, leading to increased ENaC activity due to constitutive expression of sodium channels and prolongation of the half-life, which result in net renal salt balance.</td></tr>
<tr><td>2.</td><td>Gitelman syndrome
MIM 263800 (AR)</td><td>↓↓BP, ↑↑ plasma renin activity, ↓↓K^+, ↓↓Mg^{2+}, ↓Ca^{2+}.</td><td>SLC12A3
16q13</td><td>Loss-of function mutation in the NCC co-transporter, leading to lower sodium reabsorption.</td></tr>
<tr><td rowspan="3">3. Bartter syndrome</td><td>Type 1 “antenatal”
MIM 601678 (AR)</td><td rowspan="3">↓↓BP, ↑↑ plasma renin activity, ↑↑aldosterone, ↓↓K^+, metabolic alkalosis and hypercalciuria.</td><td>SLC12A1
15q21.1</td><td>Loss of function mutation in the NKCC2 co-transporter, leading to impaired salt reabsorption in TAL.</td></tr>
<tr><td>Type 2 “antenatal”
MIM 241200 (AR)</td><td>KCNJ1
11q24.3</td><td>KCNJ1 belongs to the potassium channel gene (KCN) family that encodes potassium channels. A loss of function mutation in KCNJ1, which encodes the apical ROMK, leads to a disruption in the normal activity of NKCC2.</td></tr>
<tr><td>Type 3
MIM 607364 (AR)</td><td>CLCNKB
1p36.13</td><td>Homozygous or compound heterozygous mutation in the kidney chloride channel B gene (CLCNKB).</td></tr>
<tr><td colspan="5">Mineralocorticoid Pathway:
This pathway regulates homeostasis of blood volume and pressure by promoting renal sodium, chloride, and water reabsorption through releasing mineralocorticoid hormones, of which aldosterone represents 90%. The mineralocorticoid receptor (MR) is a major regulator of ENaC activity, which is normally activated by aldosterone in CCT. Genetic mutation in this pathway causes abnormalities in aldosterone secretion or production of other steroids that activate MR, leading to increases sodium and chloride reabsorption in the distal nephron, or enhances the effects of hormones with mineralocorticoid activity leading to low plasma renin activity, thus the elevation in BP is more likely to be salt sensitive.</td></tr>
<tr><td>4. Familial Hyperaldosteronism (FH)</td><td>FH Type 1, or glucocorticoid remediable aldosteronism
MIM 103900 (AD)</td><td>↑↑BP, ↑↑ aldosterone, ↓K^+, ↓↓plasma renin activity.
Treated by dexamethasone.</td><td>CYP11B1
8q24.3</td><td>Chimeric CYP11B1/CYP11B2 gene, which encodes a protein with aldosterone synthase activity, but it is regulated by adrenocorticotropic hormone (ACTH), rather than normal hormonal regulator. Thus, higher aldosterone is secreted to maintain normal level of cortisol, leading to expanded plasma volume and suppressed renin activity.</td></tr>
</table>

	FH Type 2 MIM 605635 (AD)		7p22.3- 7p22.1	Hyperaldosteronism because of adrenocortical hyperplasia; not suppressible by dexamethasone.
	FH Type 3 MIM 613677 (AD)		*KCNJ5* 11q24.3	Severe hypertension due to massive mineralocorticoid production, that only can be treated by adrenalectomy.
5.	Apparent Mineralocorticoid Excess (AME) MIM 218030 (AR)	↑↑BP, ↓↓plasma renin activity, ↓↓serum aldosterone, ↓↓K^+. Onset in infancy or childhood.	*HSD11B2* 16q22.1	A congenital defect in 11-beta-hydroxysteroid dehydrogenase type II (HSD11B2) activity, resulting in decreased conversion of biologically active cortisol to inactive cortisone; this defect allows cortisol to act as a ligand for the MR, resulting in sodium retention and volume expansion.
6. Pseudohypoaldosteronism (PHA)	PHA1A MIM 177735 (AD)	↓↓BP, ↓Na^+, ↑↑K^+, ↑↑serum aldosterone, ↑↑ plasma renin activity, metabolic acidosis. Onset in infancy.	*NR3C2* 4q31.2	Loss of function mutation in MR that impairs maximal salt reabsorption, with reduced ENaC activity, leading to salt wasting despite elevated aldosterone levels.
	PHA2B “Gordon syndrome” MIM 614491 (AD)	↑↑BP, ↓↓ plasma renin activity, ↑K^+, ↑↑Cl^-, ↑ aldosterone. Treated by thiazide diuretics.	*WNK4* 17q21.2	Loss-of-function mutations in *WNK4*, or *WNK1* which encode proteins involved in electrolyte homeostasis, leading to reduced potassium excretion from the kidney, despite normal renal glomerular filtration. The pathogenic sequence is similar to PHA1A; despite the high level of aldosterone, the normal target for MR activation, ENaC, is missing.
	PHA2C “Gordon syndrome” MIM 614492 (AD)	↑↑BP, ↓↓ plasma renin activity, ↑K^+, ↑↑Cl^-, ↑ aldosterone. Treated by thiazide diuretics.	*WNK1* 12p12.3	
	PHA2D MIM 614495 (AD/AR)	↑K^+,↑Cl, metabolic acidosis. Age at diagnosis 24 +/- 18 years for AD. Age at diagnosis 26 +/- 14 years for AR.	*KLHL3* 5q31.2	Loss of function missense mutations in KLHL3 that disrupt binding to WNK4, WNK1, or CUL3, leading to decreased ubiquitination and increased levels of WNK4.
	PHA2E MIM 614496 (AD)	↑↑BP, ↓ plasma renin, ↑K^+, ↑↑Cl, metabolic acidosis. Age at diagnosis 18 +/- 6 years.	*CUL3* 2q36.2	Loss of function mutations that disrupt ubiquitination of at least a subset of KLHL3 targets. Patient affected with PHA2E have the worst symptoms of PHA.
7.	Sporadic aldosterone-producing adenoma (APA), or primary aldosteronism (AD)	↑↑BP, ↑↑ aldosterone, ↓K^+, ↓Na^+, ↓↓plasma renin activity, metabolic alkalosis.	*KCNJ5* 11q24.3	Gain of function mutations in and near the selectivity filter of the potassium channel KCNJ5 produces increased sodium conductance and cell depolarization, triggering calcium entry into glomerulosa cells, the signal for aldosterone production and cell proliferation.
			ATP1A1 1p31.1	Mutations in *ATP1A1* that encodes Na^+/K^+ ATPase α subunit that is expressed in adrenal cells and control Na^+, K^+, Ca^{2+} homeostasis.

			CACNA1D 3p21.3	Mutation in *CACNA1D*, which encodes voltage-sensitive Ca^{2+} channels that regulates hormone and neurotransmitter release.
			ATP2B3 Xq28	Mutations in *ATP2B3* that encodes Ca^{2+} ATPase, which is expressed in adrenal cells and control Na^{+}, K^{+}, Ca^{2+} homeostasis.
8.	Hypertension exacerbation in pregnancy MIM 605115 (AD)	↑↑BP, ↓K^{+}, ↓↓plasma renin activity, ↓↓aldosterone.	*NR3C2* 4q31.2	Gain of function mutation in the ligand binding domain of MR that causes increased renal reabsorption and hypertension.
Glucocorticoid Pathway: Defects in enzymes of cortisol biosynthesis result in a group of autosomal recessive disorders collectively called congenital adrenal hyperplasia. In some of these syndromes, plasma ACTH will increase in an attempt to produce cortisol, and the some of the aberrant products that accumulate can result in hypertension.				
9.	11β-hydroxylase MIM 202010 (AR)	↑↑BP, ↑↑ACTH, ↓↓aldosterone, ↓↓renin, ↓↓cortisol, ↓↓K^{+}, ↑↑FSH, ↑↑deoxycorticosterone. Treated by glucocorticoid therapy.	*CYP11B1* 8q21	Loss of function mutation in *CYP11B1*, leading to decreased conversion of 11-deoxycortisol and 11-deoxycorticosterone to cortisol and corticosterone, respectively; resulting in accumulation of 11-deoxycortisol and 11-deoxycorticosterone; which is a potent salt-retaining mineralocorticoid that leads to arterial hypertension.
10.	3β-hydroxysteroid dehydrogenase (AR) OMIM 613890		*HSD3B2* 1p12	Genetic mutation in *HSD3B2*, which is important for production of all classes of steroid hormones.
11.	17α-hydroxylase deficiency MIM 202110 (AR)		*CYP17A1* 10q24.3	Loss of function mutation in *CYP17A1*, which encodes steroid 17α-hydroxylase, leading to an excessive amount of corticosterone and deoxycorticosterone resulted in hypertension and hypokalemic alkalosis.
12.	21-Hydroxylase deficiency MIM 201910 (AR)		*CYP21A2*	Genetic mutation in *CYP21A2*, that encodes 21-hydrozylase enzyme, which is essential for adrenal steroidgenesis.
Sympathetic pathway Monogenic forms of hypertension that result from disruption of the sympathetic pathways is caused by genetic mutations that leads to rare neuroendocrine tumours in the form of phaeochromocytomas and paragangliomas. This is accompanied with higher level of catecholamines that increases the sympathetic activity.				
13. Paragangliomas (PGL)	Paragangliomas 1 MIM 168000 (AD)	Tumours or extraadrenal paraganglia associated pheochromocytoma, ↑↑catecholamines level, ↑↑BP, and ↑↑heart rate.	*SDHD* 11q23.1	Rare tumours diffuse paraganglionic tissues that are located internally and centrally around the major arteries, nerves, within organs.[130] PGL is characterized by genetic mutation in any of the four subunits of the mitochondrial succinate dehydrogenase enzyme complex (SDH), which catalysed the conversion of succinate to fumarate in the Krebs cycle and serves as
	Paragangliomas 2 MIM 601650 (AD)		*SDHAF2* 11q12.2	
	Paragangliomas 3 MIM 605373 (AD)		*SDHC* 1q23.3	
	Paragangliomas 4 MIM 115310 (AD)		*SDHB* 1p36.13	

	Paragangliomas 5 MIM 614165 (AD)		*SDHA* 5p15.3	complex II of the electron transport chain. *SDHAF2* encodes a protein that is necessary for SDH function. These are collectively known as the *SDHx* genes.
14.	von Hippel–Lindau syndrome MIM 193300 (AD)	↑↑BP, retinal, cerebellar, and spinal hemangioblastoma, renal cell carcinoma, pheochromocytoma, and pancreatic tumours.	*VHL* 3p25.3	Genetics mutation in *VHL*, that encodes protein important in tumour suppression mechanisms.
15.	Multiple endocrine neoplasia, type IIA MIM 171400 (AD)	Multiple endocrine neoplasms, including medullary thyroid carcinoma, pheochromocytoma, and parathyroid adenomas. ↑↑BP.	*RET* 10q11.2	Genetic mutation in the RET oncogene, that produce constitutively activated receptors, leading to several endocrine and neural-crest-derived tumour.
16.	NOS3-pregnancy-induced hypertension (AD) MIM +163729	Hypertension that is exacerbated during pregnancy.	*NOS3* 7q36.1	Genetic mutation in *NOS3*, leading to dysfunction releases of NO, which an important role in the maintenance of cardiovascular and renal homeostasis.
Table is compiled from information in these reviews [37,38,40], direction of the arrow indicates either low level or high level.				

1.6.2 Overview of GWAS for BP and hypertension

Several GWAS have been conducted using BP as a quantitative trait, or by using a binary definition of hypertension. The first GWAS was a case-control design from the Wellcome Trust Case Control Consortium (WTCCC), published in 2007.[131] The study examined seven complex diseases of major public health importance using 2,000 cases in each, and 3,000 shared control. The study genotyped approximately 500,000 SNPs using the 500 K Affymetrix SNP chip, and reported a total of 24 significant SNPs ($p<5.0x10^{-7}$) that were associated with the six examined diseases, apart from hypertension, which was the only trait without any significant signal even in the regions that previously showed evidence of association. This failure in identifying significant SNPs also extended to the first GWAS that analysed BP as a quantitative trait in the Framingham Heart Study, in which almost 71,000 SNPs were genotyped in about 1,400 related individuals.[132] The study used six primary phenotypes for BP that were derived from single and long-term averaged (LTA) SBP and DBP, in which all the analysed phenotypes failed to produce significant association signals. These two studies represent the first attempts at applying the GWAS approach for hypertension and BP. Although no association signals were reported for hypertension, important lessons were taken from these two attempts. For instance, failing to identify any association signal for hypertension only in WTCCC emphasized the complexity of hypertension, and the need for having much larger sample size to reveal association signals for genetic marker with low effect size.

The first two successful GWAS for BP were reported in 2009 by two large consortia, the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) study [115], and the Global Blood Pressure Genetic (Global BPgen) study.[133]The limited resources of case-control samples necessitated the study of BP primarily as a quantitative trait in these two studies. The CHARGE consortium included six population-based cohorts of European ancestry with a total sample size of 29,000 individuals, and the Global BPgen consisted of 17 cohorts with a total sample size of 34,000 at the discovery phase. The two consortia tested the association of SBP and DBP as the primary phenotypes, using a cross-sectional measurement with addition of a fixed value of 15/10 mmHg or 10/5 mmHg for individuals taking antihypertensive therapies in CHARGE and Global BPgen, respectively. In order to combine the results from different cohorts, imputation

was performed to impute the missing markers between the included cohorts, allowing to merge the genotypes results from different genotyping chips and platforms. The final association tests were performed in almost 2.5 million genotyped or imputed SNPs and discovered 13 loci independently associated with SBP or DBP at a level of genome-wide significance ($p<5.0x10^{-8}$). Each study reported eight loci with three loci overlapping in both studies. These two studies have been followed by further GWASs and the results of these studies are summarised in Table 1-5. In addition, most of the loci reported in these two studies were novel except for some loci such as *CYP17A1-NT5C2* and *MTHFR-NPPB*, the former has been associated with a rare Mendelian form of hypertension, and the later lies in a region that has previously been associated with BP and hypertension.[134]

Most of the studies in CHARGE and Global BPgen consortia were then included in a larger follow-up study by the ICBP consortium in 2011.[107] The ICBP included more than 69,000 individuals in the discovery sample, followed by a replication in 133,000 individuals, making it the largest GWAS meta-analysis of BP to date. The SNPs association analyses were performed under an additive genetic model, which assumes that the effect conferred by an allele is increased by r-fold for heterozygotes and 2r-fold for homozygotes. The model was adjusted for sex, BMI, age, and its square (to account for the middle age plateau of DBP). Also, a fixed value of 15/10 mmHg was added to individuals taking antihypertensive treatment to account for treatment effect. The study identified 29 independent SNPs at 28 loci, of which 16 loci were novel and the remaining 13 loci were a replication of the previously reported loci in CHARGE or Global BPgen. Although the majority of SNPs identified by ICBP were intragenic, some loci were in gene desert regions or in genomic regions that has no gene encoding protein with a biological plausible effect on BP.

A second study was also carried out by the ICBP consortium using MAP and PP as primary phenotypes with the addition of a further six studies in the consortia, increasing the total discovery sample size to more than 74,000 individuals.[135] The study identified four novel loci associated with PP and two loci associated with MAP, with one locus associated with both traits near to *FIGN*. The importance findings of this study is that three of the four loci associated with PP

were found to have an opposite effect in SBP and DBP, unlike the majority of BP variants that exerts effect in the same direction on SBP and DBP, which suggests the presence of genetic pathways that may differentially influence SBP and DBP. The study has also showed that most of MAP variants were also associated with both SBP and DBP, suggesting a high correlation between these three BP traits.

Most of the GWAS for BP have taken the quantitative approach by studying BP as a quantitative trait, except for two studies that analysed hypertension as a binary trait.[82,136] The first study has successfully identified a novel locus located in the promoter region of Uromodulin gene (*UMOD)*, which is exclusively expressed in the kidney and may influence BP by a novel sodium homeostatic pathway.[82] An alternative strategy were taken by this study in recruiting case and control groups by selecting individuals from the extreme of the BP distribution, this strategy has allowed a sharper contrast between case and control, and hence a smaller sample size would be required compared to a normal case-control study from the general population. The identified variant has also showed a suggestive evidence for association with SBP and DBP, with a consistent direction of effect with the odds of hypertension. The second study has used a classical case-control approach using the HYPERGENES Project, and identified a new locus in the promoter region of the endothelial NO synthase gene, which is a critical mediator for cardiovascular homeostasis and BP control via vascular tone regulation[136]

GWAS for populations other than European descent were also performed with the aim of replicating the variants identified in European populations, and also finding new population-specific loci. The Asian Genetic Epidemiology Network Blood Pressure (AGEN-BP) was the largest non-European GWAS that included more than 30,000 individuals in the discovery stage and 20,000 for replication.[137] AGEN-BP identified six novel loci and confirmed seven loci previously reported in CHARGE and Global BPgen. The Continental Origins and Genetic Epidemiology Network (COGENT) study also performed another large GWAS using a trans-ethnic meta-analysis and discovery sample size of 29,000 individuals of African American (AA) origin[138] The replication sample included a mixed ethnic background of European and East Asian origins due to a lack of sufficient samples from AA. The COGENT study reported five loci associated with SBP or

DBP, three of which were not previously reported to be associated with BP. Another GWAS was also performed recently in the Chinese population and reported 3 novel loci, and replicated 14 previously reported loci.[139] The success of replicating the previously reported loci for European population in the other population suggests that the physiologic effects of these loci may be generalized across populations with diverse genetic backgrounds. Yet, identifying novel loci also suggests that populations with different genetic background may have a unique genetic factors as a result of differences in allele frequencies or population-specific factors that interact with genes to influence BP.

Several studies have used measured other than the single-time BP phenotypes in their GWAS to improve the phenotype accuracy. One study used the average of BP measurements across time instead of the single-visit measurement approach that is typically used in most of GWAS.[140] This study identified 39 association signals at 19 loci; of them, 4 loci were novel. The study has also estimated a 20% improvement in statistical power with using the LTA approach over the single-visit method (discussed in more details in the next Section 1.6.3 - p87). Another study has used the visit-to visit variability in BP and identified a cluster of genetic variants within the *NLGN1* gene (3q26.31), but further replication of this finding is required.[141] Finally, a family study of 2020 individuals used the mean of 24-hr BP measurements that were derived from the ABMP, and reported association of 24-hr DBP with a SNP in the promoter region of *MTHFR* and *CLCN6* genes.[142]

A large scale study was performed by CHARGE, Global BPgen, and ICBP consortia by taking a non-standard approach to assess the gene-age interaction at GWAS level, in the first attempt to assess the gene-age interaction on BP using common variants from GWAS data.[143] Unlike other GWAS that adjust for age by including it in the covariates, this study stratified subjects into 10-year age bins (20-29 years, 30-39 years …etc.) and then conducted the SNP association test within each age bin separately. The resulted SNP main effect estimates and their standard errors from all age bins were then entered into a meta-regression analyses through a linear regression of the SNP effect estimates onto the median age of each sub-group. The study was performed in a two-stage design using more than 99,000 individuals, and reported 20 variants using joint tests of the

SNP main effect and SNP-age interaction. The SNPs with the largest age-gene interaction in three loci (*CASZ1*, *EHBP1L1*, and *GOSR2*) displayed opposite directions of effect by increasing BP in the young and decreasing BP in the old, by a difference in the effect size that can reach up to 1.58 mmHg. In a secondary analysis, the study explored the age-specific effects by meta-analysing the results within each age bin separately, and identified 22 distinct loci with evidence of age-specific effects. An important message from this study is that pooling data from different studies with a wide range of age distribution may obscure genetic effects that are age dependent. Gene-environmental interaction were also assessed in two other studies with smaller sample size; in the first; gene-alcohol interactions were found for SNP rs10826334 near *SLC16A9* modulated by both the number of alcoholic drinks and the ounces of alcohol consumed per week, as SBP decreased by 3.8 mmHg in those consuming 14 drinks/week compared to only 0.46 in non-drinkers.[144] The same group also assessed the gene-smoking and gene-education interactions in another two studies[145,146]. However, their findings have not been replicated in any external samples and future work is required to validate the reported markers.

Table 1-5 Loci reported in GWAS for association with hypertension or BP traits

#	Locus	Nearby gene	GWAS SNP	Population: Ref	Trait	CA	CAF	β (se)
1	1p36.2	*CASZ1*	rs880315	EA: 115	SBP	C	0.35	0.89 (0.17)
				AS: 147	SBP DBP	C	0.59	1.08 (0.70 –1.46) 0.79 (0.56–1.01)
				AS: 137	DBP	C	0.65	0.56 (0.09)
				AS: 139	SBP DBP HTN	C	0.63	0.97 (0.16) 0.46 (0.09) 0.09 (0.16)
				EA: 143	SBP MAP	T	0.64	-0.43 (0.08) -0.27 (0.05)
				EA: 140	LTA SBP LTA MAP LTA PP	T	NR	-0.71 (0.1) -0.46 (0.07) -0.42 (0.07)
2	1p36.22	*MTHFR, CLCN6, NPPA, NPPB*	rs17367504	EA: 133	SBP	G	0.14	-0.85 (0.63-1.07)
				AS: 147	SBP DBP	A	0.90	0.65 (0.07–1.24) 0.34 (0.01–0.69)
				EA: 107	SBP DBP HTN	G	0.15	-0.90 -0.55 -0.10
			rs5068	EA: 134	SBP DBP	C	0.06	-0.08 (0.02) -0.08 (0.02)
3	1p13.2	*SLC16A1, CAPZA1, ST7L, MOV10*	rs2932538[a]	EA: 107	SBP DBP	G	0.75	0.39 0.24
			rs17030613	AS: 137	DBP	C	0.49	0.38 (0.07)
			rs10745332[a]	AS: 139	SBP DBP HTN	A	0.82	0.96 (0.18) 0.53 (0.1) 0.11 (0.02)
4	1q32.1	*MDM4*	rs2169137	EA: 148	DBP	G	0.27	-0.35 (0.07)
5	1q42.2	*AGT*	rs2004776	EA: 149	SBP DBP HTN	T	NR	0.42 (0.09) 0.32 (0.06) 0.08 (0.02)
				EA: 150	HTN	T	0.24	0.14 (0.02)
6	2p23.2	*KCNK3*	rs1275988	EA: 140	LTA SBP LTA MAP	T	NR	-0.60 (0.09) -0.39 (0.06)
7	2q11.2	*FER1L5*	rs7599598	EA: 140	LTA DBP	A	NR	-0.31 (0.05)
8	2q24.3	*FIGN*	rs1446468	EA: 135	SBP DBP MAP	T	0.53	-0.50 (0.07) -0.26 (0.05) -0.34 (0.05)
9	2q24.3	*FIGN*	rs13002573	EA: 135	PP	G	0.20	-0.31 (0.05)
10	2q24.3	*FIGN*	rs16849225	AS: 137	SBP	C	NR	0.75 (0.11)
11	2q24.3	*STK39*	rs6749447	EA: 151	SBP	G	0.19	1.9 (0.6)
12	2q32.1	*PDE1A*	rs16823124	EA: 152	DBP MAP	A	0.30	0.26 (0.04) 0.27 (0.05)
13	3p25.3	*HRH1-ATG7*	rs347591	EA: 148	SBP	G	0.36	-0.53 (0.11)
14	3p24.1	*SLC4A7*	rs13082711	EA: 107	DBP	T	0.78	-0.24
			rs820430	AS: 139	SBP	A	0.32	0.76 (0.11)
15	3p22.1	*ULK4*	rs9815354[a]	EA: 115	DBP	A	0.17	0.49 (0.08)
				AS: 139	DBP	A	0.19	0.67
			rs3774372[a]	EA: 107	DBP	T	0.83	-0.37
			rs1717027[a]	Mix:138	DBP	T	0.46	0.49 (0.10)
16	3p21.31	*MAP4*	rs319690	EA: 135	SBP DBP MAP	T	0.51	0.42 (0.07) 0.28 (0.05) 0.29 (0.05)
				AS: 153	MAP	T	0.67	0.38 (0.20)
				AS: 154	MAP	T	0.71	0.26 (0.11)
			rs7651237	EA: 143	DBP	G	0.71	0.30 (0.07)

#	Locus	Nearby gene	GWAS SNP	Population: Ref	Trait	CA	CAF	β (se)
					MAP			0.34 (0.05)
17	3p21.1	*CACNA1D*	rs9810888	AS: 139	SBP DBP	G	0.39	0.53 (0.10) 0.39 (0.06)
18	3q26.1	*MIR1263*	rs16833934	EA: 143	DBP MAP	G	0.26	-1.63 (0.29) -1.33 (0.27)
19	3q26.2	*MECOM*	rs419076[a]	EA: 107	SBP DBP	G	0.47	0.41 0.24
				EA: 135	SBP DBP MAP	T	0.44	0.50 (0.07) 0.30 (0.04) 0.34 (0.04)
20	4q12	*CHIC2*	rs871606	EA: 135	PP	T	0.85	0.43 (0.08)
				AS: 153	PP	T	0.79	0.40 (0.23)
21	4q21.21	*FGF5*	rs16998073[a]	EA: 134	DBP	T	0.24	0.36 (0.12)
				AS: 147	SBP DBP	T	0.31	1.51 (1.12–1.89) 0.82 (0.59–1.05)
				AS:137	SBP DBP	T	0.30	1.43 (0.20) 0.76 (0.11)
			rs1458038[a]	EA: 107	SBP DBP	T	0.29	0.71 0.45
				EA: 135	SBP DBP MAP	T	0.30	0.56 (0.08) 0.40 (0.05) 0.40 (0.05)
22	4q24	*SLC39A8*	rs13107325	EA: 107	SBP DBP	T	0.05	-0.98 -0.68
				EA: 135	SBP DBP MAP	T	0.12	-0.90 (0.14) -0.60 (0.09) -0.63 (0.10)
23	4q25	*ENPEP, PITX2*	rs6825911	AS:137	DBP	C	0.51	0.39 (0.07)
24	4q32.1	*GUCY1A3-GUCY1B3*	rs13139571	EA: 107	DBP	C	0.76	0.26
25	5p13.3	*NPR3-C5orf23*	rs1173771[a]	EA: 107	SBP DBP HTN	G	0.60	0.50 0.26 0.06
				EA: 135	SBP MAP PP	G	0.52	0.51 (0.07) 0.28 (0.05) 0.28 (0.05)
			rs7733331[a]	EA: 140	LTA SBP	T	NR	-0.55 (0.09)
			rs1173766	AS:137	SBP	C	0.60	0.63 (0.11)
26	5q33.3	*EBF1*	rs11953630	EA: 107	SBP DBP	T	0.37	-0.41 -0.28
27	6p22.2	*HFE*	rs1799945	EA: 107	SBP DBP HTN	G	0.14	0.63 0.46 0.10
				EA: 150	DBP	G	0.14	0.50 (0.12)
				EA: 148	DBP	G	0.15	0.41 (0.09)
				AS: 139	SBP DBP HTN	G	0.04	0.95 (0.36) 0.88 (0.20) 0.16 (0.04)
			rs198823	EA: 140	LTA DBP	T	NR	-0.33 (0.06)
28	6p21.33	*BAG1*	rs805303	EA: 107	SBP DBP HTN	G	0.61	0.38 0.23 0.05
29		*CYP21A2*	rs2021783	AS: 139	SBP DBP HTN	C	0.79	0.68 (0.12) 0.49 (0.07) 0.09 (0.01)
30	6p21.32	*HLA-DQB1*	rs2854275	EA: 152	DBP	A	0.13	-0.56 (0.10)
31	6p21.1	*CRIP3*	rs10948071	EA: 140	LTA PP	T	NR	-0.38 (0.07)
32	6q22.33	*RSPO3*	rs13209747	Mix: 138	SBP	T	0.19	0.85 (0.21)

#	Locus	Nearby gene	GWAS SNP	Population: Ref	Trait	CA	CAF	β (se)
					DBP			0.56 (0.12)
33	6q25.1	*PLEKHG1*	rs17080102	Mix: 138	SBP DBP	C	0.10	-1.02 (0.25) -0.74 (0.15)
34	7p15.2	*EVX1-HOXA*	rs17428471	Mix: 138	SBP DBP	T	0.14	1.20 (0.24) 0.61 (0.14)
35	7p12.3	*IGFBP3*	rs2949837	EA: 140	LTA PP	A	NR	0.40 (0.07)
36	7q21.2	*CDK6*	rs2282978	EA: 152	PP	C	0.34	0.27 (0.05)
37	7q22.3	*PIK3CG*	rs17477177[a]	EA: 135	SBP PP	T	0.71	-0.55 (0.08) -0.42 (0.06)
			rs12705390[a]	EA: 143	PP	G	0.78	-0.42 (0.06)
				EA: 140	LTA SBP LTA PP	A	NR	0.63 (0.11) 0.59 (0.08)
38	7q36.1	*NOS3*	rs3918226	EA: 150	DBP	T	0.08	0.78 (0.21)
				EA: 136	HTN	T	0.12	OR 1.54 (1.37-1.73)
39	8p23.1	*BLK-GATA4*	rs4841569	EA: 143	SBP MAP	G	0.57	0.14 (0.31)* 0.26 (0.21)*
			rs2898290	EA: 155	SBP	C	0.53	NR
40	8q24.12	*NOV*	rs2071518	EA: 135	PP	T	0.17	0.31 (0.05)
41	10p12.31	*CACNB2*	rs11014166[a]	EA: 115	DBP	A	0.66	0.37 (0.06)
				AS: 156	SBP	T	NR	−0.19 (0.69)
			rs1813353[a]	EA: 143	SBP DBP MAP	T	0.68	0.53 (0.29)* 0.58 (0.19)* 0.56 (0.20)*
			rs4373814	EA: 107	SBP DBP HTN	G	0.55	-0.37 -0.22 -0.05
			rs12258967	EA: 140	LTA SBP LTA DBP LTA MAP	C	NR	0.35 (0.06) 0.63 (0.10) 0.45 (0.07)
42	10q21.2	*c10orf107*	rs1530440	EA: 134	DBP	T	0.19	-0.44 (0.12)
			rs4590817	EA: 107	SBP DBP HTN	G	0.84	0.65 0.42 0.10
			rs12244842	EA: 140	LTA DBP LTA MAP	T	NR	-0.38 (0.06) -0.48 (0.08)
			rs7070797	EA: 140	LTA SBP	A	NR	-0.74 (0.13)
43	10q22.2	*VCL*	rs4746172	EA: 152	DBP MAP	C	0.25	0.23 (0.04) 0.28 (0.05)
44	10q23.33	*PLCE1*	rs932764	EA: 107	SBP HTN	G	0.44	0.48 0.05
45	10q24.32	*CYP17A1-NT5C2*	rs1004467a	EA: 115	SBP	A	0.90	1.05 (0.16)
			rs11191548a	EA: 134	SBP	T	0.92	1.05 (0.27)
				AS:137	SBP DBP	T	0.74	1.18 (0.14) 0.58 (0.08)
				EA: 107	SBP DBP	T	0.91	1.10 0.46
			rs12413409	AS: 147	SBP DBP	G	NR	1.58 (1.18–1.98) 0.76 (0.53–1.00)
			rs4409766a	AS: 139	SBP DBP	T	0.71	1.24 (0.15) 0.59 (0.09)
			rs3824755	EA: 148	SBP PP	C	0.10	-0.64 (0.12) -0.64 (0.12)
46	10q25.3	*ADRB1*	rs2782980	EA: 135	MAP	T	0.20	-0.39 (0.06)
			rs7076938	EA: 148	MAP	C	0.28	-0.39 (0.08)
			rs1801253	EA: 143	DBP MAP	G	0.27	-0.29 (0.20)* -0.34 (0.22)*

#	Locus	Nearby gene	GWAS SNP	Population: Ref	Trait	CA	CAF	β (se)
47	11p15.5	*LSP1-TNNT3*	rs661348	EA: 150	MAP	T	0.57	-0.65 (0.11)
				EA: 148	SBP	C	0.43	0.47 (0.10)
48	11p15.4	*ADM*	rs7129220	EA: 107	SBP	G	0.89	-0.62
49	11p15.1	*PLEKHA7*	rs381815	EA: 115	SBP	T	0.26	0.65 (0.11)
				EA: 107	SBP DBP	T	0.26	0.56 0.35
				EA: 143	SBP DBP MAP	T	0.25	0.03 (0.22)*
		PIK3C2A, NUCB2, NCR3LG1	rs757081	EA: 152	SBP MAP PP	G	0.33	0.26 (0.05) 0.32 (0.05) 0.40 (0.06)
50	11p15.2	*SOX6*	rs2014408a	EA: 150	MAP	T	0.21	0.58 (0.13)
				EA: 148	SBP	C	0.43	0.47 (0.10)
				Mix: 138	DBP	T	0.46	0.45 (0.10)
			rs4757391a	AS: 139	SBP DBP	C	0.28	0.88 (0.15) 0.49 (0.09)
51	11q13.1	*EHBP1L1*	rs4601790	EA: 143	DBP MAP	G	0.27	0.84 (0.21)* 0.91 (0.22)*
52	11q13.1	*RELA*	rs3741378	EA: 152	SBP MAP	T	0.13	-0.36 (0.07) -0.55 (0.09)
53	11q22.1	*FLJ32810-TMEM133*	rs633185	EA: 107	SBP DBP HTN	G	0.28	-0.56 -0.33 -0.07
54	11q24.3	*ADAMTS8*	rs11222084	EA: 135	PP	T	0.38	0.337 (0.05)
55	12q13.13	*HOXC4*	rs7297416	EA: 152	SBP	C	0.30	-0.33 (0.07)
56	12q21.33	*ATP2B1*	rs11105354[a]	EA: 150	HTN	G	0.16	-0.12 (0.04)
			rs2681492[a]	EA: 115	SBP DBP HTN	T	0.80	0.85 (0.13) 0.50 (0.08) 0.15 (0.02)
				AS: 154	MAP	T	0.67	0.61 (0.12)
			rs2681472[a]	AS: 147	SBP DBP	A	NR	0.99 (0.62–1.35) 0.43 (0.21–0.64)
				EA: 148	SBP MAP	G	0.17	-0.97 (0.16) -0.61 (0.11)
				EA: 140	LTA SBP LTA DBP LTA MAP	A	NR	0.52 (0.07) 0.95 (0.12) 0.69 (0.08)
			rs17249754[a]	AS:137	SBP DBP	G	0.64	1.17 (0.13) 0.58 (0.08)
				EA: 107	SBP DBP HTN	G	0.84	0.93 0.52 0.13
				AS: 139	SBP DBP HTN	G	0.64	1.03 (0.15) 0.52 (0.08) 0.084 (0.016)
57	12q24.12	*SH2B3*	rs3184504[a]	EA: 115	SBP DBP	T	0.48	0.58 (0.10) 0.48 (0.06)
				EA: 107	SBP DBP	T	0.47	0.60 0.45
				EA: 140	LTA DBP LTA MAP	T	NR	0.39 (0.05) 0.45 (0.06)
			rs653178[a]	EA: 134	DBP	T	0.52	-0.46 (0.05)
58	12q24.13	*RPL6-ALDH2*	rs11066280	AS:137	SBP DBP	T	0.74	1.56 (0.13) 1.01 (0.08)
				AS: 139	SBP DBP	T	0.81	0.96 (0.17) 0.62 (0.10)
59	12q24.21	*TBX5-TBX3*	rs35444	AS:137	DBP	A	0.75	0.50 (0.08)
				EA: 140	LTA SBP	A	NR	0.55 (0.09)

#	Locus	Nearby gene	GWAS SNP	Population: Ref	Trait	CA	CAF	β (se)
					LTA MAP			0.36 (0.06
			rs2384550[a]	EA: 115	DBP	A	0.35	-0.35 (0.06)
			rs10850411	EA: 107	SBP DBP	T	0.7	0.35 0.25
			rs1991391[a]	AS: 139	DBP	G	0.85	0.60 (0.20)
		MED13L	rs11067763	AS: 139	SBP DBP	A	0.62	0.81 (0.10) 0.51 (0.06)
60	15q21.1	*FBN1*	rs1036477	EA: 152	PP	G	0.11	-0.40 (0.08)
61	15q24.1	*CYP1A1-ULK3*	rs6495122	EA: 115	DBP	A	0.42	0.40 (0.06)
				EA: 155	DBP	A	0.43	NR
			rs1378942	EA: 134	DBP	C	0.33	0.43 (0.04)
				EA: 107	SBP DBP	C	0.35	0.61 0.42
				EA: 135	SBP DBP MAP	C	0.33	0.59 (0.07) 0.42 (0.05) 0.37 (0.05)
62	15q24.2	*COX5A*	rs11072518	EA: 140	LTA SBP LTA MAP	T	NR	0.57 (0.09) 0.43 (0.06)
			rs1133323	EA: 140	LTA DBP	T	NR	-0.33 (0.05)
63	15q26.1	*FURIN-FES*	rs2521501	EA: 107	SBP DBP	T	0.31	0.65 0.36
				EA: 135	SBP DBP MAP	T	0.37	0.58 (0.09) 0.37 (0.05) 0.34 (0.06)
64	16p12.3	*UMOD*	rs13333226	EA: 82	HT	G	NR	OR 0.87 (0.84–0.91)
65	16q22.1	*NFAT5*	rs33063	EA: 152	PP	A	0.14	0.34 (0.07)
66	17q21.31	*PLCD3*	rs12946454	EA: 134	SBP	T	0.27	0.57 (0.10)
67	17q21.32	*GOSR2*	rs17608766	EA: 107	SBP	T	0.86	-0.56
				EA: 143	PP	T	0.84	-0.52 (0.32)*
68	17q21.33	*ZNF652*	rs12940887	EA: 107	SBP DBP	T	0.38	0.36 0.27
			rs16948048	EA: 134	DBP	G	0.37	0.31 (0.05)
69	20p12.2	*JAG1*	rs1327235[a]	EA: 107	SBP DBP	G	0.46	0.30 0.30
				EA: 135	DBP MAP	G	0.58	0.26 (0.04) 0.26 (0.05)
			rs1887320[a]	AS: 139	SBP DBP	A	0.53	0.78 (0.14) 0.43 (0.08)
70	20q13.32	*GNAS-EDN3*	rs6015450	EA: 107	SBP DBP HTN	G	0.12	0.90 0.56 0.11
		C20orf174	rs6092743	EA: 140	LTA SBP LTA DBP LTA MAP	A	NR	0.84 (0.14) 0.50 (0.08) 0.64 (0.10)

SNPs in the same locus with a superscript “a” are in LD (r >0.80), NR: Not Reported
* Effect size is the theoretical SNP effect on BP in mmHg at birth (age =0).
SNPs that are genotyped in this thesis are underlined.

1.6.3 Challenges of GWAS for BP and hypertension

Despite the promising success of the second wave of GWAS for hypertension and BP traits, with findings of almost 70 loci as shown in Table 1-5, the proportion of phenotypic variance that is explained by all of these loci together is less than 2.5%.[107] This phenomenon of “missing heritability” is not restricted to BP traits, but has been observed in almost all the findings of GWAS for the complex traits. For instance, a classic complex trait such as height has a very large heritability estimate from family studies (about 80%), yet the identified loci through GWAS explain less than 10% of the phenotypic variance despite studies with very large sample sizes (>180,000 individuals).[157,158] Later, a genome-wide complex trait analysis (GCTA) approach in unrelated individuals showed that 45% of height variance can be explained by the common SNPs (h^2_{SNP}) in the human genome, leaving more than 30% of the variance in height still unexplained.[159] The GCTA approach was introduced by Yang et al. in 2010, and is based on estimating the heritability from unrelated individuals using common SNPs with the assumption that heritability estimates in unrelated individuals is only attributable to the common SNPs, while the estimation in related individuals is attributed to the entire genome.[159]

Applying the same approach to SBP has shown that h^2_{SNP} was about 24%, which is approximately 50% of the heritability estimates from other twin-studies, and about 80% of the same study heritability estimate (h^2 =30%).[160] Furthermore, the number of independent variants with similar effect size to those reported in the ICBP study was estimated to be 116 (95% CI: 57-174), which can collectively explain around 2.2% of the phenotypic variance for BP phenotypes, compared with only 0.9% explained by the 29 SNPs identified by ICBP.[107] These findings indicate that a large proportion of the heritability of BP is hidden rather than missing because of large number of common variants, each of which has too small an effect to be detected at the stringent genome-wide significance level using current sample sizes. Another possible explanation for this is that a typical GWAS does not consider the non-standard genetic contributions such as allelic heterogeneity, rare alleles, epistasis, parent-of-origin effects, and genetic variance heterogeneity, all of which can make significant contributions to the phenotypic variance.

A general trend in GWAS approaches is to increase sample sizes to enable the detection of variants with small effect size, especially for quantitative traits as they are more powered than binary traits in this setting (i.e. detection of variants with modest effect size). Basically, this relies on the hypothesis that if a part of the phenotypic variance can be explained by genetic factors, then increasing sample size would allow alleles with modest effect to gain statistical significance. Although studies with such enormous sample size offer the statistical power to detect larger number of variants with smaller effect size, it was argued that increasing the sample size may also scale the genetic heterogeneity in parallel, making it even harder to detect risk alleles.[161,162] That is, increasing sample size in GWAS without unfolding heterogeneity of complex traits such as BP and hypertension may reduce the power of GWAS. For example, a case-control study of 5,000 cases and 5,000 controls with case misdiagnosis proportion equal to 20% has equivalent power of only 3,200 cases and 3,200 controls without misdiagnosis cases or control.[163]

A challenge for the success of most GWAS is the accuracy of phenotype, as most of the common disease phenotypes suffer from low resolution and imprecision.[161] The majority of explanation and solutions of the missing heritability are related to the genetic factors, such as epistasis, CNV, gene-environment interactions, or epigenetics. However, the way complex traits are measured, and phenotypic information is modelled is at least as important in GWAS as these genetic factors.[164] For hypertension and BP traits, this represents a real challenge due to several factors that are related to the complexity of the disease *per se* and methods of measurements (discussed in Section 1.2.3.1 - p26). Typically, BP measurements for GWAS are based on a single-time visit; when more than one BP measurements are taken and the average of the last two measurements are used to calculate the measured BP level. Although this practice has proved valuable, it might be affected by several factors that can influence the BP measurements, and introduce additional sources of variability (noise) with resulting loss of statistical power to detect association signals.

One way to strengthen the phenotype accuracy as performed by Ganesh et al. was to use a longitudinal phenotype data (i.e. repeated measures) of BP.[140] This study modelled the BP phenotypes by LTA approach, in which repeated

measurements of BP were taken for each participant that were at least one-year apart and within a 15-year timespan. The LTA approach intended to reduce the BP measurement errors that may add another source of BP variability, and hence improving the phenotype accuracy and the study power. The study has identified 39 association signals at 19 loci; of them, 4 loci were not previously discovered. The study has also estimated a 20% improvement in statistical power with using the LTA approach over the single-visit method. This improvement is a result of reducing the BP variability that can arise due to different factors, for instance, in addition to the variability in BP introduced by imprecise measurement techniques, BP can also vary for the same individual during the day time following different factors such as smoking or “white-coat” effect. Remarkably, it is estimated that between 15% and 20% of the power to detect a genetic association is reduced with increasing the variance attributable to the intra-individual variation and measurement error up to 20%.[165]

When the continuous BP measure is dichotomised into hypertension and normotension, it is likely that the binary phenotype may be affected by phenotype factors that are not very important for quantitative measures. Thus it is important to recognise that whilst both traits are complex polygenic traits, they may not be entirely similar in terms of gentic architecture and phenotypic confounders. Studying hypertension as a binary trait has been performed in two GWAS, where participants with BP of 140/90 or higher, or taking antihypertensive medications were classified as cases.[82,136] Similar to the typical approach in the quantitative approach, BP measurements that are used to classify individuals are usually taken from a single visit. This is different to a clinical diagnosis of hypertension, in which a confirmation methods are required before providing a clinical diagnosis of hypertension. For instance, the National Institute for Health and Clinical Excellence 2011 guidelines advise to use ABPM for patients with a clinic measurement of 140/90 mmHg to confirm a clinical diagnosis of hypertension.[34] Moreover, the prevalence of hypertension decreases (by almost one-half) in all populations when estimates are based on BP measurements taken from two or more visits compared to prevalence based on estimates that were taken on a single-visit.[166] Thus current approach taken in the most typical hypertension GWAS is more likely to overestimate the prevalence of hypertension. Thus, the interpretation of the results of binary

trait studies need to consider the underlying basis of the phenotypic definition used. However, the two studies that used a binary trait as a phenotype have reported two loci in plausible biological pathways, and relied on extensive phenotyping characterization in selecting the cases.[82,136] For instance, the BP-extreme study has reported the *UMOD* locus by genotyping individuals drawn from the extreme of BP distribution to allow for the maximum separation between cases and controls.[82]

Studying BP as a quantitative trait in GWAS can be compromised by the effect of BP-lowering medications, which may distort the physiological BP level, leading to substantial shrinkage in the statistical estimates (discussed in Section 1.5.1.1 -p51).[87] An appropriate adjustment was proposed by Tobin et al. by adding a fixed value to the observed BP in treated subjects, another method was proposed by Rena et al. using a refined approach by adding values based on antihypertensive drug class and ethnicity.[87,88] In both methods, the accuracy of the adjustment relies on the source of subject's medication history, which is typically obtained from either questionnaires or medical records. For studies with large sample size, it may be easier to collect this information using a questionnaire than obtaining the medical records for participants, due to inability to get an access for patients medical records. Hence, the accuracy of adjustment procedure would largely depend on the reliability of information given by the participants. Therefore, phenotypic complexity of BP represents a major challenge for GWAS, and further work and explorations are required to reduce the heterogeneity of BP phenotypes.

Genetic studies with a family-design offer a powerful alternative for gene discovery, as relatives are more likely to share both the genetic background and environmental factors. Hence, the analysis of phenotypes among family members is controlled to some extent for both genetic and environmental factors.[167] This is especially true for populations that are relatively static and stable, such as the Scottish population, which provides an ideal cohort to measure heritable and lifestyle factors for complex traits. The power of family studies in dissecting the genetic architecture of complex diseases relies on the availability of extra information that can be modelled to explain the error variance, leading to higher power to estimate the model parameters.[168] One of

the largest challenges for family-based studies is the difficulty in recruiting family members, which is time-consuming and requires more resources compared to unrelated individuals. However, the presence of a pre-existing large scale family-based cohort such as Generation Scotland: Scottish Family Health Study (GS:SGHS) can overcome this issue. Furthermore, the higher rate of CVD and hypertension in Scotland compared to other European populations highlight the uniqueness of the Scottish population and the possible roles of genetic and environmental factors.[169]

1.7 Study aims

The specific aims of this study are:

- Critically analyse BP measurements and medication history in GS:SFHS, to generate highly validated BP phenotypes for epidemiological and genetically analyses.

- Conduct a detailed epidemiological analysis of BP traits and hypertension.

- Study the familial aggregation of hypertension, treatment, and BP control.

- Estimate heritability of BP traits.

- Validate SNPs previously reported in large meta-analysis of GWA studies of BP traits in the Scottish population.

2 Materials and Methods

2.1 Introduction to this chapter

This chapter describes the overall materials and methods of the thesis, with the exception of the statistical analyses, which are described separately in the relevant chapters. It explains the processing of DNA samples from the time they were received from Generation Scotland (GS) in sample plates, the genotyping, and the methods applied to generate high quality data for downstream analysis. The aim of the methods described in this chapter was to generate a homogenous sample with high quality data (i.e. phenotype and genotype) that are suitable for all subsequent analysis.

As the GS:SFHS cohort was utilised for this study, GS has played a key role in providing the DNA samples and linking the genotypes with phenotypes. Figure 2-1 depicts the overall approach of communication and the contribution of GS management team to this work. Communication was performed in two steps to maintain participant anonymity and confidentiality, as only GS management had the appropriate access to link DNA information to participant's clinical data.

The chapter starts by describing the GS:SFHS cohort and the participant recruitment process. It then provides a detailed explanation of the data collection process, including how BP was measured, and the methods to assess exposure to BP-lowering medications. Finally, it describes the method of SNPs selection for genotyping, the genotyping platforms, procedure and QC.

2.2 Generation Scotland

GS is a multi-institutional, cross-disciplinary collaboration between the Scottish University Medical Schools and the National Health Service (NHS) in Scotland. The Scottish Government funded the project in 2003 to promote research into genetics and healthcare. GS includes three biomedical resources for study of common complex disease: the Scottish Family Health Study (GS:SFSH), Genetic Health in the 21st Century (GS:21CGH), and the Donor DNA Databank (GS:3D). The number of participants recruited in these three projects is over 30,000.[170] Participants were recruited from across Scotland between February 2006 and March 2011. Although the main recruitment is completed, the resources

continue to grow with every new use of the cohort. The full details of GS resources are available on the study website (www.generationscotland.org).

The major study of GS is GS:SFHS, which is an extended family-structured, population-based, intensively phenotyped cohort study. The study has biological samples, socio-demographic information, and clinical data for approximately 24,000 participants, aged between 18 and 98 years, recruited from across Scotland. The GS:SFHS protocol was published in 2006, and the full cohort profile description was published in 2013.[171,172] GS:SFHS is characterized by the family-based recruitment procedure through grouping individuals into family units, and it is considered to be one of the largest family-based genetic epidemiology studies. The cohort includes a wide spectrum of ages, lifestyle, and demography. It includes breadth and depth of phenotype information that allows population-based genetic and epidemiological research on several important diseases and risks. Furthermore, participants have given consent to anonymously link their data with NHS datasets such as prescribing records, hospital attendance, cancer and death registration.[171,172] The collected data are a combination of clinical measurements taken by trained staff, and self-reported data obtained from a pre-clinic questionnaire (PCQ) that was completed by each participant.

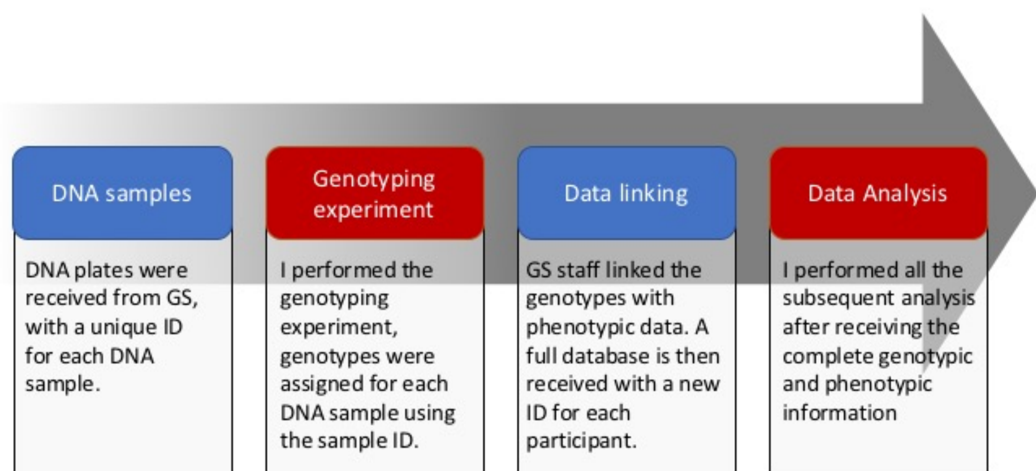


Figure 2-1 The overall strategy of the study.

2.2.1 Participants

The participant recruitment strategy is divided into two phases based on the geographic location and recruitment time. The first phase of the study covers the period between the start of the study in 2006 and 2008. Potential participants, or probands, during this phase were recruited from the Glasgow and Tayside areas if they were aged between 35 and 65 years; the term “proband” here refers to the first person who invited other relatives into the study and does not imply affected individual, as participants were not ascertained via any particular disease. In the second phase (2009-2011), the study was extended to include Ayrshire, Arran, and North East Scotland, along with broadening the age range of the probands to between 18 and 65 years. In both phases, probands were recruited at random from lists of collaborating general medical practices, and were asked to provide details of at least one first-degree relative who was over 18 years old and would also participate in the study, and so on, to create a “snowball” sampling effect. This sampling method is approximate to general population family sample, since more than 96% of the UK population are registered at general medical practices. In addition to recruitment by invitation, volunteers were also recruited if they were over 18 years with at least one first-degree relative. Throughout the study period, the methods of recruitment, identification, and approach were performed under an extensive public consultation exercise.[173,174]

The total number of invited probands was 126,000. Of that group, only 12.3% responded and met the study criteria. Not all of the responders were recruited for practical reasons such as inability to give informed consent, or recruit another family member. The overall successful response rate was 5.3% (6665 probands), plus 1,288 volunteered participants without an invitation, and 16,007 relatives, giving a sample size of 23,960 individuals. The number of participants who attended the research clinics to complete and provide blood samples for DNA extraction was 21,476. The remaining participants who were unable to attend the clinic sent their PCQ with a saliva sample, for DNA extraction, by post (n =2,484). All participants signed the consent form and received the participant information leaflet prior to their enrolment.

In comparison with the Scottish population, the GS:SFHS sample was generally healthier, wealthier (39% lived in areas with above average Scottish-index of multiple deprivation [SIMD]), older, and included higher proportion of women (59%) (Table 2-1).[171] Most of the participants were born in the UK or Ireland (96%), with the vast majority born in Scotland (87%). Although the sample cannot be considered a truly representative of the Scottish population, it includes a wide range of socio-demographic and clinical features. In addition, the large sample size increased the proportion of participants from all socio-economic classes, with many or multiple disease traits.

Table 2-1 Comparison between GS:SFHS cohort and the Scottish population

	GS:SFHS	Scottish population
Median age (years)		
Males (years)	47	37
Females (years)	48	39
Gender (% Male)	41	48
Ethnicity (% White)	99	98
Employment (those aged up to 75 years) (%)		
Unemployed	1.7	4[a]
Retired	15.1	12.9[a]
Employed	62.8	58[a]
Education (%)		
Degree	33	20
No qualifications	5	33
Overweight or obese (BMI >25) (%)		
Males	65	68
Females	53	61
Current smokers (%)		
Males	19	25
Females	16	25
Alcohol intake (mean units/week)		
Males	15.8	17.5
Females	7.1	7.8
Hypertension (%)		
Males	37.4 (Measured)[b] 14 (Self-reported)	35[c]
Females	24.7 (Measured)[b] 13.2 (Self-reported)	31[c]
Heart disease (%)	Total 3.9 (Self-reported)	
Males	5.6	7.3
Females	2.7	5.5
Diabetes (%)	Total 3.3 (self-reported)	
Males	4.1	5.7
Females	2.8	4.3
Stroke	Total 1.5 (Self-reported)	
Males	1.7	2.7
Females	1.3	2.2

[a] People aged 16-74 years
[b] Systolic >140 or diastolic >90.
[c] Treated or untreated hypertension.
Table is reproduced with modification from [171]

2.3 Phenotypic data

Multiple measures and clinical tests were performed at the clinic for all the participants by trained clinical staff, according to rigorous standard operating procedures. All participants signed “broad” consent forms before any data collection, in addition to having an appropriate discussion about the future of their participation in the study. Participants were also asked to provide consent for linkage of their data and samples to routine datasets, such as NHS prescription records. Although GS:SFHS has a wide range of phenotype information, this thesis only focused on data related to BP. Hence, the related measurements and clinical assessments that were of interest to this thesis include BP measurements, anthropometric and demographic measures, treatment history provided by either PCQ or electronic-prescription records (EPRs), and health and lifestyle information obtained from the PCQ.

2.3.1 Blood pressure phenotypes

Blood pressure was measured twice, consecutively, with a three minutes interval, using Omron HEM-7051T digital BP monitor. The participants were asked to sit quietly for five minutes before the first reading. The readings were recorded in the clinical record form (available at http://www.generationscotland.org), along with the time at which the readings were taken. PP was calculated as the difference between SBP and DBP. MAP was calculated as $^1/_3$ of SBP plus $^2/_3$ of DBP. The BP phenotypes that were analysed in this thesis are either the observed BP values or the adjusted BP values. The observed BP values are calculated as the average of the two readings, for SBP and DBP. Likewise, the observed PP and MAP were calculated based on the values of the observed SBP and DBP. The adjusted BP values are the observed BP values plus a fixed value if individual was on BP-lowering medication when BP was measured. The rationale of adding a fixed value to the observed BP values for treated participants is explained in the next Section (2.3.1.1).

2.3.1.1 Adjustment for BP-lowering medications

Studying BP as a quantitative trait is a powerful method to examine the genetic and environmental factors, as it can overcome the issue of inconsistent diagnostic criteria for hypertension. However, a preliminary analysis of

prescriptions records for 8,238 individuals in GS:SFHS showed that about 12% of them were taking BP-lowering medications.[171] Using the observed BP value without any adjustment for treated individuals can lead to substantial shrinkage in the estimated effect of the aetiological factors and reduce statistical power.[87,175] This is because the outcome of primary interest, i.e. BP level without the medication effect, cannot be directly observed. Therefore, the observed BP measurements in treated individuals represent a biased distortion in quantitative analysis. Yet, excluding treated subjects from the analysis wastes important information regarding the familial components of BP variance, and reduces the effective sample size.[175] Hence, the contribution of the treated subject is very important in the study, and can be lost if they were excluded, or obscured by the treatment effect if no adjustment was taking during the analysis.

The size of the added fixed values is based on the recommendation of Tobin et al. by adding 15 mmHg to the observed SBP, and 10 mmHg to the observed DBP to account for the treatment effect.[87] This method of adjustment for blood pressure-lowering medication was applied in several BP genetic studies such as ICBP.[107,135,152] It was also shown that this method of adjustment restores the familial components of BP variance, and the adjusted BP values more closely reflects BP values without treatment effects.[175,176]

2.3.2 Drug exposure

In order to adjust for the BP-lowering medication effect, participants taking these medications first needed to be identified. GS:SFHS has two sources that can be used to search for medication history; first, EPRs obtained by linking the participant's information to the NHS prescription information system database. Second, self-reported medications (SRMs) history that can be extracted from the PCQ. The vast majority of the participants completed the current medication history Section in PCQ, EPRs were also available for smaller number of participants. Based on these two sources of treatment- exposure history, the participants were classified into two groups; first group included individuals who had both types of medication history data, and second group included individuals with medication history based on one medication history only. An additional complexity in the SRMs data occurred as the SRMs history was

differently obtained during the two phases of the study (detailed explanation in Section 2.3.2.3 - p101). The next section explains the methods to identify individual who were taking BP-lowering medication, based on each medication history sources.

2.3.2.1 Definition of BP-lowering medications

The British National Formulary (BNF 58) was used to create a database of medications classifications and indications.[177] This database included 1694 items in the form of medication's approved name, BNF code, BNF description, and pharmacological class. All of the medications indicated for hypertension in the BNF were considered as BP-lowering medications, and individuals taking any of these medications were eligible for the adjustment of the observed BP value by addition of the fixed value.

2.3.2.2 Electronic-prescription records (EPRs)

Using EPRs as a source to assess drug exposure can avoid the potential of recall bias that may occur in SRMs.[178,179] However, important factors need to be considered carefully to avoid misclassification of participants. The first factor is to differentiate between participants with missing prescription records, and participants that were not in the prescription record because they have not been prescribed any medication. In other words, to be sure that missing individuals are such because they have not been prescribed any medication and not because their data were not available. To address this factor, a list of "eligible" participants was generated to identify the ID of participants with prescription records (i.e. whether they have been prescribed any medication not just antihypertensive drugs). The eligibility criteria were based on the recruitment area and date, that is all Tayside participants (area code T), and Glasgow participants (area code G) recruited from 2008 onwards. Individuals from other areas or recruited from Glasgow before 2008 were set as missing for prescription records because of the incomplete prescription data.

The second factor is the definition of drug exposure, which is based on the type of medication and the time window. The time window refers to the period prior to a reference date that is reviewed to search for prescription of BP-lowering medication. The recruitment date was used as the reference date because it

was the day when BP was measured and each participant completed the PCQ. Several time windows were considered in studies that investigate the accuracy of prescription record as a source to measure drug exposure; it ranged from 30 days to one year prior to the reference day.[179-183] Although the 90 days time window has shown high sensitivity for medication of chronic disease, this thesis used a time window of 120 days prior to the day of recruitment.[181,183] This was to have a less stringent period, and choose a value that lay between the two most common time windows (i.e. 90 days and 6 months). Also, a preliminary analysis showed that increasing the time window beyond 120 days did not demonstrate any additional improvement in concordance with the SRMs. Participants were then coded as “1” or “0” if they were taking a BP-lowering medications or not, respectively; this value was named as ($Treatment^{prescr}$), which indicates if the individual was taking a BP-lowering medication based on EPRs.

2.3.2.3 Self-reported medications (SRMs) history

The PCQ form was sent to the participants to be completed at home, and participants were required to bring with them during the clinic visit along with any regular medications. The participants completed the PCQ before performing any medical assessment, with help of the research nurse to clarify any queries. The PCQ included questions related to demographic details, occupational history, lifestyle, personal and family medical history, pain, and current medication history. The PCQ form was slightly amended in 2009, dividing the study into two phases, where the period up to this revision was termed phase 1 (n =9,016), and the period thereafter was termed phase 2 (n =11,305). One aspect of this revision was changing the question structure from open-ended questions to closed questions, in particular the part about current medication history. The participants were asked to write “Name of Prescribed or Bought Pills or other Oral Medication” in PCQ-1, and to tick the appropriate Yes or No box if they “regularly taking any blood pressure lowering medication” in PCQ-2. In order to offset this alteration in the structure of question, each PCQ-phase was analysed separately to identify participants who were taking BP-lowering medication. Subjects who have written a BP-lowering medication name in PCQ-1 or answered “Yes” in PCQ-2 were coded as “1”, or “0” otherwise for the field ($Treatment^{Ques}$); separate coding was constructed based on each PCQ. The two

PCQ forms can be accessed from the GS website (http://www.generationscotland.org). The following section explains the method of analysis in each PCQ.

2.3.2.4 PCQ-1

The current medication history section in this form contained four questions about names of the regular medications based on their dosage forms (i.e. oral, cream, inhaler, and injection). Participants were also asked to tick a box if they were not taking any medications. Because participants manually wrote their medication names, entries were inconsistent for medications across individuals. For instance, the name of the same medication can be written in several forms such as a generic, trade, medication name plus the strength, or a combination of these names. Also, the entries were not free from spelling errors or irrelevant symbols. Hence, the written names were first formatted in the same style to improve the accuracy of the retrieved information. This was performed by checking each participant answer to be reformatted into the correspondence medication approved name; for unclear answers, an attempt to correct for the written medication names was undertaken when it possible. Although the vast majority of the participant entries were retrieved, a few answers were ignored as they were impossible to be guessed. The result of this review was a new database that contained the participants ID and their answers to the current medication history, but in the form of the medication's approved name. This database was then linked to the same database of BNF medication in Section 2.3.2.1, and subjects taking BP-lowering medication were identified based on the same criteria. Participants who have completed the PCQ-1 and were identified to be receiving a BP-lowering medication were coded as “1”, or “0” otherwise for the field ($Treatment^{PCQ-1}$).

2.3.2.5 PCQ-2

The current medication history question has changed in this phase to be a closed question (i.e. yes/no). This has made the analysis straightforward as each subject is coded as either “1” if answered “Yes”, or “0” if answered “No” to the question of taking a BP-lowering medication. Yet, some participants have left the question unanswered and they were coded as missing “-9”. Participants who

have completed the PCQ-2 and were identified to be receiving a BP-lowering medication were coded as “1”, or “0” otherwise for the field (Treatment^{PCQ-2}).

2.3.3 Anthropometric measures

The anthropometric measurements obtained at the recruitment visit were height, weight, waist, hip, and body fat composition. These measurements were used to calculate BMI, as weight in kilograms (kg) divided by height in meter squared (m^2).

2.3.4 Social and demographic data

Socio-demographic data for each participant were retrieved from the PCQ. Age and sex of each individual were obtained during the clinic visit, along with information about place of birth and ethnicity background. Furthermore, participants answered the same demographic questions about their parents. In addition, participants reported their address postcode, which was then linked to the SIMD 2009 report.[184] SIMD is a measure of deprivation that ranks Scotland’s area into data zones based on seven domains and indicators, which are income, employment, health, education, access to service, housing, and crime.[184] The data zones are then ranked based on their score on these seven domains. SIMD quintile classifies the postcodes into five quintiles, each containing 20% of the data zones. Areas in the first quintile (SIMD =1) are the most deprived area in Scotland, and areas in the fifth quintile (SIMD =5) are the least deprived in Scotland.

2.3.5 Family health history

The two forms of PCQ contain questions regarding previous clinical diagnoses of certain conditions including high blood pressure of the participants, their father, mother, brothers, sisters, and grandparents. The question in PCQ-1 combined brother and sister in one single question, meaning that the participants would tick the same box if they have a “brother or sister” with hypertension. This was slightly different in PCQ-2, where the answer for brother and sister were collected separately in two boxes. Also, a question about number of brothers and sisters were only available in PCQ-2. To overcome this discrepancy, I created a new field named as “sibling” that combined the information from the

two forms, though indicating no sibling with hypertension in PCQ-1 cannot be distinguished from the case where an individual has no sibling.

2.3.6 Phenotype Quality Control

Clinical and questionnaire data were only retrieved for individuals with quality checked genotypes, which meant that individuals without genetic information or those who did not pass genotype QC were excluded before this step. The phenotype QC aimed to check the completeness of clinical data for the genotyped individuals, in particular BP values, anthropometric measurements, and medication history, for which the procedure of collection was previously explained. As the availability of these three components is critical for all the analysis, any individual without any of these components was excluded from the study. Thus, the study sample will be homogenous with regards to the availability of genotypic and phenotypic data. The importance of these three components is summarized as follows: first, BP measurement is essential to do the quantitative analysis, as well as the qualitative analysis in which subjects are dichotomized into a binary trait (i.e. hypertensive or not) based on a specific definition. Second, presence of anthropometric measurements is necessary to calculate BMI and is used as a covariate in most of the subsequent analysis. Third, ascertainment of current medication history is essential to assess if the observed BP is influenced by BP-lowering medication or truly reflects the underlying BP (that is, the BP that treated subjects would have if they were not taking BP-lowering medication). Individuals were not excluded if they had other missing information, and will be reported as missing. Figure 2-2 shows the flowchart for the steps of genotype and phenotype QC, starting from receiving the DNA until completing of the QC.

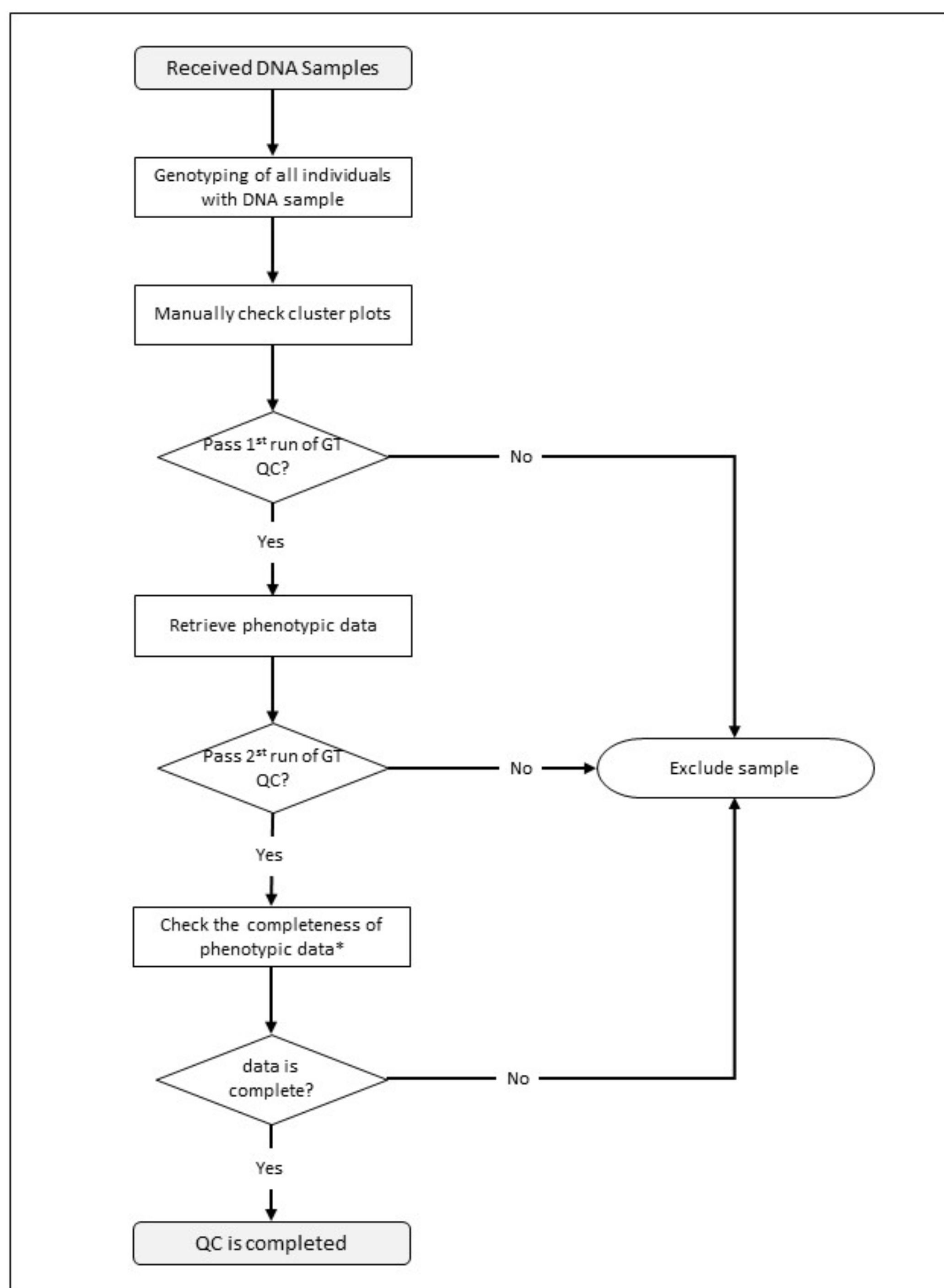


Figure 2-2 Flowchart of the QC procedure.
*** Subjects with BP readings, BMI measurements, and medication history were considered as having a complete phenotypic data.**

2.4 Genotypic data

2.4.1 Extraction and storage of DNA

The process of collection and storing of DNA for all the samples was performed using standard operating procedures and managed through a laboratory information management system. Blood samples were obtained in the research clinic from each consenting participant using standard venipuncture procedures and collected in a 9 ml Ethylenediaminetetraacetic acid (EDTA) tube. DNA extraction was performed using a Nucleon Kit on 9 ml of blood sample (Tepnel Life Science) with the BACC3 protocol. The precipitated DNA was hooked out and placed directly into a labelled 2.0 ml microtube (Scientific Specialties Inc) containing 1 ml TE buffer pH 7.5 (10 mM Tris-Cl pH 7.5, 1 mM EDTA pH 8.0). For postal participants and participants with insufficient blood sample, DNA was extracted from the provided saliva sample using an Oragene OG-250 saliva kit into similar microtubes by a standard protocol (DNA Genotek). Microtubes were incubated at room temperature for two weeks with rotation until DNA was fully re-suspended. To assess the quality of DNA extraction, 8 samples of every batch of 92 samples were electrophoresed on 1% agarose gel and quantified by NanoDrop (Thermo Scientific) to confirm DNA yield. The Picogreen method (Invitrogen) was finally used to quantify the DNA concentration in ng/µl, and 500 µl of each DNA master stock was transferred to a deep well plate and then normalized to 50 ng/µl to make working stock plates. The DNA extraction procedure was performed at the Centre for Molecular Medicine, University of Edinburgh, by the GS laboratory staff.

For this study, DNA samples were transferred in 384-well PCR plates to the British Heart Foundation-Glasgow Cardiovascular Research Centre, and stored at -4°C until genotyping. Plates were numbered from 1 to 55, and matched with Plates map sheet that was provided by GS. The Plates map sheet contains information about the ID sample of each well in each plate in the form of; Res ID, CRF Barcode, destination plate ID, and destination well. This plate map sheet was used to link the genotype information with individual phenotype data, and to track the sample information of each well. The plates numbered from 1 to 54 contained 380 DNA samples in the concentration of 50 ng/µl, and 4 empty wells to be used as no-template controls (wells number A1, A24, P1, and P24).

Plate number 55 contained only 233 DNA samples, and the remaining wells were all empty, which gave a total of 20,753 samples that were available for genotyping.

2.4.2 Genotyping procedure

2.4.2.1 Genotyping platform

The genotyping procedure was performed using a TaqMan® OpenArray® Genotyping System (Applied Biosciences), which provides an automated and rapid qualitative detection of bi-allelic SNP. The assay relies on a fluorescence-based polymerase chain reaction (PCR) that uses the 5` nuclease activity of the AmpliTaq Gold® DNA polymerase to cleave a perfectly matched probe and emit a signal (Figure 2-3). Each assay consists of a normal PCR primers that flank the target polymorphic region, in which two allele-specific TaqMan® probes are hybridized.[185,186] The fluorescence signals for each intact probe is quenched by the physical proximity of the reporter and quencher dyes through fluorescence resonance energy transfer. During the PCR extension, the AmpliTaq Gold® DNA polymerase extends the normal PCR primers until reaching the TaqMan® probes, in which it can only cleave the hybridized probe (i.e. matched), and release the reporter dye. On the other hand, the mismatched probe remains intact and shows no fluorescence.[186-188] The TaqMan® probes incorporate 3`-minor groove binder (MGB) technology that enhances the stability of the formed duplex DNA when probes are hybridized.[185] This is achieved by increasing the differences in the melting temperature (Tm) between the perfectly matched and mismatched probes, which leads to improvement in genotyping accuracy. The higher Tm of the MGB probes is achieved without increasing probe length, which allows for designing of shorter probes that are more sensitive to a single base mismatch.

The TaqMan® OpenArray® plate contains 3,072 microscopic through-holes (divided into subarrays of 8 x 8) that are coated with a hydrophilic material from inside, and can accommodate a reaction volume of 33 nl.[189] The dimensions of plates are similar to the conventional microscope slides, and the number of loaded samples per plate depends on the number of assays in the plate, which can range from 144 samples in the 16-plex assays format to only 12 samples in the 256-plex Assay format (Figure 2-4). The experiments in this thesis were

performed using the 64-plex assays, with which up to 48 samples can be loaded on a single OpenArray® Genotyping plate.

The advantage of this system is that it only requires a single enzymatic reaction with a simple operating procedure. The experiment can be completed within eight hours with TaqMan® assay reliability and a projected overall call rate of 99%.[189] The assay materials include TaqMan® assays, primers, and probes were designed by Applied Biosystem based on SNPs information that were selected for this thesis.

2.4.2.2 Genotyping workflow

The overall procedure of the experiment involved transferring the DNA sample to the TaqMan® OpenArray® 384-Well Sample Plate, and then to the TaqMan® OpenArray® Genotyping Plate (OpenArray plate), which is then placed in the thermal cycler before scanning the OpenArray plates to call the genotypes by using the appropriate software (Figure 2-5). The complete procedure that was followed during the genotype experiment was as following:

1. Preparing the Sample Information File (*.csv) for sample integration
    a. Before starting the experiment, the sample information file was created that contains information about the sample ID of each well in each plate. Each one of the 55, 384-Well DNA Sample Plates has a unique sample information file, which is created based on the information provided by GS. This sample information file is used in the sample integration process that is performed when sample is loaded into the TaqMan® OpenArray® Genotyping by the OpenArray® Accufill™ System, as discussed in next steps.
2. Loading DNA samples and master mix into the TaqMan® OpenArray® 384-well Plate
    a. The 384-well PCR plate that contains the DNA samples was thawed to room temperature, and then centrifuged at 1600 rpm for 1 minute.
    b. 2.5 µl of the TaqMan® OpenArray® Master Mix was added to the TaqMan OpenArray 384-well sample plate.

c. 2.5 µl of the DNA sample from the 384-well PCR plates was transferred to the same position in the TaqMan OpenArray 384-well sample plate, with gently mixing by pipetting up and down.
d. 2.5 µl of DNase-free, sterile-filtered water was added to the no-template wells.
e. The TaqMan OpenArray 384-well sample plate was covered with the provided sealing tape, and then centrifuged for 1 minute at 1500 rpm to eliminate bubbles.

3. Prepare the TaqMan® OpenArray® Genotyping plate
    a. The genotyping plate is stored at -20 °C until the day of the experiment.
    b. The required amount of genotyping plates are removed from the freezer, and thawed at room temperature for 5 minutes.
4. Sample plate preparation for loading into OpenArray Genotyping plate using OpenArray® Accufill™ System
    a. Prepare the TaqMan® OpenArray® Accessories Kit, which contains the genotyping case, sealing glue, and immersion fluid.
    b. Place the genotyping case in the case rack and fill the case approximately $^{3}/_{4}$ with immersion fluid.
    c. Place the OpenArray AccuFill Loader Tips, and Plate Holder at their assigned locations.
    d. Initialize the Accufill™ system for the loading operation.
    e. To enable the sample integration option, the sample information file (*.csv) and the plate setup file (*.spf) were imported into the appropriate fields. The plate setup files were entered in the same order as their physical position in the plate holder. This file contains information on assay location within the OpenArray plate. The output of sample integration is the assignment of all samples to their respective assays in one new updated plate file.
    f. Hold the OpenArray Plate by the edges, at the end opposite from the barcode, and place it into the Plate Holder.
    g. Start the loading process
    h. After completion of the loading, the software create a newly plate setup file that start by the prefix "Loaded_[plate serial number].spf". This file is then copied into a USB drive to process

the OpenArray® Genotyping plates in the OpenArray® NT Cycler during the imaging step.

5. Seal the TaqMan OpenArray® Genotyping plates
    a. Slide the OpenArray Genotyping plate into the genotyping case, guided by the grooves in the case.
    b. Use the sealing glue to seal the genotyping case by filling the top of the case.
    c. Cure the glue by placing the genotyping case in the sealing station for 90 seconds.
6. Perform thermal cycling using the Block GeneAmp® PCR System 9700
    a. The sealed OpenArray genotyping plates were then loaded into the thermal cycler, by keeping the barcode face-up.
    b. The thermal cycler can accommodates up to 8 OpenArray plates, and the run takes about 4 hours using this profile
        i. 10 minutes at 93°C.
        ii. 45 seconds at 95 °C.
        iii. 13 seconds at 94 °C.
        iv. 2:14 minutes at 53 °C.
        v. Repeat steps (ii to iv) for 50 cycles.
        vi. Hold at 4°C forever.
7. Image of the OpenArray® Genotyping plate by OpenArray® NT Imager
    a. Start the OpenArray® NT Imager and wait for the system to fully boot up.
    b. Import the newly created plate setup file that corresponds to the OpenArray® Genotyping plate position in the machine, and repeat for position 2 and 3.
    c. When the imaging is completed, a new folder is created that contains all the related files;
        i. SNP plate data (*.spd) file for downstream analysis in TaqMan® Genotyper Software V1.3. This file is copied into the USP drive for later analysis and QC.
        ii. Four images that are important to check the quality of each Genotyping plate; all the four images were carefully checked in each run.
8. The experiment for each plate was tracked using the designed form.

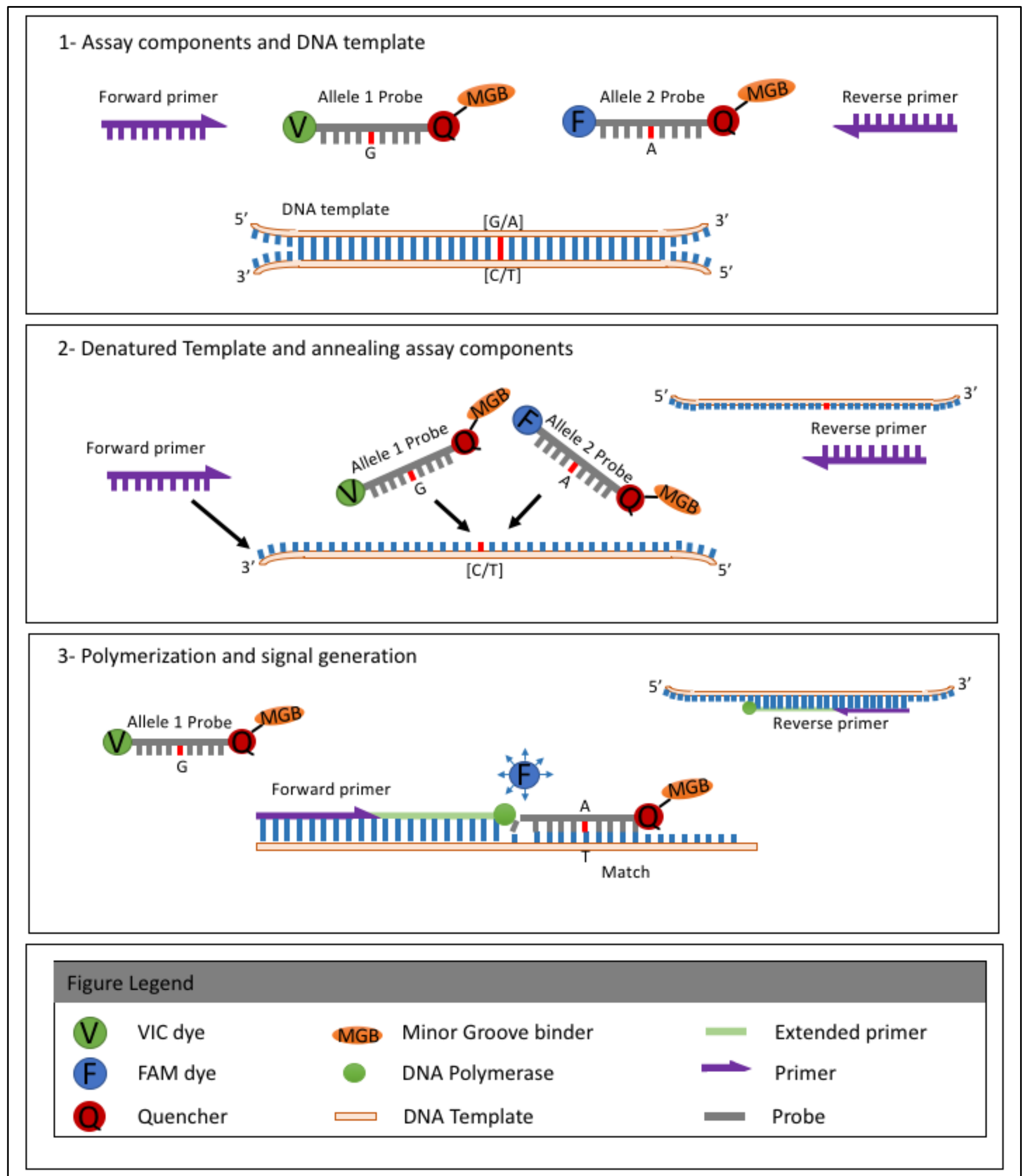


Figure 2-3 Dynamics of allelic discrimination in the TaqMan® genotyping assay.
Each allele has a specific probe that carries a nonfluorescent quencher (NFQ) at the 3′ end, and either VIC (green) dye, or FAM (blue) dye at the 5` end. During PCR primer extension, each probe anneals specifically to its complementary sequence between the forward and reverse primer sites. A DNA polymerase that possesses 5` exonuclease activity is used in the assay to cleave the hybridized probe that is perfectly matched, freeing the reporter dye from the NFQ and allowing it to fluorescent signal in the absence of its NFQ. However, the mismatch probe is not hybridized and remains intact with no florescence. In this figure, a substantial increase in VIC dye indicates homozygote for Allele 1, and a substantial increase in FAM dye indicates homozygote for Allele 2, and an equal amount on VIC and FAM dyes indicates heterozygotes.

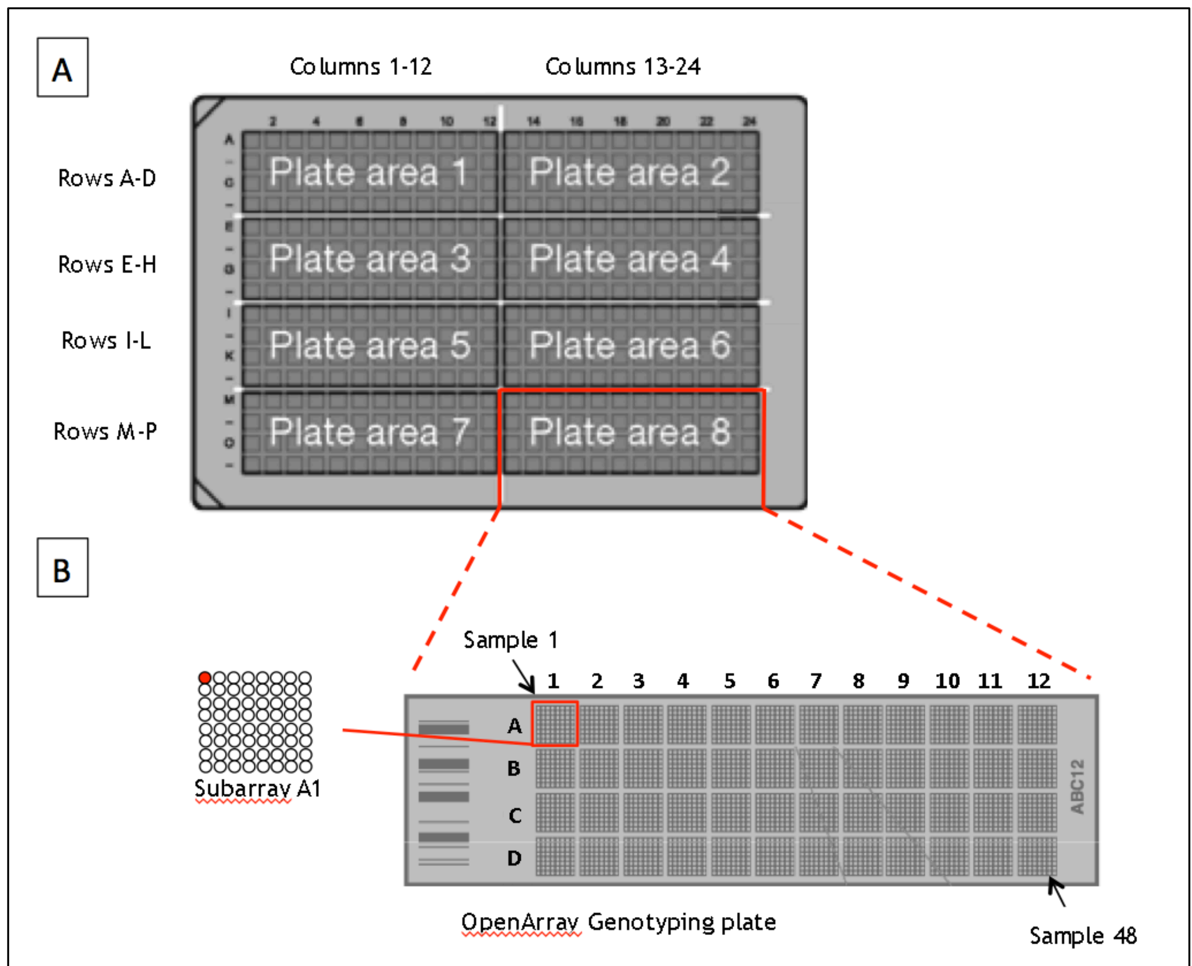


Figure 2-4 Layout of the OpenArray 384-Well Sample Plate, and the OpenArray Genotyping Plate.
(A) The 384-Well sample plate is used to mix the DNA with the Master Mix before using the AutoLoader system to transfer the mixture into the genotyping plate. It is divided into eight areas; each area has 48 wells (12 x 4). During the loading process, the AutoLoader transfers samples from one area into the Genotyping plate. (B) The OpenArray Genotyping Plate has 3,072 through-holes divided into smaller groups of 8 x 8 subarrays (48 subarrays). The subarrays are addressed by a capital letter represents the row (A-D), and a number represents the column (1-12). Within subarrays, each through-hole is addressed by a small letter (a-h) to represent the row, and number (1-8) to represent the column. For instance, the through-hole filled with red can be addressed as A1a1. For the 64-plex format, each subarray contains one sample only.

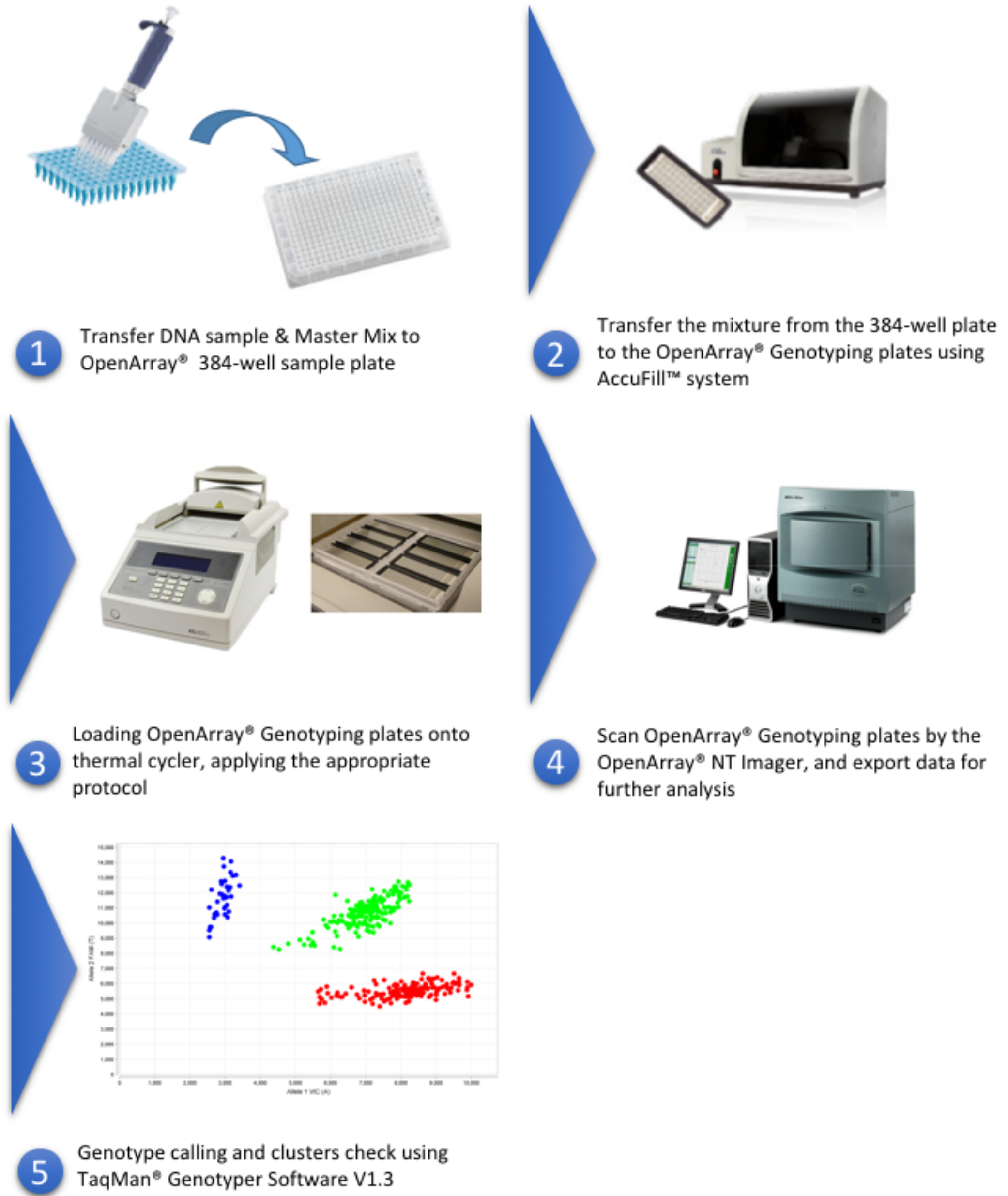


Figure 2-5 Genotyping experiment workflow.

2.4.2.3 Genotype calling

Genotype calling was performed using the TaqMan® Genotyper Software V1.3 (Life Technologies), which uses the generated raw data from the OpenArray® NT Imager system in the form of (*.spd) file. The genotype-calling algorithm is based on the clustering strategies that combine sample with similar dye fluorescence intensities at each SNP to one cluster. The program then assigns a discreet genotype to each cluster according to its position on the plot. Samples are then clustered into three clusters that vary along the X-axis (homozygosity for Allele 1), Y-axis (homozygosity for Allele 2), or diagonal (heterozygosity). One feature of the TaqMan® Genotyper Software is its ability to overlay and analyse raw data from several genotyping plates together, to increase the confidence of the genotype call.[190] Therefore, it is recommended to combined between 6 to 8 genotypes experiments in one *study*, a *study* is the technical name for the collection of the raw data from several OpenArray® Genotyping plates. For the convenient of sample tracking, each 8 OpenArray® Genotypes plates that originated from the same DNA sample plate were pooled into one *study*. The cluster plots for each *study* of each SNP were then manually checked to exclude samples with low genotype QC, as indicated by a call rate lower than 80%. In addition, a manual genotype calling procedure was carried out on any SNP that showed clear clustering to one of the three genotype clusters, but was not automatically called by the program. The procedure of excluding samples with very poor genotyping quality can be considered as the first step of genotype QC. The output across all plates was exported to the appropriate format to be linked with phenotype data by GS management staff, before commencing the complete genotype QC.

2.4.2.4 Genotype quality control

The genotype QC was performed using PLINK software after merging all the samples in one file.[191] All the samples were pooled together to do more stringent QC based on a total of 60 SNPs, including 16 SNPs that shared the same OpenArray Genotype plate but were related to a different project. Sample QC was performed first to exclude samples with a genotype call rate ≤95% (using PLINK's function --mind 0.05); samples with very large proportion of failed SNP assays may be indicative of a poor quality DNA sample, which could lead to

aberrant genotype calling. Based on this call rate threshold, any sample with more than three missing genotypes was excluded. The next step was to remove families with Mendelian errors ≥5% (using PLINK's function --me 0.05 0.01); that is to exclude family with high proportion of Mendelian inconsistencies, which indicate genotyping errors if pedigree information is correct. Checking of Mendelian error is only possible in samples of related individuals, with pedigree information. This procedure aims to check that genotype transmission for each SNP follows Mendelian inheritance laws. For instance, two homozygous parents with a genotype of AA for a SNP cannot have an offspring with heterozygous or homozygous for the other allele (i.e. AB or BB), and this scenario will be highlighted as a Mendelian error in the QC. The aim of the this step is to exclude any family with three or more Mendelian errors, erroneous genotypes in individuals with less than three Mendelian errors were set as missing genotype (using PLINK's function --set-me-missing 1 1). A second run of checking of genotypes for Mendelian inconsistency was performed using the program PEDSTATS (Version 0.6.12; sph.umich.edu/csg/abecasis/PedStats; Goncalo Abecasis; Abecasis Lab, Center for Statistical Genetics, School of Public Health, University of Michigan, Ann Arbor, Michigan).[192] This step will be described in more details in the next Section (2.4.2.5), as it was followed by imputation of the missing genotypes.

The QC for SNPs started by checking the SNP genotyping efficiency (the proportion of samples with a genotype call for each marker), to exclude SNPs with a high proportion of missing individuals; a threshold of 90% of call rate per SNP was applied in this step (using PLINK's function --geno 0.1). The next step was to check the MAF using the function (using PLINK's function --freq), and exclude any SNP with MAF below 1% (using PLINK's function –maf 0.01), as statistical power is extremely low for rare SNPs and they are more prone to error.[122,190,193] The final step in SNPs QC was to check the Hardy-Weinberg Equilibrium (HWE), as departure from this equilibrium may result from genotyping errors or population stratification.[194] In PLINK, the HWE statistics is only calculated in founders (i.e. individuals for whom the paternal and maternal individuals are coded 0). In other word, the test considers the parents only and ignores the offspring to have an estimate that is free from bias, which might be introduced by the presence of correlated genotypes in one nuclear family. PLINK

outputs the HWE exact test p-value of the null hypothesis that genotype frequency follows the HWE expected proportions (i.e. p^2, $2pq$, q^2); against the alternative hypothesis that observed genotypic proportions are significantly different from the expected. A threshold of p-value <0.001 was set to indicate a departure from HWE.

2.4.2.5 Pedigree based Imputation

After completing the genotype QC, one more step was performed to impute the missing genotypes in a subset of individuals, who have a maximum of three missing genotypes or a maximum of three genotypes set to missing due to Mendelian inconsistency. Two approaches were used to impute the missing genotypes. The first approach was applied for individuals with genotyped relatives (i.e. parents or offspring) by using the genotype inference feature in MERLIN software (Version 1.1.2; sph.umich.edu/csg/abecasis/Merlin/index.html; Goncalo Abecasis; Abecasis Lab, Center for Statistical Genetics, School of Public Health, University of Michigan, Ann Arbor, Michigan).[195] In this method, PEDSTATS was initially used to check the pedigree information and confirm that all genotypes follow Mendelian law of inheritance. The next step was to use the genotype inference parameter (--infer) that is integrated in MERLIN to estimate the missing genotypes. The output of this analysis is a table that contains; the most likely genotype, the expected number of copies for the tested allele (0, 1, or 2 if the genotyped is not missing; or a fractional count if the genotype is missing), and the posterior probabilities for the three alternative genotypes. The genotype with highest posterior probabilities was imputed instead of the missing genotype. The second approach was applied for individuals without a genotyped relative, for whom the genotype inference feature in MERLIN cannot be used, and the missing genotypes were imputed using the average coded allele frequency.

2.5 SNP Selection

Two parallel methods were performed to identify SNPs to be assayed in the chosen genotype platform. In the first method, the National Human Genome Research Institute (NHGRI) catalogue of Published GWAS (updated as of July 2012) was searched.[196,197] The GWAS catalogue provides a publicly available

collection of SNPs identified by means of GWAS and catalogued based SNP-trait association with p value less than $1.0x10^{-05}$.[196] The second approach was done by expanding the search method by snowballing from the identified GWAS, to include any possible studies that were not represented in the GWAS catalogue.The search was restricted to SNPs reported from studies of European ancestry, searching for these key words: “SBP”, “DBP”, “MAP”, “PP”, “hypertension”, or “BP”. The selected SNPs were entered in the SNAP pairwise LD analysis to test independent SNPs that are not in LD.[198] Similarly, SNAP tool was used to identify proxy SNPs in complete LD with the selected SNP, to be used if no predesigned TaqMan assay is available for the selected SNP by Life Technologies. In total, 44 SNPs in 39 loci were selected for assay in the genotyping platform (Table 2-2)

2.6 Ethical approval

All components of GS:SFHS have received ethical approval from the NHS Tayside Committee on Medical Research Ethics (REC Reference Number: 05/S1401/89). GS:SFHS has been granted Research Tissue Bank status by the Tayside Committee on Medical Research Ethics (REC Reference Number: 10/S1402/20) providing generic ethical approval for a wide range of uses within medical research.

Table 2-2 List of selected SNPs for genotyping

No.	GS SNP	Locus	Main Ref.	Other studies reported the same SNP	Other Studies reported the same locus
1	**rs17367504**	**1p36.22**	107	133, 155	133, 155, 147, 142, 150
2	**rs5068**		134		
3	**rs17030613**	**1p13.2**	137	133	
4	**rs2932538**		107		
5	**rs2004776**	**1q42.2**	150,149		
6	**rs1446468**	**2q24.3**	135		137
7	**rs13082711**	**3p24.1**	107		
8	**rs3774372**	**3q22.1**	107		115
9	**rs319690**	**3p21.31**	135,137		
10	**rs419076**	**3q26.2**	107		
11	**rs871606**	**4q12**	135		
12	**rs1458038**	**4q21.2**	107		133, 137
13	**rs13107325**	**4q24**	107		
14	**rs13139571**	**4q32.1**	107		
15	**rs1173771**	**5p13.3**	107		150, 137
16	**rs11953630**	**5q33.2**	107		
17	**rs1799945**	**6p22.2**	107	150	
18	**rs805303**	**6p21.33**	107		
19	**rs12705390**	**7q22.3**	135		
20	**rs3918226**	**7q36.1**	136	150	
21	**rs2071518**	**8p23.1**	135		
22	**rs4373814**	**10p12.3**	107		
23	**rs1813353**		107		155, 115
24	**rs4590817**	**10q21.2**	107		
25	**rs1530440**		133	155	
26	**rs932764**	**10q23.33**	107		
27	**rs11191548**	**10q24.3**	107		133, 155, 147, 137, 115
28	**rs2782980**	**10q25.3**	135		
29	**rs661348**	**11p15.5**	150		
30	**rs7129220**	**11p15.4**	107		199
31	**rs381815**	**11p15.1**	107	155,115	137, 150
32	**rs633185**	**11q22.1**	107		
33	**rs11222084**	**11q24.3**	135		
34	**rs17249754**	**12q21.3**	107		115, 147, 150, 155
35	**rs3184504**	**12q24.1**	107	155, 115	
36	**rs653178**		133	155	
37	**rs10850411**	**12q24.2**	107		137, 115
38	**rs1378942**	**15q24.1**	107	133, 155	115
39	**rs2521501**	**15q26.1**	107		
40	**rs13333226**	**16p12.3**	82		
41	**rs12940887**	**17q21.33**	107		133, 155
42	**rs17608766**	**17q21.32**	107		
43	**rs1327235**	**20p12.2**	107		
44	**rs6015450**	**20q13.32**	107		

3 Quality control of GS:SFHS

3.1 Introduction about this chapter

This chapter describes the data cleaning procedure applied in this thesis to ensure high quality data were used in the subsequent analyses. This procedure started with performing a QC step for genotyping data, and then explored the primary variables in this study (i.e. BP traits). Each participant was checked for the availability of high quality genotypes, phenotypes, and medication history to be included in the final study population.

3.2 Method

The general methods used in this chapter are described in detail in Chapter 2. Briefly, inclusion and exclusion criteria for individuals in the QC procedure are based on the availability of three components: genotype information, clinical data, and medication history (Figure 2-2). Each one of these three components has its own inclusion/exclusion criteria. Briefly, individual's exclusion criteria based on genotypes were; sample call rate ≤95%, and individuals from family with Mendelian errors ≥5% (see Section 2.4.2.4 -p102). The clinical data were then checked for completeness in individuals with complete genotypes. This was first accomplished by checking the availability of BP and BMI measurements; individuals without these values were excluded as well as those without medication history (see Sections 2.3.2 -p99, and Section 2.3.3 -p103). Analysis was performed only on participants who passed these criteria, to produce a sample with a complete genotype and phenotype information. The procedure of SNP QC started by checking the genotype call rate, MAF, and Mendelian errors as described previously in Section 2.4.2.4 -p114 and Section 2.4.2.5 -p116.

Exploratory data analysis for BP values was performed first by generating graphical summaries for each variable. Extreme values were then double checked to distinguish between those values that are genuine extreme value (i.e. outliers) fromincorrect values. Standard methods to remove outliers are based on considering any values above multiples of standard deviation (i.e. 3 SD) as outliers. However, this thesis used the method reported by Welch et al. to remove outliers based on sensible boundaries defined by representative survey data.[200] In this method, acceptable age- and gender-specific ranges of each variable were derived by adding/subtracting 10% to the most extreme values in

the Health Survey for England (HSE). A table for these ranges was provided by Welch et al. for both variables of SBP and DBP.[200] Graphical summaries and data exploratory analysis were performed using SPSS (version 22.0; SPSS Inc. Chicago, Illinois, USA) and ggplot2 package in R statistical package version 3.1.1.[201]

3.3 Results

3.3.1 Excluding individuals based on genotype call rate

The total number of individuals with a DNA sample and available for genotyping was 20,753 individuals; of them, only 18,470 individuals passed the QC criteria and were available for subsequent analysis (Figure 3-1). The first run of QC was performed manually by checking the cluster plots, and then checking the genotype call rate for all the individuals, in which 456 individuals were excluded due to low genotype call rate. The second run of QC excluded a further of 1,261 individuals based on the predefined exclusion criteria; this resulted in a total of 19,027 individuals with high quality genotype information.

3.3.2 Excluding individuals based on missing phenotypes

The total number of individuals with genotypes (n =19,027) were then checked for availability of phenotypes (i.e. BP and BMI), and 146 (0.7%) individuals were excluded because of missing BP measurement or BMI values (Figure 3-1).

3.3.3 Excluding individuals based on medication history

The remaining individuals were then checked for the availability of medication history data in the form of either an EPRs or SRM. Of 18,881 individuals, a total of 411 (2.7%) individuals were excluded because of missing medication history (Figure 3-1). This resulted in a total of 18,470 individuals with BP measurements, BMI values, and at least one source for medication history. The total number of participants with EPRs was 13,732 individuals. The total number of participants who completed PCQ-1 was 8,253 individuals, of them 3,225 individuals ticked the box of not taking any medication pills. The total number of participants who completed PCQ-2 was 9,918 individuals, of them 1,086 individuals did not answer the question regarding taking BP-lowering medication

(i.e. not answering yes or no), which resulted in a total of 8,832 individuals with a valid answer for PCQ-2.

The participants were then classified based on the number of medication history sources; the first group included individuals with at least one source of medication history (n =18,470). The second group included individuals with two sources of medication history (i.e. subjects with EPRs and SRMs in PCQ; n =12,347).

3.3.4 QC for SNP

All SNPs have MAF ≥0.05, except for rs17608766 with a MAF of 0.00691. All SNPs were in HWE (P >0.0001) based on founders only, and no SNPs were removed because of low call rate (<90%) (Table 3-1). Furthermore, GS management staff performed an independent validation of a subset of the genotypes. This was accomplished by comparing the generated genotypes in this project against another project that genotyped a similar 19 SNPs for 590 individuals common to both the two projects, providing a total of 11,210 genotypes to be cross-validated. The discrepancy in genotypes between the two projects was reported to be less than 1%.

Imputation was performed based on the family information when possible, or based on the average coded allele frequency for persons without other relatives. For the 44 SNPs related to this thesis, 13,997 individuals had complete genotype data without requiring any genotype imputation. Genotype imputation in the remaining individuals ranged from only one SNP for 3,340 individuals to six SNPs for 74 individuals. The proportion of the total imputed genotypes to the chip genotypes was less than 2%.

3.3.5 Family data review

The family structure of the participants composed of 7,025 extended families, which range from 1 member per family in 2,396 families to 29 members per family in a single pedigree. The average family size was 2.63 members per family, and the average number of generations per family was 1.91 (Figure 3-2).

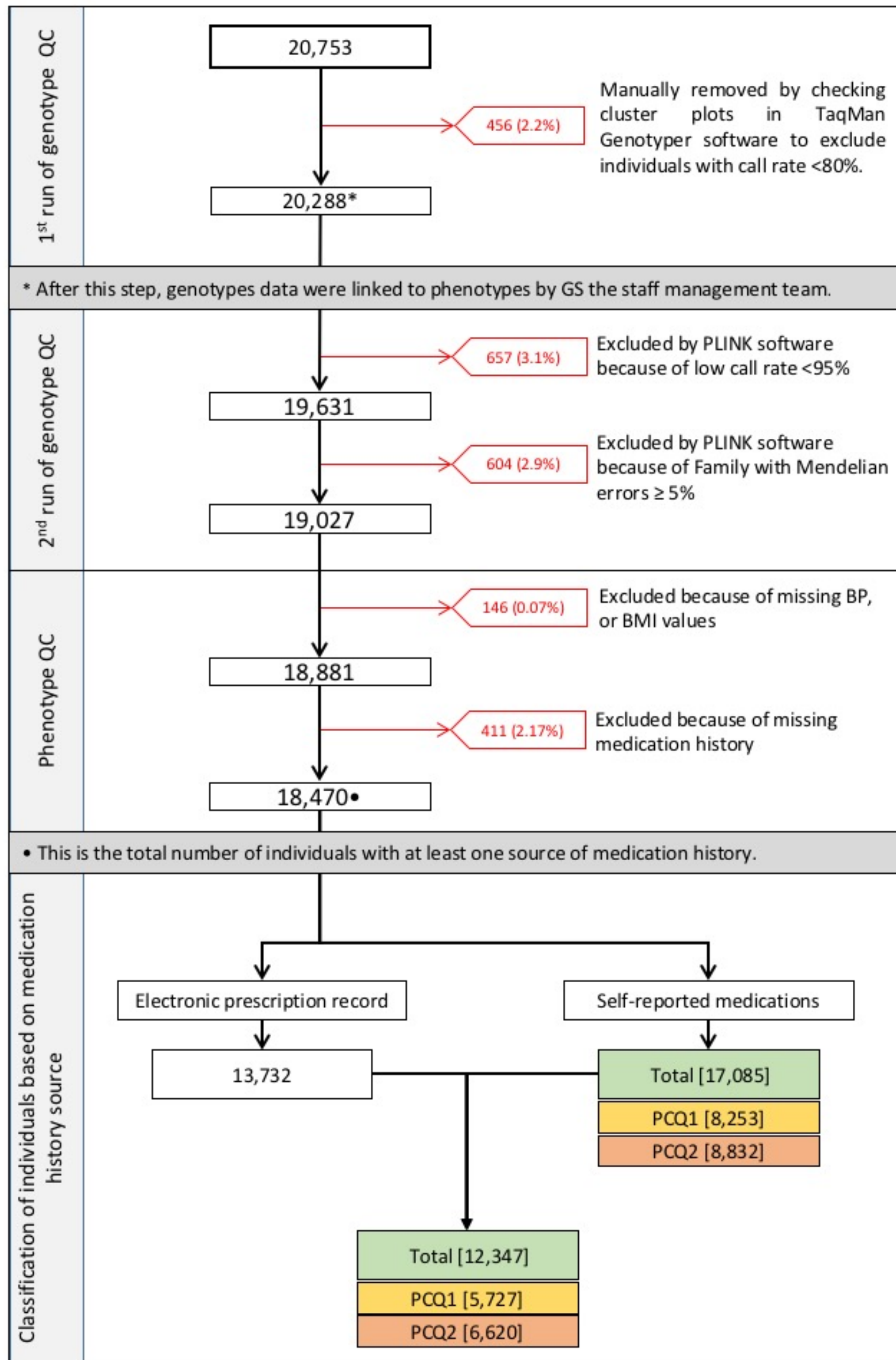


Figure 3-1 Flowchart of QC results and classification of individuals based on the sources of medication history.

Table 3-1 Summary statistics of SNPs QC

No	Chr.	SNP	Minor allele	Major allele	MAF	Call rate
1	**1**	rs17367504	G	A	16.2%	99.6%
2	**1**	rs5068	G	A	5.05%	99.8%
3	**1**	rs17030613	C	A	19.9%	99.5%
4	**1**	rs2932538	A	G	24.3%	99.3%
5	**1**	rs2004776	T	C	26.4%	99.3%
6	**2**	rs1446468	A	G	46.6%	98.1%
7	**3**	rs13082711	C	T	25.2%	99.5%
8	**3**	rs3774372	C	T	15.0%	99.8%
9	**3**	rs319690	G	A	32.0%	95.6%
10	**3**	rs419076	T	C	48.8%	99.5%
11	**4**	rs871606	C	T	10.9%	99.8%
12	**4**	rs1458038	T	C	29.1%	98.1%
13	**4**	rs13107325	T	C	5.55%	94.0%
14	**4**	rs13139571	A	C	24.6%	99.3%
15	**5**	rs1173771	A	G	39.0%	99.3%
16	**5**	rs11953630	T	C	36.6%	99.1%
17	**6**	rs1799945	G	C	15.1%	98.7%
18	**6**	rs805303	A	G	38.1%	99.6%
19	**7**	rs12705390	A	G	19.8%	99.8%
20	**7**	rs3918226	T	C	8.41%	99.1%
21	**8**	rs2071518	T	C	28.4%	99.6%
22	**10**	rs4373814	C	G	41.7%	99.6%
23	**10**	rs1813353	G	A	33.5%	99.4%
24	**10**	rs4590817	C	G	17.9%	99.6%
25	**10**	rs1530440	T	C	19.6%	99.4%
26	**10**	rs932764	G	A	44.5%	99.5%
27	**10**	rs11191548	C	T	8.19%	99.8%
28	**10**	rs2782980	T	C	27.6%	99.1%
29	**11**	rs661348	C	T	41.1%	99.4%
30	**11**	rs7129220	A	G	12.3%	99.8%
31	**11**	rs381815	T	C	29.6%	99.6%
32	**11**	rs633185	G	C	27.8%	99.6%
33	**11**	rs11222084	T	A	36.9%	99.6%
34	**12**	rs17249754	A	G	17.3%	99.5%
35	**12**	rs3184504	C	T	50.0%	99.2%
36	**12**	rs653178	C	T	50.0%	99.5%
37	**12**	rs10850411	C	T	31.6%	98.6%
38	**15**	rs1378942	C	A	32.1%	99.4%
39	**15**	rs2521501	T	A	31.2%	99.7%
40	**16**	rs13333226	G	A	18.7%	99.4%
41	**17**	**rs17608766**	**C**	**T**	**0.69%**	**98.6%**
42	**17**	rs12940887	T	C	35.7%	99.4%
43	**20**	rs1327235	G	A	49.4%	99.5%
44	**20**	rs6015450	G	A	12.0%	99.2%

Abbreviations: Chr.: Chromosome, MAF: Minor allele frequency, HWE: Hardy-Weinberg equilibrium. SNP rs17608766 (**in bold**) was excluded because of low MAF <1%.

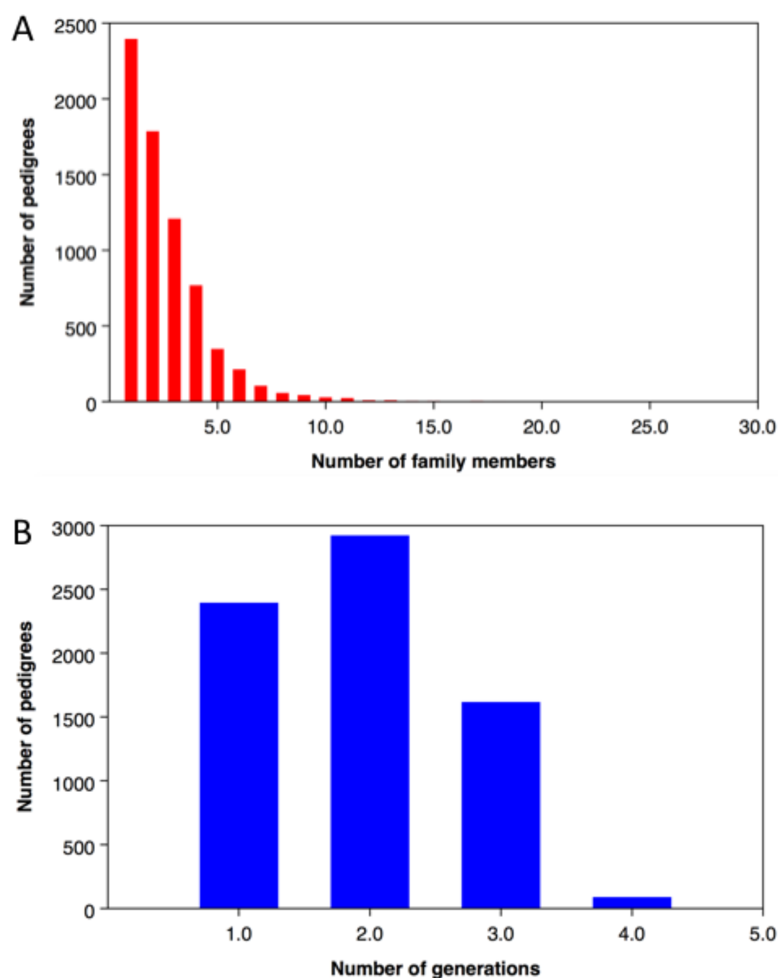


Figure 3-2 Family size and number of generations per pedigree in GS:SFHS.
A. shows the number of individuals per family (i.e. family size). Families are groups in the study sharing a family identity number; a family identity number was given to groups where each member was a first-degree relative of at least one other individual in the group. B. shows the number of generation per pedigree, this figure of generation structure may include family members that are not in the final study population, but their presence is required to draw the pedigree (figures were produced using Pedstats Software).

3.3.6 BP values data check

The quality of BP measurements was checked to identify any outliers or erroneous entries by two different methods. First, a data exploratory step was performed by drawing a graphical summary for each BP variable as shown in Figure 3-3 and Figure 3-4. Second, the age- and gender-specific ranges of BP values were compared to a reference ranges provided in HSE,[200] in which all values of BP in GS:SFHS were within the range (Table 3-2). Hence, all the measurements were considered valid values without any value labelled as an outlier. The departure from normality for BP traits was minimal except for PP, as measured by skewness (0.7, 0.3, 0.4, and 1.1 for SBP, DBP, MAP, and PP, respectively) and kurtosis (0.9, 0.3, 0.3, and 2.2 for SBP, DBP, MAP, and PP, respectively).

3.3.7 Identifying participants taking BP-lowering medications based on EPRs

Of the individuals who passed QC (n =18,470), the EPRs contained 1,507,927 NHS prescriptions, for 13,732 subjects, including 4,319 items (trade names) or 1,333 pharmacological items (approved names). The total number of prescriptions that contained a BP-lowering medication was 208,408 prescriptions for a total of 91 pharmacological items, for 4,787 individuals. However, this number of individuals encompassed the total number of individuals who were taking a BP-lowering medication at any time point before the reference data (i.e. recruitment day). The definition of BP-lowering medication exposure was restricted to 120 days before the reference date. Hence, searching of prescriptions within this period revealed 7,550 BP-lowering medication prescriptions for 1,965 individuals (detailed description of BP-lowering medication is discussed later in Section 4.4.2 -p143).

3.3.8 Identifying participants taking BP-lowering medications based on SRMs

The total number of participants with SRMs history was 17,085 individuals; of them, 8,253 individuals completed the PCQ-1, and 8,832 completed the PCQ-2. While the retrieved information from PCQ-2 was straightforward and simple as subjects answered yes/no questions, more information was retrieved from

PCQ-1 such as medication name and pharmacological classes. In PCQ-1, the number of written medications names by participants was 20,715 entries; of them, 2,426 medications were identified as BP-lowering medication. The number of participants who reported taking a BP-lowering medication was 1,334 individuals (16.16%); 701 (52%), 449 (33.6%), 148 (11.1%), 33 (2.5%), 2 (0.2%), and 1 (0.07%) participants reported one, two, three, four, five, and six BP-lowering medications, respectively. The most common reported monotherapy were ACEIs 184 (26.2%), beta-blockers (BBs) 182 (26%), calcium-channel blockers (CCBs) 118 (16.8%), and thiazides diuretics 87 (12.4%). In PCQ-2, out of the 8,832 individuals who completed the PCQ-2, 1,517 (17.2%) had positively answered the question of taking a regular BP-lowering medication.

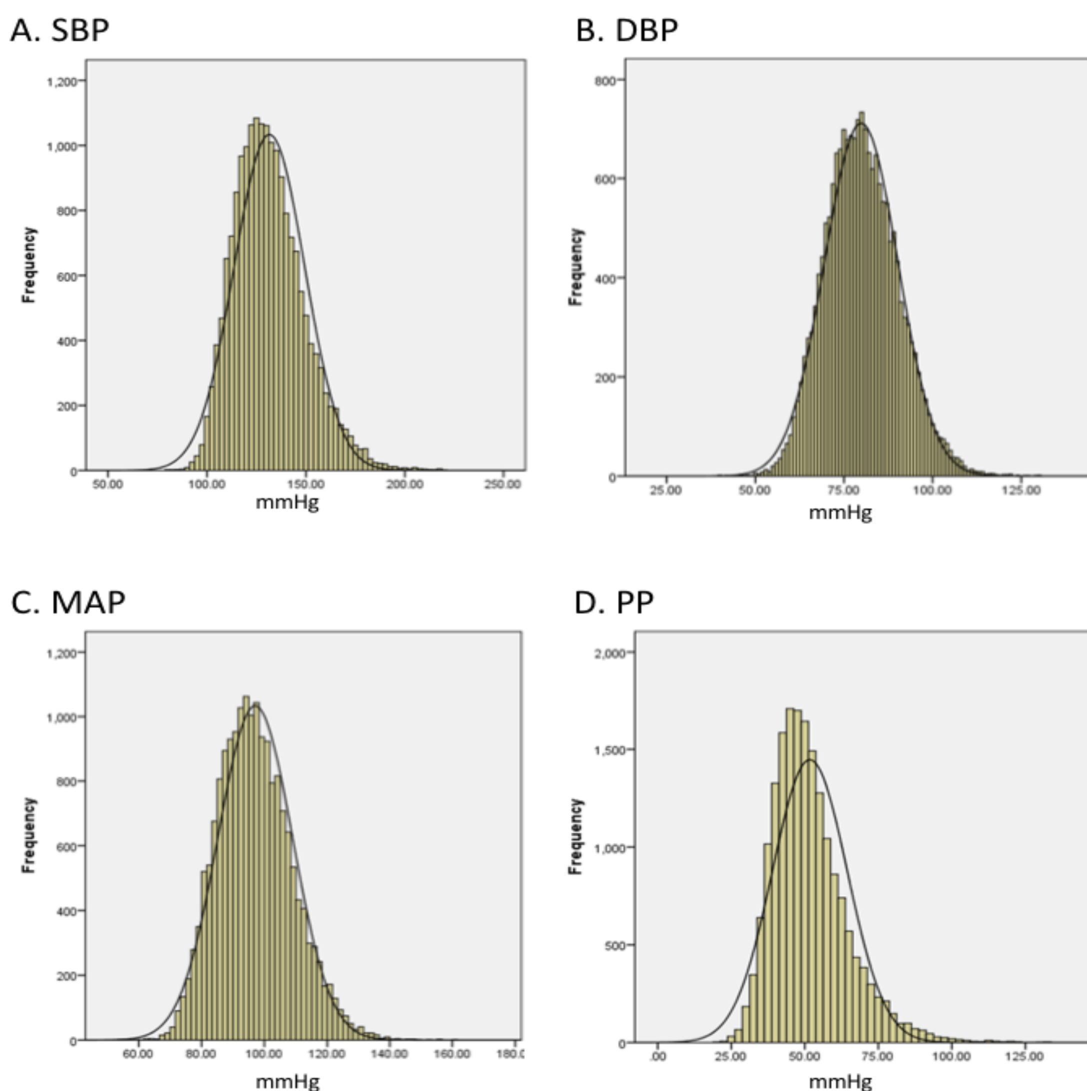


Figure 3-3 Histogram for BP traits with normal distribution curve on the plot.
The vertical axis (y-axis) shows the frequency of individuals, and the horizontal axis (x-axis) shows BP values in mmHg. Histograms represent A. SBP, B. DBP, C. MAP, and D. PP.

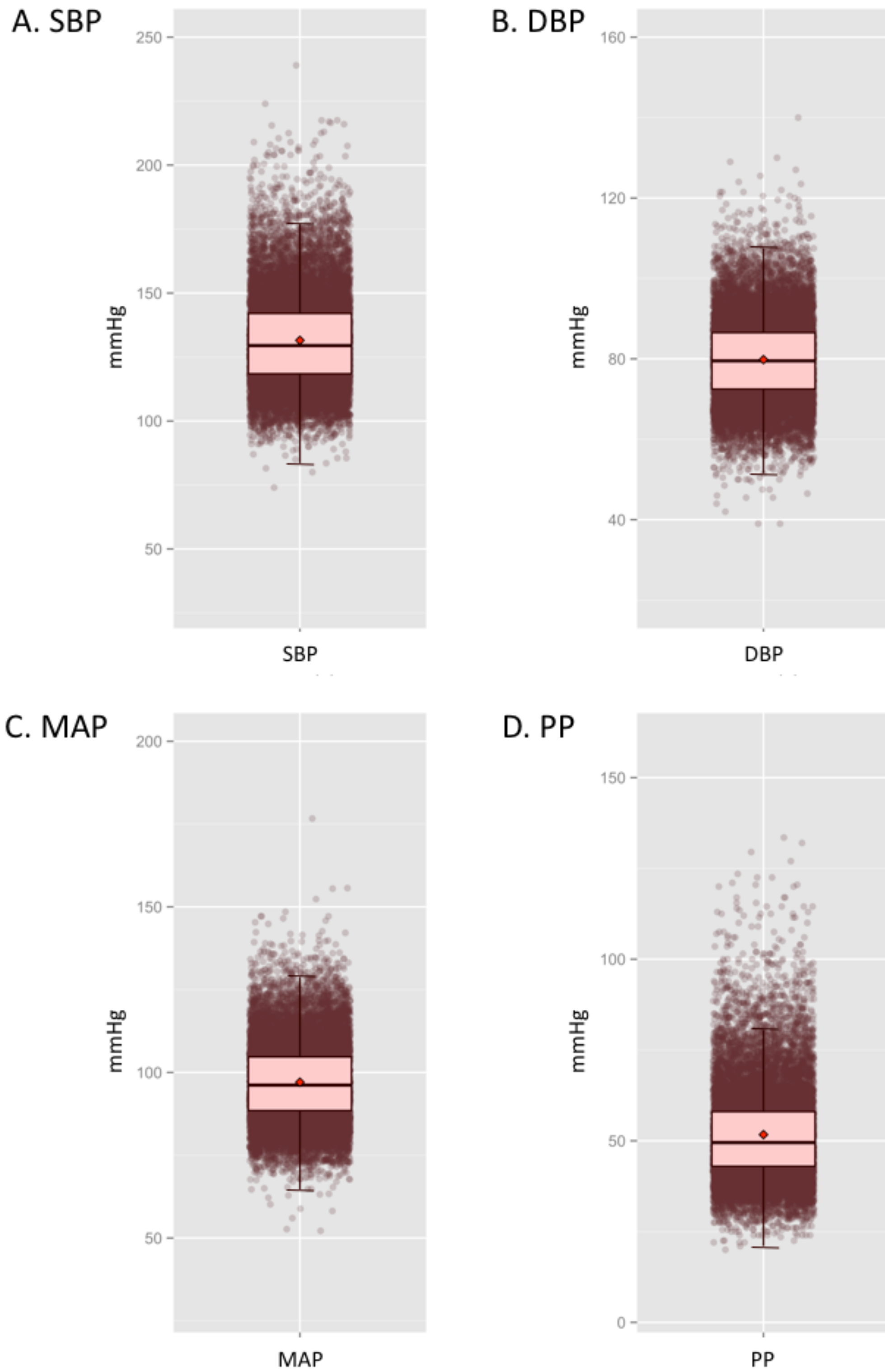


Figure 3-4 Box-Whisker plot with jitter-dots for BP traits.
Red diamond inside the box plot represents the mean. The red diamonds inside the box plots represents the mean, and the lines within the boxes indicate the median, the edges of the pink boxes are the first and third quartiles (the interquartile range or (IQR)), the whiskers extend to 1.5 times the IQR. BP values above/below the boxplot are plotted in with small jitter-dots to show data distribution. Boxplots represent A. SBP, B. DBP, C. MAP, and D. PP.

Table 3-2 Age- and gender-specific ranges of GS:SFHS and HSE

Variable	Age group (years)	Men				Women			
		GS:SFHS		HSE (+/- 10%)*		GS:SFHS		HSE (+/- 10%)*	
		Min	Max	Min	Max	Min	Max	Min	Max
SBP (mmHg)	17–24	85.5	187.5	68.4	207.9	80.0	143.5	52.2	149.7
	25–34	93.5	180.0	61.2	214.5	88.5	181.0	72.0	221.1
	35–44	85.0	188.5	55.8	218.9	87.0	188.5	64.8	235.0
	45–54	98.5	197.0	55.8	242.0	74.0	239.0	54.9	269.5
	55–64	86.5	216.5	63.0	256.3	81.5	210.5	72.0	256.3
	65–74	94.5	213.0	72.0	253.0	97.5	217.0	54.9	255.2
	75+	83.5	224.0	78.3	249.7	100.5	217.5	61.2	265.1
DBP (mmHg)	17–24	35.5	107.5	29.7	128.7	44.0	105.5	27.9	127.6
	25–34	52.0	112.5	36.0	139.7	50.5	114.0	30.6	125.4
	35–44	46.0	122.0	29.7	141.9	36.0	119.5	35.1	144.1
	45–54	55.0	124.0	27.9	159.5	42.0	145.5	27.9	163.9
	55–64	50.0	125.5	28.8	157.3	48.5	130.0	35.1	167.2
	65–74	45.5	121.5	35.1	154.0	55.5	120.0	34.2	165.0
	75+	46.5	104.5	33.3	158.4	50.0	107.5	34.2	161.7

*age- and gender-specific ranges of SBP and DBP from HSE were derived from data reported in Welch et al.[200] HSE: Health Survey of England

3.4 Discussion

This chapter described the data exploratory and QC processes used in this thesis. This data exploration is a critical first step in analysing data to detect any possible incorrect, implausible, missing, or duplicate values to ensure high data quality. The availability of a large health record database such as GS:SFHS offers a unique resource for epidemiological and medical research. However, primary variables need to be carefully examined before conducting analysis to exclude errors that may potentially affect the interpretation of the results. For instance, a study that compared the error rate generated by two independent investigators reported a rate that ranged from 2.3% up to 26.9%, with errors clustering in non-random fashion.[202] Hence, several guidelines have reported the importance of data cleaning in good clinical practices.

Quality of phenotypes, DNA samples, genotyping, and family structure of GS:SFHS has already been examined in other studies.[171,203] The quality of phenotypes and clinical data of the complete cohort of GS:SFHS (n =23,960) was published by Smith et al.[171] Several reasons can explain the difference in the sample size between this thesis and the complete cohort reported by Smith et al. First, this thesis originally started with a smaller sample size due to the availability of DNA samples. Second, the final study population was restricted to individuals with specific inclusion criteria (i.e. BP and BMI measurements along with medication history). Third, the full cohort has reported presence of 1,400 singletons due to recruiting participants with registered relatives who then failed to participate in the study. This number increased to 2,396 singletons in this thesis due to exclusion of extra participants from the study.

While Smith et al. have mainly examined the phenotype data quality, Kerr et al. has examined the DNA, genotyping, and pedigree quality, by genotyping 32 SNPs in DNA from 10,450 individuals.[203] SNPs genotyping call rate ranged from 91.6% to 99.5%, with an average sample call rate of 96.2% for DNA samples derived from saliva, and an average sample call rate of 97.1% for DNA samples derived from blood. The quality of genotyping in this thesis has also been assessed by an independent investigator, who compared the genotypes of 19 SNPs for 590 individuals that were generated by this project to the same set of SNPs and

individuals in another project. This has provided 11,210 genotyped to be cross-validated and the discrepancy was found to be less than 1%.

The primary aim of this thesis is to study whether the genetic variants that have been previously reported to be associated with BP and hypertension also play a role in the Scottish population. Hence, the first inclusion criterion was the presence of SNP genotypes. This would allow a consistent sample size across the different analyses performed in this thesis. The proportion of individuals excluded due to genotyping errors was about 6% compared to only 2% due to phenotype missing information, which is mainly due to the applied stringent genotype QC criteria. Mendelian errors are more likely to be a result of genotype errors rather than errors in the recorded pedigree, as reported by Kerr et al. after checking the inheritance patterns of 32 SNPs in 925 parent-child trios in GS:SFHS.[203] Considering the total of 925 trios, only 16 trios had two or more Mendelian errors, and 13 of them had offspring-paternal Mendelian errors, indicating a maximum estimated non-paternity rate in GS:SFHS of less than 1.5% (13 trios out of 925 trios).[203] The other possible explanations of Mendelian errors include sample handling errors in the clinic or laboratory, or errors in pedigree data collection, especially that family structures were recorded during the participant clinic visit without an independent verification from another source.

A higher proportion of individuals were excluded because of missing medication history (2.1%) than those excluded because of missing phenotypes (0.7%). The importance of the presence of a medication history relies on the fact that studying BP as a quantitative trait is seriously compromised without proper adjustment for treatments (see Section 1.5.1.1 -p51).[87] Hence, adjustment for treatment can only be performed for subjects with a reliable and complete medication history, and including individual with vague treatment status would introduce inconsistency between individuals, which may affect the precision of the phenotype and thence statistical power. At this stage, the source of medication history (i.e. EPRs or SRMs) did not influence the decision of including individuals, and each participant with at least one source of medication history was included.

The only reason for excluding individuals based on phenotype quality was missing BP measurements in about of 0.7%. BP measurements in all of the remaining participants were all considered within the age- and gender-specific ranges provided by Welch et al.[200] This method was found to be more efficient than the traditional method of removing values above/below 3 SD, as it is based on sensible boundaries as defined by a representative population, while keeping individuals with genuine extreme values that may represent the true variation in the study. The distribution of BP traits were normally distributed with a slight skewness to the right, especially for PP.

In conclusion, this chapter aimed to ensure the availability of high quality data for each participant in the final study population. Hereafter, the term of GS:SFHS will be used to describe the final study population that includes only participants with validated phenotypes and genotypes. This chapter did not expand to describe the baseline characteristics of participants.

4 Study of BP phenotypes

4.1 Introduction

4.1.1 Introduction to this chapter

This chapter describes the baseline characteristics of the study population. It then focuses on BP-related phenotypes, starting with the assessment of the prescription patterns of BP-lowering medications and the quality of SRMs history by comparing it with the information from EPRs. Next, the prevalence of hypertension, treatment, control, and awareness in GS:SFHS were calculated. Lastly, the influence of different risk factors on hypertension-related traits was evaluated using a logistic mixed model that accounts for the familial correlation.

4.1.2 Pharmacoepidemiological analysis

Self-reported methods are prone to various types of bias, such as recall bias, that may influence recall accuracy. Several studies have reported different factors that influence medication recall accuracy in self-reported methods such as medication type, pattern of drug use, questionnaire design, respondent characteristics, and the study method of analysis.[183,204,205] Typically, pharmacy prescription records are considered the "gold standard" of drug exposure information compared with self-reported information.[205,206]

As shown in Chapter 3 (Sections 3.3.7 -p126, and Section 3.3.8 -p126), participants of GS:SFHS can be divided into two groups based on the number of medication history sources. While two-thirds of participants (n =12,347) have both SRMs history and EPRs, the vast majority of remaining participants have only SRMs history. In order to combine information of medication history across all participants, it is important to determine the quality of SRMs in GS:SFHS by comparing it to the information retrieved from the EPRs. For this thesis, the analysis was restricted to assess the reliability of reporting BP-lowering medication by self-reported methods.

The availability of different sources of medication history in GS:SFHS has allowed different kinds of pharmacoepidemiological analysis to be performed. The aim of the pharmacoepidemiologic analysis is to make use of the EPRs and SRMs to understand the prescription patterns for BP-lowering medications in

GS:SFHS. Further analyses include assessment of the accuracy of SRMs history. Finally, the influence of changing the PCQ forms from open-ended questions in PCQ-1 to closed-end questions in PCQ-2 will be assessed by comparing each phase with the EPRs.

4.1.3 Prevalence of hypertension-related health outcomes in GS:SFHS

Sections 1.3.3 to 1.3.7 in the introduction chapter discussed the prevalence of hypertension and its related health outcomes globally and in the Scottish population. Historically, Scotland has been called the “sick man of Europe”, owing to the higher mortality rate compared to other western European countries. Although life expectancy in Scotland was comparable to other western countries up until 1950, it improved at a slower rate thereafter compared to other wealthy nations, before further faltering after 1980.[207] Basically, mortality rates from CVD, stroke and cancer were higher before 1980s, after which an increase in mortality rates from violent, drug-related and suicide deaths was observed. Relative to England, Scotland had 40% higher mortality rates for all causes, 60% higher mortality rates for coronary heart disease, and 42% higher mortality rate for stroke.[169] These higher mortality rates were observed across the socio-economic spectrum, suggesting a minimum contribution of socio-economic factors in explaining the mortality rates differences. Indeed, it was found that only a quarter of the excess mortality rates in Scotland could be attributed to socioeconomic, behavioural, anthropological or biological factors.[169] The remaining excess mortality rate that is not explained by differences in levels of socio-economic deprivation is known as the “Scottish Effect”.[208] These findings highlight the uniqueness of the Scottish population, and the importance of identifying the factors that contribute to the higher mortality rates compared to their English counterpart. In this section, the aims are to estimate the prevalence of hypertension in the GS:SFHS, and the extent of treatment, control, and awareness of hypertension. Also, the influence of some determinants in these hypertension-related outcomes will be examined.

4.2 Aims of this chapter

The aims of this chapter are to:

(1) describe the baseline characteristics of the study population.

(2) Perform a pharmacoepidemiological analysis to identify those individuals taking BP-lowering medications, and assess the reliability of SRMs history in exposure measurements of BP-lowering medication.

(3) Calculate the prevalence of hypertension, treatment, control, and awareness.

(4) Estimate the relationship of BP levels and prevalence of hypertension with socio-economic factors (as defined by SIMD quintiles).

4.3 Methods

4.3.1 Overview of the method

Since the EPRs were not available for all participants, the reliability of SRMs was performed only on the subset of individuals with this information. The SRMs source can be compared against the EPRs; in which, the EPRs were considered the "gold standard". Because hypertension is defined based on BP measurements and medication history, as will be discussed in Section 4.3.2, participants taking BP-lowering medications were firstly identified as explained in Section 2.3.2. The second analysis was carried out to determine the prevalence of hypertension, awareness, treatment, and control of hypertension; and finally using a multilevel multivariate mixed model that accounts for familial correlation to assess the effect of socio-economic factors in hypertension-related indicators.

4.3.2 Assessing the validity of the SRMs

The agreement between each source was evaluated using several methods, each of which addresses a different question regarding the concordance between the sources.[205,209-211] Generally, comparing two methods of data collection or two

sources can be performed in two ways; measurement of validity or measurement of reliability.[205] Although the terms validity and reliability can be used interchangeably to describe the agreement between two sources of information, they differ in that the validity analysis requires that one of the two sources is superior to the other, or one source can be described as a “gold standard”. Whereas, the reliability analysis assumes that the two sources are similar without *a priori* assumption that one is superior to the other. The important metrics of validity analysis are; sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) (Table 4-1). It is important to consider that the denominators of the first two measures represent the actual number of subjects with or without the outcome. However, the denominators of the last two measures (i.e. predictive values) are the number of subjects classified as having or not having the outcome. Hence, the predictive values measure the performance of classification not the validity of the test, and are influenced by the prevalence of the outcome which may differ between populations.

The second method to assess the concordance between the two methods of data collection is applied when both of them are considered equivalent (i.e. neither of them can be considered as gold standard). In this situation, the measurement of reliability statistics evaluates the agreement between the two methods using Cohen’s Kappa coefficients (κ) statistics.[209-211] Kappa is designed to improve the simple measure of proportional agreement between two observers (p_o) by subtracting the proportion of agreement that is expected by chance (p_e) from the optimal agreement (1). Hence, the calculation of kappa is based on constructing a ratio between the chance-corrected observed agreement and the chance-corrected perfect agreement.[212] That is, subtracting the value of the expected agreement by chance (p_e) from both the observed agreement (p_o) and the perfect agreement with a value of (1). The kappa value can be interpreted as follows: poor agreement if κ <0.20, fair agreement if κ =0.21-0.40, moderate agreement if κ =0.41-0.60, good agreement if κ =0.61-0.80, and very good agreement if κ >0.80.[209-211] Although the kappa statistic is quite simple and is increasingly reported in the literature, the value of kappa is influenced by the skewed distribution of the outcome or unbalanced marginal total, resulting from high prevalence of the outcome or systemic bias. Hence, other measures are

reported here to highlight the reliability results. The effect of prevalence occurs when the proportion of positive agreement is greatly different from the proportion of negative agreement, and it is expressed by the prevalence index (PI). In the presence of higher PI, the agreement just by chance is also higher and thus kappa is lower (this illustrates the first paradox of kappa).[210,211] The effect of bias occurs when the two sources disagree on the proportion of the outcome, and it is expressed by the bias index (BI). In the presence of higher BI, large bias exists between the two sources, which leads to higher kappa. An adjusted kappa value has been used in some literature, known as prevalence-adjusted bias-adjusted kappa (PABAK), which is calculated by substituting the positive and negative agreement values by their average to adjust for prevalence, and the other two disagreement values by their average to adjust for bias. Yet, Hoehler has criticized this value as he showed through simulation studies that the effect of bias and prevalence are informative and should not be disregarded by adjusting.[213] Finally, Cicchetti and Feinstein (1990) proposed two indices to be presented with kappa for better describing the observed agreements; that are the proportion of the observed positive agreement (P_{pos}) and the proportion of the observed negative agreement (P_{neg}). These two indices are considered as analogous to sensitivity and specificity, in the scenario of no *a priori* “gold standard”.[212,214]

In this thesis, the two methods of measuring the validity and reliability were applied as following; the EPRs ($Treatment^{prescr}$) was considered the “alloyed gold standard” in one analysis, and the validity of SRMs ($Treatment^{Ques}$) was assessed by applying the method of measuring validity (i.e. sensitivity, specificity, and predictive values). The second analysis was performed under the assumption of no gold standard method and the reliability measure was used by applying kappa statistics. In order to compare the two PCQ phases and to examine the effect of questionnaire structure, the two PCQ phases were compared using the measurement of reliability methods. However, because the two PCQ phases ($Treatment^{PCQ-1}$) and ($Treatment^{PCQ-2}$) were mutually exclusive, no subjects have completed both of them. An indirect assessment was performed by comparing both of them independently against ($Treatment^{prescr}$). The contingency 2x2 tables were constructed using SPSS (version 22.0; SPSS Inc. Chicago, Illinois, USA) after coding participants into “1”, “0”, or “NA” if they were taking BP-

lowering medication, not, or not applicable, respectively (i.e. if participant was coded “1 or 0” for $Treatment^{PCQ-1}$, he will be coded as “NA” for $Treatment^{PCQ-2}$). Hence, each participant received at least three codes for the previous four sources (i.e. $Treatment^{prescr}$, $Treatment^{Ques}$, $Treatment^{PCQ-1}$, and $Treatment^{PCQ-2}$). This analysis was performed using the subset of individuals who were identified as hypertensive using the definition of [a mean BP value ≥ 140/90 mmHg, taking a BP-lowering medication based on either EPRs or SRMs, or answered “yes” to the question about a previous medical diagnosis of high BP].

Table 4-1 Definitions and formulas of the concordance analysis measures

<table>
<tr><td rowspan="7">Notation</td><td rowspan="5">SRMs</td><td></td><td colspan="3">EPRs</td></tr>
<tr><td></td><td>Treated</td><td>Not treated</td><td>Total</td></tr>
<tr><td>Treated</td><td>A</td><td>B</td><td>g_1</td></tr>
<tr><td>Not treated</td><td>C</td><td>D</td><td>g_2</td></tr>
<tr><td>Total</td><td>f_1</td><td>f_2</td><td>N</td></tr>
<tr><td rowspan="2">Others</td><td>m</td><td colspan="3">$(A + D)/2$</td></tr>
<tr><td>n</td><td colspan="3">$(C + B)/2$</td></tr>
</table>

Measure	Formula and definition
Sensitivity	The degree to which the self-reported source correctly identifies participants taking BP-lowering medication [$\frac{A}{f_1}$].
Specificity	The degree to which the self-reported source correctly identify participants not taking BP-lowering medication [$\frac{D}{f_2}$].
Positive predictive value (PPV)	The probability that subjects with a positive test are truly positive [$\frac{A}{g1}$].
Negative predictive value (NPV)	The probability that subjects with a negative test truly do not have the outcome $\frac{D}{g2}$.
Observed agreement (p_o)	The proportion of observed agreement in both sources [$\frac{A+D}{N}$].
Expected agreement (p_e)	The proportion of agreement that is expected by chance [$\frac{(f1*g1)+(f2*g2)}{N^2}$].
Cohen's Kappa coefficients (κ)	The proportion of agreement beyond that is expected by chance [$\frac{(p_o - p_e)}{(1 - p_e)}$].
Prevalence index (PI)	The proportion of agreements that is due to the prevalence of the outcome only [$\frac{\lvert A-D \rvert}{N}$].
Bias index (BI)	The proportion of disagreement between the two sources that is due to systemic difference (bias) [$\frac{\lvert B-C \rvert}{N}$].
Bias-adjusted and prevalence-adjusted Kappa (BAPAK)	Kappa statistics after adjusting for the bias and prevalence effects [$\frac{\left(\frac{2n}{N}\right) - p_e}{1 - p_e}$].
Proportion of positive agreement (P_{pos})	The observed proportion of positive agreement [$\frac{2A}{f_1 + g_1}$].
Proportion of negative agreement (P_{neg})	The observed proportion of negative agreement [$\frac{2D}{f_2 + g_2}$].

4.3.3 Definitions of hypertension related indicators

Several hypertension-related indicators were evaluated for each participant: hypertension status, treatment, control and awareness of hypertension. The definition of each indicators is:

- Hypertension: SBP ≥140 mmHg or a DBP ≥90 mmHg, or receiving treatment.
- Treatment: taking at least one BP-lowering medication based on EPRs where available, or based on SRMs for the remaining participants.
- Controlled hypertension: taking a treatment with BP level below 140/90 mmHg if aged <80 years, or below than 150/90 mmHg if aged ≥80 years.
- Awareness: self-report of a previous diagnosis of hypertension by a healthcare professional.
- BP category: the seventh report of the Joint National Committee was used to categorize participant based on their BP values[215]:
    - Prehypertension: SBP 120-139 mmHg, or DBP 80-89 mmHg
    - Stage-1 hypertension: SBP 140-159 mmHg, or DBP 90-99 mmHg
    - Stage-2 hypertension: SBP ≥160 mmHg, or DBP ≥100 mmHg
- High-risk population: it includes the subset of participants with a complete medical history (i.e. SRM, EPR, and self-reported history of hypertension) who were defined as hypertensive based on the study definition and/or self-reported history of clinically diagnosed hypertension.

4.3.4 Estimating of hypertension-related indicators prevalence

For descriptive analysis, continuous variables are expressed as mean ± SD, and categorical variables were expressed as number of subjects and percentage.

Baseline variables were compared using a chi-square test for categorical variables, and analysis of variance (ANOVA) for continuous variables; a two-tailed p value of <0.05 was considered significant. Simple descriptive analysis was performed using SPSS (version 22.0; SPSS Inc. Chicago, Illinois, USA).

The prevalence of hypertension, awareness, treatment and control are represented as crude proportions, and as age- and sex- standardized prevalence. In addition, they were stratified by different risk factors - BMI, SIMD quintiles, smoking status, number of parents with hypertension as reported by participants in the PCQ, and number of years at school. The standardization procedure was performed by applying a direct method using age and sex sub-groups distribution of mid-year 2011 census household estimates for Scotland as a standard population; the age sub-groups were: 18–24, 25–34, 35–44, 45–54, 55–64, 65–74, and +75 years (Appendix 1). The age-standardized proportion ($p`$) was calculated as follows:

$$p` = \frac{\sum_i N_i P_i}{\sum_i N_i}$$

Where $p`$ is the age specific rate in the age group *i* and N_i is the standard population in age group i. The variance of the standardized proportion was estimated by:

$$var(p`) = \frac{\sum_i (N_i^2 \; p_i \, q_i \, / \, n_i)}{(\sum_i N_i)^2} \quad ,$$

Where $q_i = 1 - p_i$

4.3.5 Assess the effect of socio-economic factors in hypertension-related indicators

A multilevel mixed effect model that accounts for the family structure of the sample was used to examine the factors associated with selected outcome variables. The model can be considered as two levels where individuals were treated as the level one unit and families were treated as the level two unit. This would allow the simultaneous examination of the effects of family-level and individual-level variables on individual-level outcome, while controlling for the

correlation within family members.[216] The family units were constructed to include members of a nuclear family only. A generalized linear-mixed model were applied to assess that effect of socio-economic factors in the different hypertension-related binary outcomes; in which, individuals were nested within families in the logistic-mixed model that included the hypertension-related indicators (i.e. hypertension status) as dependent variables after adjusting for age, sex, and BMI. The results are presented in tables showing odd ratios (OR) with 95% CI for the final models; ORs are showed relative to reference category. Mixed-models were estimated using R package “lme4”.[217]

4.4 Results

4.4.1 Participants baseline characteristics

The baseline characteristics of the GS:SFHS final populations (n =18,470) are provided in Table 4-2. The final study population had a higher proportion of women (58.7%) than men (41.3%). The mean of age was 45.3 ± 15.0 years, and 45.8% of the participants were middle aged (40–59 years old). A higher percentage of men were active smokers compared to women (18.4% and 15.8%, respectively), and 51.7% of GS:SFHS participants had never smoked. The average BMI was 26.6 ± 5.1 kg/m^2, and 21.2% of the participants had a BMI of 30 kg/m^2 or more (obesity). Only 3% had a reported a history of diabetes and 4.5% reported a history of other cardiovascular diseases. The mean SBP and DBP were 131.5 ± 17.8 mmHg, and 79.8 ± 10.3 mmHg, respectively. Approximately half of the participants (52.4%) lived in area with above average SIMD score. The high risk group included 5,181 (28%) individuals, and 5,071 (98%) of them were defined as hypertensive based on the study definition of hypertension.

4.4.2 Pharmacoepidemiological analysis

For all the 13,732 individuals with EPRs, 1,956 (14.3%) individuals had at least one BP-lowering medication that was prescribed within the time-window of 120 days before the recruitment day. The number of BP-lowering medication classes per individual ranges from a single medication in 1011 (52%) individuals to a maximum of five BP-lowering medications in 12 (1%) individuals (Table 4-3); the number of individuals taking two, three and four classes of BP-lowering

medication were 648 (33%), 223 (11%), and 62 (3%), respectively. The most commonly prescribed monotherapy class was ACEIs, followed by BBs, CCBs, and thiazide diuretics. In dual combination therapy, ACEIs and diuretics were the most common combination therapy followed by ACEIs and BBs; ACEIs & CCBs; and BBs & diuretics. The proportion of controlled hypertension according to the number of BP-lowering classes were 47%, 46%, 43%, 39%, and 50% for individuals taking one, two, three, four, and five BP-lowering medications, respectively. Furthermore, only 53% participants who were taking a single BP-lowering class declared a previous diagnosis of hypertension, comparing to 71%, 78%, 85%, and 83% for the participants taking two to five medications, respectively. Similarly, the proportion of participants who declared other cardiovascular or diabetes is higher in the groups of two or more BP-lowering classes.

The number of individuals reported taking BP-lowering medication by SRMs was 2,851 (16.9%). Of the 8,253 individuals who completed PCQ-1, 1,334 individuals (16.2%) reported taking BP-lowering medication by writing the medication names, and of the 8,832 individuals completed PCQ-2, 1,517 (17.2%) reported BP-lowering medication by giving affirmative answer to the corresponding yes/no question.

Table 4-2 Baseline characteristics of study participants

Category	All		Men		Women	
Total	18470	100%	7632	41.3%	10838	58.7%
Age, mean±SD, years	47.6 ± 15.0		47.0 ± 15.1		47.6 ± 14.8	
BMI, mean±SD, kg/m^2	26.6 ± 5.1		26.9 ± 4.5		26.5 ± 5.5	
Average SBP, mmHg	131.6 ± 17.8		136.3 ± 15.8		128.3 ± 18.3	
Average DBP, mmHg	79.8 ± 10.3		81.8 ± 10.4		78.4 ± 10.1	
eGFR, ml/min/1.73m^2	87.3 ± 17.6		89.9 ± 17.6		85.4 ± 17.1	
Total Cholesterol, mmol/l	5.1 ± 1.1		5.0 ± 1.1		5.2 ± 1.1	
HDL Cholesterol, mmol/l	1.5 ± 0.2		1.3 ± 0.3		1.6 ± 0.4	
Glucose, mmol/l	4.8 ± 1.1		5.0 ± 1.2		4.7 ± 1.0	
Diabetic*	548	3.00%	293	3.80%	255	2.40%
Cancer*	397	2.20%	107	1.40%	290	2.70%
Other CVD diseases*	831	4.50%	477	6.30%	354	3.30%
Age group (years)						
≤ 39 years old	5734	31.0%	2481	32.5%	3253	30.0%
40–59 years old	8468	45.8%	3366	44.1%	5102	47.1%
≥ 60 years old	4268	23.1%	1785	23.4%	2483	22.9%
Smoking						
Active	3123	16.9%	1407	18.4%	1716	15.8%
Ex-smoker	5348	29.0%	2299	30.1%	3049	28.1%
Never	9549	51.7%	3716	48.7%	5833	53.8%
missing	450	2.4%	210	2.75%	240	2.20%
BMI categories (kg/m2)						
≤ 18.5	292	1.6%	75	1.0%	217	2.0%
18.5–24.9	7405	40.1%	2616	34.3%	4789	44.2%
25–29.9	6854	37.2%	3391	44.4%	3463	32.0%
≥ 30	3919	21.2%	1550	20.3%	2369	21.9%
SIMD quintile						
1	2287	12.4%	860	11.2%	1247	11.51%
2	2375	12.9%	908	11.9%	1467	13.54%
3	2725	14.8%	1176	15.4%	1549	14.29%
4	4362	23.6%	1831	23.9%	2531	23.35%
5	5320	28.8%	2283	29.9%	3037	28.02%
Number of years at school						
≤ 11 years	6281	34%	2665	37.0%	3616	35.16%
11–17 years	9392	51%	3707	51.5%	5685	55.2%
≥ 17 years	1825	10%	824	11.5%	1001	9.70%

Underlined percentages are significantly different between sex
*Based on participant's self-report answer in the family-health history section of PCQ, data were available for 18171 individuals only.

Table 4-3 Frequency of BP-lowering medications classes based on EPRs

	Medication class	Total[a]	Female	Controlled[b]	HTN[c]	CVD[c]
1	ACEIs	281 (0.14)	143 (0.51)	118 (0.42)	186 (0.68)	40 (0.15)
	ARBs	96 (0.05)	45 (0.47)	29 (0.30)	72 (0.77)	9 (0.10)
	BBs	264 (0.13)	172 (0.65)	166 (0.63)	70 (0.27)	46 (0.18)
	CCBs	163 (0.08)	90 (0.55)	66 (0.40)	102 (0.65)	21 (0.13)
	Thiazides diuretic	120 (0.06)	91 (0.76)	38 (0.32)	80 (0.67)	9 (0.08)
	Other diuretics	54 (0.03)	45 (0.83)	40 (0.74)	13 (0.24)	11 (0.20)
	Other classes	33 (0.02)	25 (0.76)	17 (0.52)	5 (0.16)	1 (0.03)
	Total of one medication	1011 (0.52)	611 (0.60)	474 (0.47)	528 (0.53)	137 (0.14)
2	ACEIs & BBs	84 (0.04)	26 (0.31)	51 (0.61)	44 (0.55)	47 (0.59)
	ACEIs & CCBs	80 (0.04)	30 (0.38)	41 (0.51)	59 (0.75)	17 (0.22)
	ACEIs & diuretics	112 (0.06)	71 (0.63)	52 (0.46)	84 (0.76)	18 (0.16)
	ARBs & BBs	25 (0.01)	16 (0.64)	9 (0.36)	15 (0.63)	11 (0.46)
	ARBs & CCBs	37 (0.02)	18 (0.49)	12 (0.32)	30 (0.81)	5 (0.14)
	ARBs & diuretics	70 (0.04)	47 (0.67)	26 (0.37)	50 (0.74)	6 (0.09)
	BBs & CCBs	45 (0.02)	24 (0.53)	21 (0.47)	29 (0.64)	16 (0.36)
	BBs & diuretics	80 (0.04)	65 (0.81)	39 (0.49)	57 (0.72)	13 (0.16)
	CCBs & diuretics	66 (0.03)	45 (0.68)	28 (0.42)	56 (0.85)	5 (0.08)
	Diuretics combinations	18 (0.01)	16 (0.89)	12 (0.67)	4 (0.22)	2 (0.11)
	Other combination	31 (0.02)	22 (0.71)	6 (0.19)	27 (0.87)	5 (0.16)
	Total of two medications	648 (0.33)	380 (0.59)	297 (0.46)	455 (0.71)	145 (0.23)
3	ACEIs, CCBs & diuretic	37 (0.02)	24 (0.63)	18 (0.47)	29 (0.78)	13 (0.35)
	ACEIs, BBs & CCB	20 (0.01)	3 (0.14)	7 (0.33)	15 (0.75)	13 (0.65)
	ACEIs, BBs & diuretics	42 (0.02)	16 (0.38)	19 (0.45)	32 (0.76)	23 (0.55)
	ARBs, BBs& CCB	11 (0.01)	2 (0.18)	2 (0.18)	9 (0.82)	6 (0.55)
	ARBs, BBs & diuretics	18 (0.01)	12 (0.67)	11 (0.61)	11 (0.61)	8 (0.44)
	ARBs, CCBs & diuretics	22 (0.01)	16 (0.72)	8 (0.35)	19 (0.86)	4 (0.18)
	BBs, CCBs & diuretics	28 (0.01)	15 (0.54)	16 (0.53)	22 (0.79)	6 (0.21)
	Other 3combinations	40 (0.02)	23 (0.58)	15 (0.38)	32 (0.80)	9 (0.23)
	Total of three medications	223 (0.11)	111 (0.5)	96 (0.43)	169 (0.78)	82 (0.38)
4	AA1Blocker, ARBs, BBs & diuretics	5 (0)	4 (0.80)	2 (0.4)	5 (1)	1 (0.20)
	ACEIs, AA1Blocker, CCB & diuretics	11 (0.01)	1 (0.09)	4 (0.36)	10 (0.91)	4 (0.36)
	ACEIs, BBs, CCBs & diuretics	14 (0.01)	4 (0.29)	5 (0.36)	11 (0.79)	5 (0.36)
	ARBs, BBs, CCBs & diuretics	6 (0)	3 (0.50)	4 (0.67)	6 (1)	4 (0.67)
	Other combinations	25 (0.01)	14 (0.54)	9 (0.35)	20 (0.80)	7 (0.28)
	Total of four medications	62 (0.03)	26 (0.42)	24 (0.39)	52 (0.85)	21 (0.34)
5	ACEIs, AA1Blocker, BB, CCB & diuretic	3 (0)	1 (0.33)	0	3 (1)	0
	Other combinations	9 (0)	4 (0.44)	6 (0.67)	7 (0.78)	6 (0.67)
	Total of five medications	12 (0.01)	5 (0.42)	6 (0.50)	10 (0.83)	6 (0.50)
Overall total		1956	1133 (0.58)	897 (0.46)	1214 (0.63)	391 (0.20)

(a) The numbers in brackets represent the proportion of the category in each medication regimen (for example, the proportion of female taking ACEIs was 143/281=0.51), except for the total column in which the proportion is calculated using the overall number as denominator (1956). (b) Controlled is defined as having a BP <140/90 mmHg for individuals aged up to 80 years, or <150 if the participant was ≥ 80 years. (b) Based on self-reported information of having hypertension (HTN), cardiovascular diseases (CVD), or diabetes.

4.4.3 Validity of SRMs

This analysis was restricted to the high risk population (n =5,181). In total 2,166 participants self-reported taking BP-lowering medications, of them 1,624 (75.0%) could be confirmed with EPRs. The remaining 542 (25%) were only identified as taking BP-lowering medication by SRMs but without a confirmation from the EPRs. Using EPRs only, the number of treated hypertensive individuals was 1,808, with 1,624 (89.2%) confirming their BP therapy in SRMs (Table 4-4). The sensitivity of self-reported BP-lowering medication was 90% and the specificity was 83%, with PPV and NPV of 75% and 64%, respectively. Asking participants to write their medication names as in PCQ-1 had better sensitivity (92% vs. 82%) and NPV (69% vs. 62%) compared to asking them to only indicate if they were regularly taking BP-lowering medication or not as in PCQ-2. However, the closed-end question in PCQ-2 showed better specificity (80% vs. 87%) and PPV (68% vs. 81%) compared to the open-end question in PCQ-1.

In the second analysis where no source was considered as superior, kappa statistics showed a “good agreement” with a value of 71% (P_{pos} =82% and P_{neg} =89%) for concordance between SRMs and EPRs. Between the two PCQ phases, PCQ-2 showed better agreement with the EPRs than PCQ-1, with a kappa value of 74% (P_{pos} =84% and P_{neg} =97%), compared to 66% (P_{pos} =78% and P_{neg} =96%) for PCQ-1.

Out of the 5,181 participants in the high risk population, 1,168 (22.5%) showed complete agreement with the three metrics of SRMs, EPRs, and self-reported diagnosis of hypertension (Figure 4-1); of those with complete agreement, 719 (61.5%) participants had a BP level equal to 140 mmHg or higher. Almost half of the participants (2,558) were classified as hypertensive based only on BP criteria, as they were not on BP-lowering medication based on any medication history sources, and they did not report a previous clinical diagnosis of hypertension. The total number of participants who declared a previous hypertension diagnosis was 1,840 (35.5%); of which 273 participants were not on treatment, based on both sources, despite more than half of them having a high BP level (≥140 mmHg) (Figure 4-1).

Table 4-4 Concordance analyses of the two medication history sources in participants with BP-lowering medication history sources

			Treatment[Prescr]		
			Yes	No	Total
Treatment[Ques]	Yes	n	1624	542	2166
		%	31.0%	10.5%	41.8%
		CI	[30.1–32.6%]	[9.7–11.3%]	[40.5–43.5%]
	No	n	184	2831	3015
		%	3.6%	54.6%	58.2%
		CI	[3.1–4.1%]	[53.3–56.0%]	[56.8–59.5%]
Total			1808 34.9% [33.6–36.2%]	3373 65.1% [63.8–66.4%]	5181
Measurements of validity	Sensitivity		90%	Specificity	84%
	PPV		75%	NPV	65%
Measurements of agreement	p_o		86%	p_e	52%
	K		71%	PI	23%
	BI		7%	PABAK	72%
	P_{pos}		82%	P_{neg}	89%
Validity analysis by PCQ					
PCQ-1 Treatment[PCQ1]	Yes	n	673	316	1269
		%	28.8%	13.5%	54.3%
		CI	[28.0–31.7%]	[12.6–15.5%]	[54.1–58.2%]
	No	n	60	1287	1347
		%	2.6%	55.1%	57.7%
		CI	[2.0–3.3%]	[53.1–57.1%]	[55.7–59.7%]
Total			733 31.4% [29.5–33.3%]	1603 68.6% [66.7–70.5%]	2336
Measurements of validity	Sensitivity		92%	Specificity	80%
	PPV		68%	NPV	69%
Measurements of agreement	p_o		84%	p_e	53%
	K		66%	PI	26%
	BI		11%	PABAK	68%
	P_{pos}		78%	P_{neg}	87%
PCQ-2 Treatment[PCQ2]	Yes	n	951	226	1177
		%	33.4%	7.9%	41.4%
		CI	[31.7–35.2%]	[7.0–9.0%]	[39.6–43.2%]
	No	n	124	1544	1668
		%	4.4%	54.3%	59.3%
		CI	[3.7–5.2%]	[52.4–56.1%]	[57.5–61.1%]
Total			1075 37.8% [36.0–39.6%]	1770 61.2% [60.4–64.0%]	2845
Measurements of validity	Sensitivity		88%	Specificity	87%
	PPV		81%	NPV	62%
Measurements of agreement	p_o		88%	p_e	52%
	K		74%	PI	21%
	BI		4%	PABAK	75%
	P_{pos}		84%	P_{neg}	90%

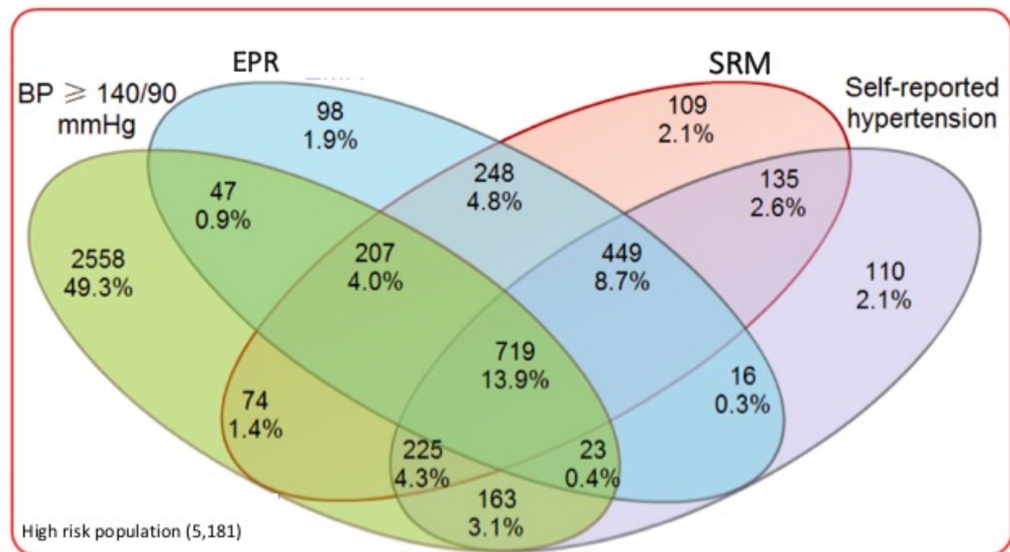


Figure 4-1 Venn diagram of the status of high-risk population in GS for BP level, treatment based on EPRs and SRMs, and self-reported hypertension.
The green circle includes participants with a BP level that is 140/90 mmHg or over, blue circle includes participants identified as taking BP-lowering medication based on EPRs, red circle includes participants identified as taking BP-lowering medication based on SRMs, purple circle includes participants indicated that they have received a clinical diagnosis of hypertension by medical staff.

4.4.4 Distribution of BP in GS:SFHS

The distribution of SBP and DBP by age groups and sex is shown in Figure 4-2. SBP increases with age in both men and women, with higher SBP in women only after the age of 75 years. The differences in mean SBP between men and women were statistically significant for all age groups up to the age group bin of 65–74 years, and thereafter SBP did not show statistically significant difference between sexes. The mean DBP also increased with age but reached a plateau at 55–64 years, and then decreased with increasing age. The difference in DBP between sexes were statistically significant in age groups bins between 25–34 and 65–74 years, and there was no statistical difference between men and women for both the youngest and oldest age group bins.

Comparison of the observed BP level between normotensive and hypertensive (treated and untreated) is depicted in Figure 4-3. On average, women had a lower BP than men by 8/3.4 mmHg (P <0.0001) overall; and by 6.9/1.8 mmHg (P <0.0001) in the normotensive group. The differences between sexes were not statistically significant in the two hypertensive groups (i.e. treated and untreated). The age-adjusted prevalence of the four BP level classes showed a higher prevalence of normotensive class in women compared to men [36.8% (CI: 35.6–38.0%), and 14.0% (CI:13.1–14.9%), respectively].The prevalence of the remaining three classes were higher in men than women (Table 4-5).

Table 4-5 Age-standardized prevalence of BP classes in men and women

BP category	Men % (CI%)	Women % (CI%)
Normotensive	14.0 (13.1–14.9)	36.8 (35.6–38.0)
Pre-hypertensive	46.4 (44.7–48.0)	34.4 (33.2–35.6)
Stage 1	29.5 (28.2–30.9)	19.5 (18.5–20.6)
Stage 2	10.1 (9.3–11.0)	9.3 (8.5–10.1)
Age- standardization was performed by applying the direct method, using the mid-year 2011 census household estimates for Scotland as standard population, Appendix 1		

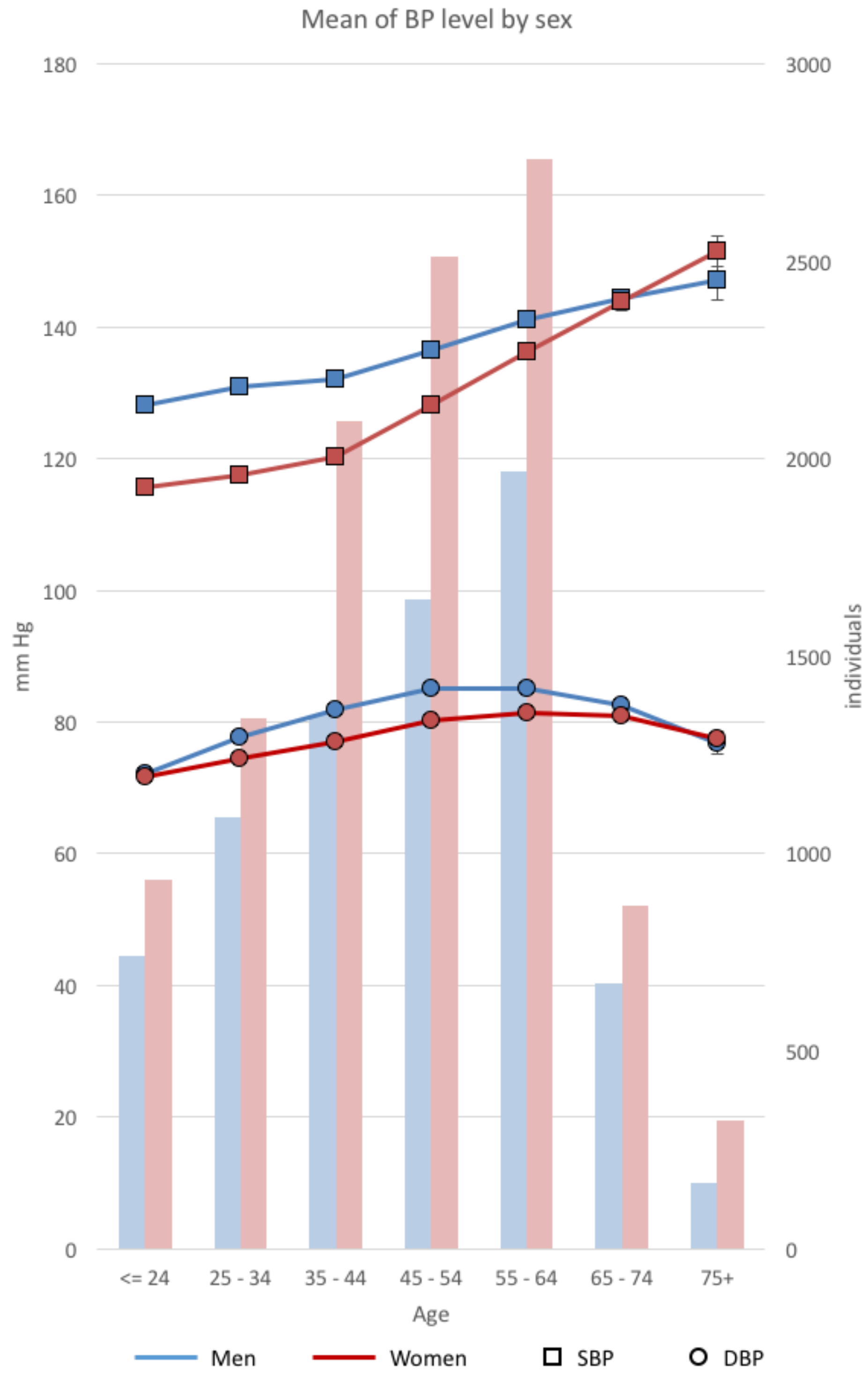


Figure 4-2 Distribution of the mean of SBP and DBP by sex and age.
The bar graph with its axis title on right shows the number of individuals in each age bin.

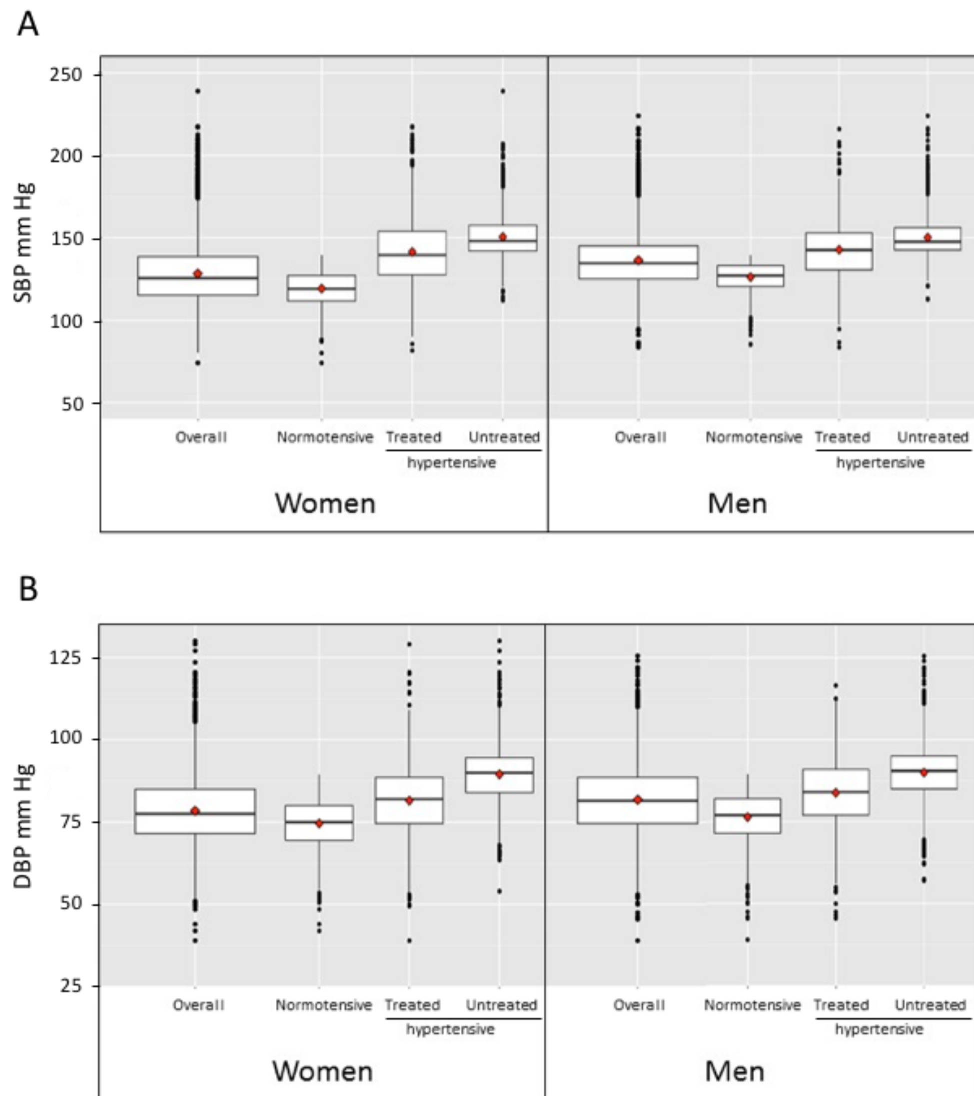


Figure 4-3 Box plots of SBP and DBP levels by sex and hypertensive status.
(A) box plots show SBP level in the overall and groups of normotensive, treated, and untreated hypertension groups in women and men. (B) box plots show the same information for DBP. The red diamonds inside the box plots represents the mean, and the lines within the boxes indicate the median, the edges of the white boxes are the first and third quartiles (IQR), the whiskers extend to 1.5 times the IQR, values above/below that are represented with dots.

4.4.5 Prevalence of hypertension

The crude prevalence of hypertension, awareness, treatment and control are shown in Table 4-6. The overall crude prevalence of hypertension was 38.7%, with a prevalence of 47% and 33% in men and women, respectively. The crude prevalence table is presented here to show the numbers of individuals in each sub-group; however, all quoted prevalence data hereafter are age- and sex-adjusted (Table 4-7). The overall age-and sex adjusted prevalence in GS:SFHS was 40.8% [CI: 39.7–41.9%]; the prevalence of hypertension was higher among men in every other age group below 75 years (Figure 4-4). The difference in prevalence between genders was large in those under 54 years. Hypertension prevalence increased with age in both sexes. In women, the prevalence was relatively low in the age groups younger than 45 years old (4.2%–15.8%), and then doubled in those aged 45–54 (31.6%). In men, however, the increase in prevalence of hypertension was steady with age; the most observed difference was between age bins of 35–45 and 45–54 years old. Examining the prevalence of hypertension across different risk factor categories showed a correlation with BMI, with prevalence increasing in each BMI group from 20.3% for BMI ≤18.5 up to 55.9% for BMI ≥30 (Table 4-7). The prevalence of hypertension among participants stating that none of their parents were hypertensive was 37.7%, compared to 44.3% and 52.6% in participants with one and two hypertensive parents, respectively. Furthermore, the prevalence of hypertension tended to be lower in the least deprived area of Scotland (SIMD =5; 38.5%), compared to area of the most deprived area (SIMD =1; 43.4%). Active smokers had lower prevalence of hypertension compared to ex-smokers or never smokers, though the difference was very small. Finally, hypertension was higher in people with a lower level of education as examined by the number of years at school.

4.4.6 Prevalence of hypertension awareness

The overall rate of awareness was 25.3% [CI: 23.9–26.7%]; and the proportions of aware hypertensive increased with age in both sexes (Table 4-7). Women were more aware than men in every age group except for the age groups >55-74 years old, where awareness was similar in both men and women (Figure 4-4). The proportions of aware participants across the different risk factors categories had generally similar trends to hypertension status with some exceptions. First,

awareness was lower in the normal BMI group (BMI 18.5–24.9 kg/m2) compared to the underweight group (BMI <18.5 kg/m2). Second, there is a general trend of higher awareness of hypertension in the most deprived areas of Scotland (SIMD quintiles 1 and 2). Finally, a clear difference was observed in the proportion of awareness between active (21.2%) and ex-smokers or never smokers (25.7%).

4.4.7 Prevalence of treated hypertension

The overall prevalence of treated hypertension was 31.2% [CI: 29.5–32.9%], with a higher prevalence of treatment in older people. Women had higher prevalence of treated hypertension in the younger groups, and both sexes had similar prevalence at the age of 55–64 years, after which men were more frequently treated (Figure 4-4). The prevalence of treated hypertensive was lower in the normal and overweight groups compared to the underweight and obese groups (Table 4-7). In addition, participants with one and two hypertensive parents had higher prevalence of treated hypertension compared to participants without any hypertensive parents (32.2% and 39.7% compared to 29.2%, respectively). With the exception of participants living in area of the fourth SIMD quintiles, areas with the lowest SIMD quintiles tend to have a higher prevalence of treated hypertension compared to areas with above average SIMD quintiles.

4.4.8 Prevalence of controlled hypertension

The overall prevalence of controlled hypertension among treated hypertensive was 54.3% [CI: 46.5–62.1%], women had a higher prevalence of controlled hypertension up to the age group of 55–64 years old, while men started to have higher prevalence of controlled hypertension after 55–64 years (Figure 4-4). The prevalence of controlled hypertension in women sharply decreased with age; for instance, all the treated women younger than 24 years were controlled compared to only 29.5% of women older than 75 years old. This trend was less obvious in men with only a general trend of lower control in older groups. Lower BMI categories had a higher prevalence of controlled hypertension, and participants with both parents had hypertension had a lower controlled rate. Finally, participants living in the most deprived areas had a higher prevalence of controlled hypertension compared to those living above the average.

Table 4-6 Crude prevalence of hypertension, awareness, treatment, and control by baseline characteristics

Category			Prevalence among all participants		Awareness among hypertensive		Treatment among hypertensive		Control among treated hypertensive	
	N	%	n	%	n	%	n	%	n	%
Overall										
Men	7632	41%	3565	47%	980	27%	1152	32%	468	41%
Women	10838	59%	3581	33%	1223	34%	1489	42%	698	47%
All	18470	100%	7146	39%	2203	31%	2641	37%	1166	44%
BMI (Kg/m^2)										
≤18.5	292	1.60%	39	13%	6	16.0%	13	33%	10	77%
18.5–24.9	7404	40%	1762	24%	421	24%	551	31%	282	51%
25–29.9	6854	37%	3014	44%	855	30%	1065	35%	465	44%
≥30	3920	21%	2331	59%	891	39%	1012	43%	410	41%
Number of parents with hypertension[§]										
None	11354	61%	4153	37%	1060	26%	1461	35%	641	44%
One	5855	31%	2378	41%	856	36%	896	38%	399	45%
Two	1261	0.70%	615	49%	287	47%	284	46%	130	46%
SIMD										
1	2287	12%	892	39%	295	34%	365	41%	202	55%
2	2375	13%	941	40%	301	34%	368	39%	144	39%
3	2725	15%	1039	38%	299	29%	374	36%	174	47%
4	4362	24%	1680	39%	518	31%	649	39%	297	46%
5	5320	29%	2043	38%	595	30%	673	33%	283	42%
Smoking										
Active	3123	17%	1026	33%	262	8%	366	36%	211	58%
Ex-smoker	5348	29%	2472	46%	854	16%	998	40%	412	41%
Never	9549	52%	3474	36%	1070	11%	1208	35%	513	42%
Number of years at school										
≤11 years	6281	34%	3063	49%	1092	36%	1332	43%	586	44%
11–17 years	9392	51%	3153	34%	866	27%	998	32%	448	45%
≥17 years	1825	10%	546	30%	146	27%	152	28%	67	44%

Awareness, treatment, and control were assessed among hypertensive participants, control was assessed among treated participants.

§ Based on self-reporting PCQ.

Table 4-7 Age- and sex-adjusted prevalence of hypertension, awareness, treatment, and controlled hypertension

Category	Prevalence among all participants		Awareness among hypertensive		Treatment among hypertensive		Control among treated hypertensive	
	%	SE	%	SE	%	SE	%	SE
Overall*								
Men	46.3	0.89	21.7	0.64	25.7	0.89	47.8	6.17
Women	35.8	0.74	28.6	1.17	33.2	1.46	60.2	5.08
All	40.8	0.53	25.3	0.72	31.2	0.87	54.3	3.93
BMI categories (years)								
≤18.5	20.3	5.06	20.7	11.05	32.6	12.2	48.3	16.79
18.5–24.9	31.0	0.89	18.8	1.12	28.9	1.79	64.8	7.23
25–29.9	42.5	0.92	24.8	1.36	28.8	1.35	50.8	7.62
≥30	55.9	1.50	30.5	1.41	34.6	1.44	42.6	6.87
Number of parent with hypertension§								
None	37.7	0.66	19.4	0.82	29.2	1.16	57.9	5.84
One	44.3	1.30	30.4	1.46	32.2	1.71	56.1	8.19
Two	52.6	4.98	45.9	6.95	39.7	4.74	48.4	8.49
SIMD quintile								
1	43.4	1.70	27.9	1.84	34.1	2.14	60.9	9.00
2	44.0	1.72	25.7	1.75	32.6	2.24	58.1	12.94
3	40.9	1.72	23.3	1.86	29.3	1.93	44.2	6.21
4	40.0	1.22	25.4	1.52	33.9	1.97	56.3	8.43
5	38.5	1.06	23.3	1.55	29.5	2.22	51.2	9.74
Smoking								
Active	39.5	1.82	21.2	1.67	33.5	2.35	67.0	8.48
Ex-smoker	42.0	1.00	25.7	1.20	32.5	2.52	51.8	7.47
Never	40.2	0.86	25.7	1.05	29.8	1.20	53.5	7.72
Number of years at school								
≤11 years	41.4	0.88	27.4	2.06	35.0	2.80	60.5	10.08
11–17 years	39.1	0.95	23.4	1.05	28.48	1.26	53.4	5.67
≥17 years	34.4	2.23	23.0	2.64	25.8	3.03	47.5	8.76

Age- and sex-standardization was carried out by applying the direct method using the mid-year 2011 census household estimates for Scotland as standard population.
*the overall prevalence was only standardized for age and stratified by sex.
§ Based on self-reporting PCQ.

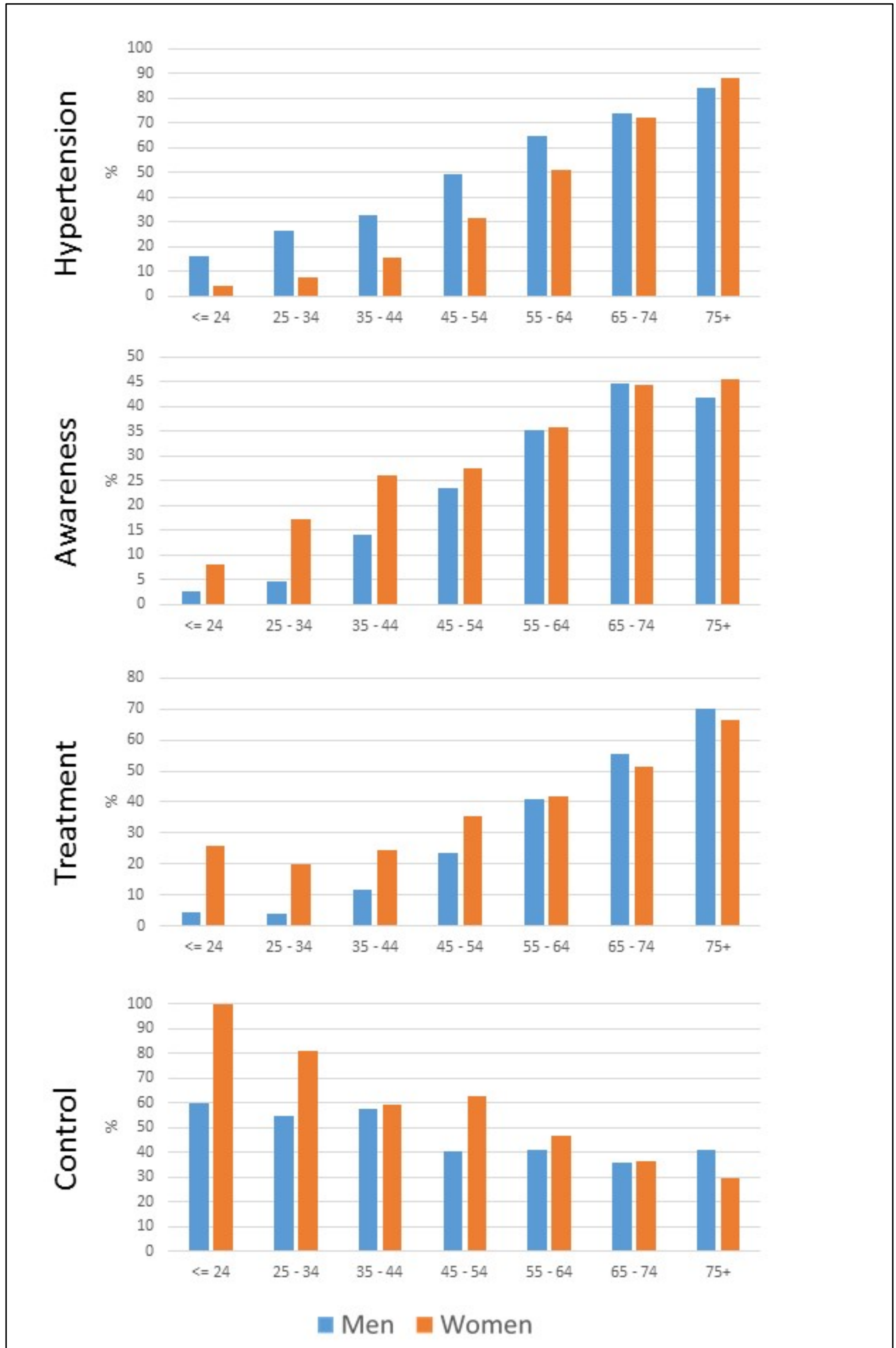


Figure 4-4 Age- and sex stratified prevalence of hypertension, awareness, treatment, and control by age groups

4.4.9 Determinants of hypertension related indicators

Mixed model logistic regression models were used to assess the independent effect of a range of risk factors with hypertension-related health outcomes after adjusting for age, sex, and BMI. Table 4-8 shows the OR and CI for the association of the tested factors with hypertension prevalence, awareness, treatment and control. The odds of hypertension were higher in men, overweight (BMI ≥25 kg/m^2), participants with at least one hypertensive parent, and those living in the most deprived area of Scotland (SIMD quintiles 1 and 2). Among hypertensive participants, men were less likely to be treated than women [OR 0.88 (95% CI:0.78–0.98)], and no statistically significant association was found between treatment and BMI category. The odds of receiving treatment were 1.46 (95% CI:1.28–1.66) times higher in participants with one hypertensive parent, and 2.65 (95% CI:2.15–3.26) times higher in participants with two hypertensive parents compared to hypertensive participants without history of parental hypertension. Participants living in the least deprived area of Scotland were significantly less likely to be treated for hypertension compared to all the other regions, except for area of average SIMD. Furthermore, awareness among hypertensive participants was statistically higher among obese and subjects with parental history of hypertension and among those living in the most deprived areas of Scotland. Among treated participants, men were less likely to be controlled than women [OR 0.79 (95% CI: 0.66–0.94). Obese subjects were also less likely to be controlled compared to participants with normal BMI level, and people living in the most deprived area had higher odds of controlled hypertension than those living in the least deprived area.

Table 4-8 Multivariate association of characteristics with hypertension, awareness, treatment, and control

Variable	Hypertension	Treatment[a]	Control[b]	Awareness[a]
	OR (95% CI) P value	OR (95% CI) P value	OR (95% CI) P value	OR (95% CI) P value
Sex				
Women	1	1	1	1
Men	**2.31 (2.13–2.50) <2e-16**	**0.88 (0.78–0.98) 0.03**	**0.79 (0.66–0.94) 0.008**	0.95 (0.84–1.07) 0.41
BMI (kg/m2)				
Normal (18.5–24.9)	1	1	1	1
Underweight (≤ 18.5)	1.19 (0.79–1.80) 0.40	1.59 (0.71–3.54) 0.26	1.44 (0.40–5.15) 0.57	0.71 (0.27–1.88) 0.49
Overweight (25–29.9)	**1.25 (1.12–1.41) 0.00015**	1.08 (0.91–1.28) 0.39	0.78 (0.61–1.01) 0.06	1.18 (0.89–1.41) 0.0746
Obese (≥ 30)	**1.39 (1.12–1.73) 0.0024**	1.30 (0.99–1.74) 0.063	**0.65 (0.44–0.96) 0.03**	**1.44 (1.08–1.91) 0.0118**
Number of parents with hypertension				
No parents	1	1	1	1
one parent	**1.45 (1.33–1.57) <2e-16**	**1.46 (1.28–1.66) 9.46e-09**	0.87 (0.72–1.05) 0.13	**2.11 (1.85–2.42) <2e-16**
two parents	**2.02 (1.74–2.34) <2e-16**	**2.65 (2.15–3.26) <2e-16**	0.76 (0.57–1.01) 0.057	**4.09 (3.31–5.06) <2e-16**
SIMD quintiles				
SIMD 5 (least deprived)	1	1	1	1
SIMD 4	1.05 (0.94–1.16) 0.4	**1.33 (1.14–1.56) 0.00027**	1.18 (0.94–1.49) 0.15	1.07 (0.91–1.26) 0.4
SIMD 3	1.06 (0.94–1.19) 0.35	1.16 (0.97–1.39) 0.11	1.04 (0.79–1.36) 0.78	0.96 (0.79–1.16) 0.66
SIMD 2	**1.16 (1.03–1.32) 0.0187**	**1.34 (1.11–1.62) 0.0019**	0.93 (0.71–1.22) 0.6	1.11 (0.92–1.35) 0.28
SIMD 1 (most deprived)	**1.16 (1.02–1.32) 0.022**	**1.61 (1.33–1.94) 1.21e-06**	**1.73 (1.31–2.28) 0.0001**	**1.27 (1.04–1.55) 0.016**

a. the model only included hypertensive participants.
b. the model only included individuals taking BP-lowering medication(s).

4.5 Discussion

This chapter provides a descriptive analysis of BP and hypertension phenotypes in GS:SFHS, with assessment and validation of BP-lowering medication exposure. The results of this chapter provide an estimate of the quality of the BP and hypertension phenotype in GS:SFHS, and enable a comparison of the epidemiological findings in GS:SFHS to other populations. GS:SFHS included participants from Scotland with a wide range of socio-demographic and clinical features, and is considered to be one of the largest family-based genetic epidemiology studies. The importance of this chapter relies on the fact that participants will be classified as hypertensive or not based on their exposure to BP-lowering medication (i.e. participants taking BP-lowering medication would be classified as hypertensive) and their blood pressure measurements. Hence, assessing the quality of self-reported data that are subject to recall bias is essential to have an estimate of the data quality, especially for the subset of participants without EPRs. Several analyses were performed in this chapter to achieve its aims based on the available resources, as discussed in the following sections.

4.5.1 Quality of SRMs history

The assessment of self-reported BP-lowering medication history used the subset of individuals with both medication exposure sources for 68% of GS:SFHS participants. The analysis showed suboptimal agreement between the two sources for BP-lowering medication history. The sensitivity analysis showed that almost 90% of all the BP-lowering medication users based on EPRs recalled their medications in the PCQ (i.e. sensitivity 90%), and 84% of high-risk participants who did not recall their medication in the PCQ were truly not on treatment based on EPRs (i.e. specificity 84%). Based on treatment status only and by considering EPRs as a superior source, self-reported history of BP-lowering medication would classify 10.5% of the high risk population as hypertensive though they were not treated by EPRs, and overestimate the exposure to BP-lowering medication by 25%. However, some people were already classified as hypertensive based on their BP values. Hence, the proportion of participants who were classified as hypertensive only because they indicated a treatment in PCQ was 4.7% (i.e. represented in Figure 4-1 by area with 109 and 135

participants). These findings are based on the assumption that EPRs is a gold standard or an “alloyed gold standard”, which is typically assumed in studies that verify drug exposure by self-report.[205] In the UK, blood pressure lowering therapy can only be obtained by prescription and through the pharmacy. Patients cannot access BP lowering drugs over the counter. Thus pharmacy refill data are a good indicator of antihypertensive therapy, though not an indicator of true drug adherence. The study of agreement or concordance between the medication exposure sources revealed “good agreement” with a total agreement between the two sources in 86% of the observations and a kappa value of 71% (P_{pos} =82% and P_{neg} =89%).

The influence of the questionnaire structure in concordance showed that the open-ended question (PCQ-1) provided a better sensitivity to self-report BP lowering medication, but the closed-ended question (PCQ-2) provided a better agreement with EPRs. This finding may be explained by the differences in the nature of the two PCQ phases, especially as the procedure of assignment to the corresponding treatment status was different between the two PCQ. For instance, participants who completed PCQ-1 were assigned a treatment status by the investigator (i.e. myself) if they wrote at least one of the BP-lowering medications; whereas, participants who completed PCQ-2 were assigned to a treatment status based on their awareness of their medication type. In the latter case, a participant prescribed one of the BP-lowering medications in our list (e.g. doxazosin) but for a diagnosis other than hypertension (e.g. benign prostatic hyperplasia) would answer “No” to the question “do you regularly takes any BP-lowering medication?”. These participants were taking a drug with BP-lowering effects, but they were unaware of there medication proprieties as they were prescribed for an indication other than hypertension.

The agreement of SRMs and medical records has been examined in several studies. Generally, several factors were identified that influence the accuracy of recalling a medication in self-report questionnaire including type of medication, frequency of drug administration, questionnaire design, and the respondent characteristics.[183,205,206,218,219] For CVD drugs and particularly antihypertensive drugs, self-reported methods were considered a reliable source to assess drug exposure with a sensitivity ranging from 87% to 94%.[179,220,221] This is in

accordance with findings that medication for chronic diseases or serious conditions that are used in a regular basis are better recalled in the questionnaire, and have higher agreement with pharmacy records compared to drug used as needed or for shorter period.[179,222] Furthermore, the higher sensitivity of antihypertensive medications recall can be extended to a longer time window, as the sensitivity was only decreased by 2% (from 92% to 90%) when increasing the exposure time window from 6 months up to two years.[221]

The influence of question structure and the questionnaire design on the recall accuracy was also evaluated in several studies. The method of addressing this matter was substantially different between these studies, with heterogeneity in the methodological approaches, populations studied, drugs evaluated, and in presentation of results. Yet, the general finding is that recalling drugs is improved when drug names or indications are used instead of an open-ended question.[223] For instance, Klungel et al. has examined the impact of a specific-indication and an open-ended question on drug recall accuracy on 372 hypertensive subjects.[204] This was performed by designing the questionnaire to have five specific-indication questions, including a question about BP-lowering medication, followed by an open-ended question to write names of medication not already mentioned. The sensitivity of the specific-indication question for antihypertensives was 90.6% compared to 16.7% for the open-ended question, and the general sensitivity for the specific-indication question was almost twice as high as for the open-ended question. For all the assessed medication classes, a direct recall with a specific indication had much better sensitivity (88%) compared to the open-ended questions (41%).

The analysis of the concordance between SRM and EPR is very important due to the differences in the type of recrods of medication-exposure between participants. It is very important to assess the reliability and concordance of each source before combining all the results together, and defined participants as treated or not. Generally, this analysis has showed that information from the differernt medication-histroy sources can be combined together in GS:SFHS. A strength of this analysis is the large sample size of the participants. In similar studies, the method of assessing the agreement is usually performed by reporting the kappa statistics. However, kappa statistics suffers from some

limitations and is influenced by the prevalence of the outcome.[211] Hence, the raw numbers and sensitivity analysis were reported here to enable a better interpretation of the results. Furthermore, the analysis was restricted to the high-risk population to reduce the number of normal participants who could have skewed the results by showing higher agreement (results using the total population is shown in Appendix 2 -p251). In the other hand, this approach has underestimated the specificity and NPV due to targeting the high risk population. A number of difficulties arise in performing the analysis that ranged from the study design to choosing the best methods for analysis. First, the EPRs were available for about two-thirds of the participants only, making the analysis only possible for those participants with two medication-history sources. Second, subjects were mutually exclusive for one of the two PCQ phases. Hence, a comparison between the two question structures (i.e. open-ended to closed-ended) in the same participant was not possible, and a general approach was taken to compare the overall agreement between PCQ-1 and PCQ-2, with EPRs was taken. Third, the influence of participants' characteristics on the recall accuracy was not examined in this analysis as this aimed to assess the overall quality of PCQ.

4.5.2 Prevalence of hypertension-related health indicators

This analysis showed that the overall prevalence of hypertension in GS:SFHS was 40.8%, with lower prevalence of hypertension in women (35.8%) compared to men (46.3%). These numbers are higher than the reported prevalence of hypertension in the SHeS for the years of 2008/2009, and 2010/2011; in which average prevalence of hypertension for the two surveys in overall, women and men were 32.7%, 31.7%, and 33.7%, respectively (discussed in Section 1.3.5 -p38).[66] The comparison was made with these two surveys because it covers the time of recruitment and data collection for GS:SFHS participants. The higher prevalence of hypertension in GS:SFHS compared to the Scottish population is in concordance with the full GS:SFHS cohort profile description reported by Smith et al., though their hypertension definition was based only on BP measurement (BP >140/90 mmHg) regardless of treatment status.[171] Compared with other populations, the prevalence of hypertension in GS:SFHS lies well within the range of prevalence of hypertension (31–60%) reported in other European countries.[224] However, it is higher than prevalence reported in the USA (29.1%)

and Canada (19.5%).[225] Furthermore, comparing the population of GS:SFHS with populations of Scotland, England, USA, and Canada showed that both hypertension prevalence and average of BP levels by groups are higher in GS:SFHS (figures are shown in Appendix 3 -p252 and Appendix 4 -p253). Several factors may contribute to the higher prevalence of hypertension and BP levels in GS:SFHS compared to other populations. First, the measurements used for GS:SFHS are based on averaging two available readings from a single visit, whereas most other studies are based on averaging the second and third measures.[225] Second, the prevalence of main risk factors of hypertension differs between these populations, which may largely contribute to the variation in hypertension prevalence. This would include differences in study design, BP measurement methods, and clinical systems between these countries.

The prevalence of hypertension in GS:SFHS was higher in men, older people, obese, most deprived area of Scotland, less educated participants, and in non-smokers participants. These findings are consistent with the results of several others studies in different countries.[226] Moreover, the prevalence of hypertension in participants with two hypertensive parents was much higher than participants without any hypertensive parents, with an intermediate prevalence for participants with only one hypertensive parent. The multi-level mixed model also showed that participants with one or two hypertensive parents have 42% or 2 folds higher risk of having hypertension compared to participants without any hypertensive parents. The association of parental history of hypertension with BP level and development of hypertension in offspring were also reported in the John Hopkins Precursors Study; in which, men with both parents with hypertension or with one hypertensive parent were at higher risk of developing hypertension.[42] This emphasizes the important of taking the family history of hypertension and reviewing of all eligible family members for hypertension status and treatment if required.

The rate of hypertension awareness and treatment in GS:SFHS were 25.3% and 31.2%, respectively; these are the lowest rates compared to the reported rates in other European and North American population-based studies.[62,63,225,226] These rates are comparable to rates reported by Lloyd-Sherlock et al. in LIC and MIC.[67] For awareness, a comparison with SHeS is not possible as the rate of awareness

was not reported in any edition. However, the 2012/2013 survey compared the prevalence of doctor-diagnosed hypertension (i.e. informed from participants) with the survey-defined hypertension, which can be used as an approximate estimate of hypertension awareness as following; the prevalence of doctor-diagnosed hypertension was 22.8% compared with 29.1% survey-defined hypertension, suggesting that the overall rate of awareness was 78% (22.8/29.1), based on the assumption that all the doctor-diagnosed hypertensive were also defined as hypertensive by the survey.[66] Similarly, the rate of treated hypertension in GS:SFHS was much lower than the prevalence of treated hypertension reported in the SHeS of 2008/2009 and 2010/2011 (49% and 48%, respectively).

The lower rate of awareness and treatment in GS:SFHS may be explained by the fact that almost 50% of hypertensive participants were classified as hypertensive based on their BP values only (Figure 4-1). In other words, their average BP level was 140/90 mmHg or higher, but they were not on BP-lowering medication. Moreover, participants were classified as hypertensive based on a single visit and two BP measurements, which is not the criteria for a clinical diagnosis of hypertension which requires multiple BP measurements on different occasions or using ambulatory BP monitoring.[215] The rate of awareness and treatment was higher in women, overweight, parental history of hypertension, greater deprivation, and in less educated persons. This highlights that people with any of these risk factors are more aware of their hypertension status, and were more likely to seek a clinical advice, or these individuals have higher comorbidity levels requiring greater engagement with doctors and hence more likely to be diagnosed by a clinician. On the other hand, people without these risk factors were more likely to have undiagnosed hypertension.

The control rate of hypertension in GS:SFHS was 54.3%, with higher rates in women (60%) compared to men (47.8%). This rate is slightly higher than the reported controlled rate in SHeS for the years of 2008/2009 (53%) and 2010/2011 (50%).[66] In contrast to the findings in SHeS, the rates of controlled hypertension in women in GS:SFHS were higher than men. Compared to other populations, the prevalence of controlled hypertension is one of the highest out of the European countries, but lower than observed in North American countries.[47,62,63] Control

rate was higher in younger groups compared to older groups; for instance, all the treated young women (<25 years) were controlled compared to only 30% of women above 75 years old. This is opposite to what has been observed in other populations, such as Portugal, England, USA, and Canada.[47,226] Controlled hypertension was not statistically associated with number of parents with hypertension, though there are a trend of lower proportion of controlled hypertension in participants with at least one hypertension parent. Similar to the pattern of awareness and treatment, control rate was higher in the most deprived area compared to the least deprived area. This is might be explained by a higher risk factor burden and hence more follow-up and treatment. This is especially true as a reduction in inequality of care between socio-economic strata have been observed in England and Scotland.[227,228] Indeed, Hammouche et al. has reported a higher achievement of Quality and Outcome Framework (QOF) – indicator 4 (QOF4) (BP <150/90) in the most deprived area of Scotland compared to the least deprived.[227] Moreover, participants in the most deprived area were 34% more likely to receive a lifestyle advice compared to participants in the least deprived area.

An important consideration for the association between hypertension-related health indicators and deprivation is that these were only observational findings, and that more works are needed to explore these findings. Several potential confounders can play role in such findings, leading to differences in the prevalence between the most and the least deprived area of Scotland. For this analysis, SIMD was used as index for the deprivation status but other confounders such as alcohol consumption, smoking, physical activity, or BMI were not accounted for in this analysis.

4.5.3 Conclusion

This chapter has shown that SRMs history can be used as a surrogate to assess the exposure to BP-lowering medication in GS:SFHS participants. Although both sources suffer from some limitations, they can be considered the best available source to estimate the drug exposure history. Hypertension management indicators in GS:SFHS are lower than similar reported studies, with the exception of prevalence of controlled hypertension, which achieved a better rate compared to other populations. In GS:SFHS, almost half of men and a third of

women had hypertension based on the study definition, and only one of four (25%) hypertensive participants were aware of their condition. Furthermore, about one-third of the hypertensive participants were treated, and more than half of the treated hypertensives were controlled. Interestingly, participants with parental history of hypertension were significantly more likely to be treated and aware of their hypertension compared to those without any parental history of hypertension. Therefore, a history of hypertension should instigate a review of all eligible family members for hypertension status and treatment if required.

5 Heritability and familial aggregation of BP and hypertension

5.1 Introduction to this chapter

This chapter focuses on examining the familial aggregation of BP and hypertension traits in GS:SFHS. Such studies are performed to estimate the genetic determinants of a trait, and dissect the contribution of genetic and environmental factors. These can be performed either for single traits or as bivariate analysis to assess the shared genetic correlation between two traits.

5.1.1 Familial correlation and heritability of BP

Familial aggregation and correlation of BP are a result of complex interactions between several factors that include genetic and/or shared familial environmental factors. Several studies have reported a higher correlation of BP levels between relatives than others; the Montreal adoption study reported a correlation coefficient of 0.38 between biological siblings, and 0.16 between adoptive siblings.[229] The Victorian Family Heart Study, a study of health adult families that included 783 families enriched with monozygotic and dizygotic twins, demonstrated a correlation coefficient of 0.44 for non-twin siblings, 0.50 for dizygotic siblings, 0.78 for monozygotic siblings, and 0.12 for spouse-spouse pairs for SBP.[230] The study has also reported that the genetic factors accounted for 41% and 46% of the variation in SBP and DBP, respectively. The relative risk of developing hypertension was found to be 4.1 in men and 5.0 in women aged 20–39, if they had two or more 1st–degree relatives diagnosed with hypertension before the age of 55 years.[231] Furthermore, evidence for familial aggregation of treatment and control of hypertension were also demonstrated a significant siblings concordance in both treatment (OR =1.61) and control of hypertension (OR =1.51).[232]

The general assumption in these studies is that siblings and parent-offspring share 50% of their genome, in addition to shared environmental factors. Under random mating assumption, spouses are expected to share familial environmental factors only without any shared genetic factors. Therefore, a significant correlation between siblings and parent-offspring but not between spouses, means that familial aggregation of the traits is primarily resulting from shared genetic factors. Alternatively, if the spouse correlation is significant as

well, then it suggests that the trait is influenced by both genetic and shared environmental factors.

Estimating the heritability components of a trait is another method to quantify familial aggregation. Heritability can be defined as the proportion of total phenotypic variance explained by genetic factors (explained in more details in Section 1.5.3.2 –p56). For BP, several studies have estimated the heritability using different approaches. Generally, the difference between these studies relies on using different family structures (e.g. family-study or twin study), or different BP measurements and traits (e.g. single office BP, ambulatory BP, or long-term BP). For instance, family studies that used single office measurements reported heritability that ranged from 15–40% for SBP, 15–30% for DBP, 33–40% for MAP, and 18–24% for PP.[233-236] However, twin studies have reported higher heritability estimates for BP traits in monozygotic twins that ranged from 48–74%, and 51–72% for SPB and DBP, respectively.[237] Other studies have also taken another approach by using ABMP instead of the office BP, knowing that ABPM is a better predictor for CVD end-organ damage.[24,25] Heritability in these studies were around 70% and 68% for SBP, and 70% and 64% for DBP at day-time and night-time, respectively.[234,238-240] These findings emphasize the multifactorial and complexity of BP traits, which exhibit a complex mode of inheritance.

5.1.2 Sibling recurrence risk ratio (λ_S)

λ_S is an alternative approach to estimate the familial aggregation of the trait by calculating the probability that a sibling of an affected person is also affected (previously discussed in Section 1.5.3.1 -p55). This measure is widely used in genetic studies as a supporting evidence for the role of genetic factors in the disease or trait, and has an important meaning in genetic counselling of the family of affected individuals. λ_S for hypertension was found to be around 1.2 – 3.5, indicating a trait with modest genetic effects.[131,241]

Typical genetic studies that estimate λ_S obtains the sibship or family data by recruiting non-randomly selected family based on certain criteria, such as families of affected individuals who attended the clinic. Such a way of sampling or ascertainment is performed to maximize the number of affected individuals with lower cost and effort. However, this procedure can produces a bias known

as “ascertainment bias”, which can alter the expected ratio of affected to unaffected siblings, and hence result in a potentially biased estimate of λ_S.[242] One way to avoid the ascertainment bias is by recruiting a random sample or a census of sibships from the population of sibships with at least one affected individual.[243] This ascertainment scheme is called “complete ascertainment”, where every sibship with at least one affected member has an equal probability of being ascertained, independent of the number of affected or the total size of the sibship.[243] Essentially, it refers to the method in which family are firstly ascertained and then the family members are examined for the presence of the disease of interest. Thus, it is not required to ascertain all the affected individuals in the population to achieve complete ascertainment, providing that every sibship in the population has an equal ascertainment probability. The original pedigree and participants of GS:SFHS study were not ascertained via any particular phenotypes, and were recruited from the general population given that another family member is willing to participate. Hence, no ascertainment correction is required for this study.[243]

5.2 Chapter aims

The aims of this chapter are to estimate the familial aggregation of BP and hypertension traits by;

(1) Estimating the familial correlation of BP traits

(2) Calculating the univariate heritability of BP traits, and bivariate genetic correlation between BP traits

(3) Calculating the λ_S of hypertension and hypertension treatment

5.3 Statistical analysis

This study used the full sample of GS:SFHS (n =18,470), which includes individuals described in the previous chapters. In particular, Section 3.2 described the method of phenotype QC. Unlike the descriptive analysis performed in the previous chapter, this chapter has used the adjusted BP values (as described in Section 2.3.1.1-p98), by adding a fixed values of 15 mmHg to

the measured SBP value and 10 mmHg to the observed DBP value, for participants taking BP-lowering medications. This follows the recommendation of Tobin et al. to adjust for the treatment effect for quantitative trait analysis.[87] The means of BP traits per family was calculated only in families with at least three members, and data is shown as a ranked mean from the lowest to the highest across the families. The following sections describe the methods used in this chapter to assess the familial aggregation and heritability of BP and hypertension.

5.3.1 Estimate of familial correlation of BP traits

The family structure was assessed using two methods. The first method is based on graphically assessment of the familial correlation of BP and hypertension traits using the graphical function in PEDSTATS software.[195] PEDSTATS was firstly used to verify the structure of each family (as performed for checking genotypes Mendelian inconsistency in Section 2.4.2.4 -p114), and to analyse the generic behaviour of the variables among the different members of the family, and to estimate the number of and types of available relative pairs. The second method is based on calculation of the familial correlation coefficients (r^2) among different types of family pairs. This was performed by the family correlation (FCOR) function in S.A.G.E (Statistical Application for Genetic Epidemiology).[244]

FCOR considers all pairs of relatives if both members have at least one trait in common, and excludes the pairs where one of the members has missing data. An available option in the function is to calculate two types of correlation that are “main type” and “subtypes”; the main type correlation does not depend on the sex of the pairs, for example it will only produce one correlation coefficient value of parent-offspring pair, by ignoring the sex of the pairs. The subtype’s correlation depends on the individuals’ sex, for the previous example, the parent-offspring main type has four subtypes; father-son, mother-son, father-daughter, and mother-daughter. FCOR always calculates the subtype correlation first and has the option to calculate the main type if requested by recalculating the correlation coefficients by ignoring sex. FCOR can also output the Chi-square statistics and p-values of the test for homogeneity of correlation among subtypes, in which the hypothesis that all subtypes within a main type have the same correlation. The familial correlation values were calculated for two

models; the first model was not adjusted for any covariate, and the second model was adjusted for age, age^2, sex, and BMI. The residualization values in the second model was calculated by the software Sequential Oligogenic Linkage Analysis Routines (SOLAR v. 7.2.5), as part of heritability calculation that is explained in more details in the next section.

5.3.2 Calculating the univariate heritability of BP traits

Univariate heritability and bivariate (genetic and environmental correlation) were estimated using standard quantitative genetic variance-component model implemented in SOLAR software (SOLAR v. 7.2.5).[245] The variance-components model is a classical quantitative genetic model that decomposes phenotypic variance (σ^2_P) into additive genetic components (σ^2_G) and non-genetic components (i.e. environmental; σ^2_E), as previously explained in Section 1.5.3.2 -p56; this model assumes that $\sigma^2_P = \sigma^2_G + \sigma^2_E$. Hence, the proportion of phenotypic variance attributable to additive genetic factors (h^2) can be calculated as [$h^2 = \sigma^2_G / \sigma^2_P$], and the proportion of variance attributable to environmental factors as [$e^2 = 1 - h^2$]. The heritability estimates represented here refers to “narrow-sense heritability”, which is concerned with the additive genetic component rather than all the genetic components such as dominance.

The univariate heritability of BP traits were estimated for three models; the first model was not adjusted for any covariates; the second model was adjusted for age, sex, age^2; and the third model was adjusted for the same covariates as the second model in addition to BMI. Including covariates in the model means that the model will firstly account for the effects of the covariates, and then estimates h^2 and e^2 from the remaining residual phenotypic variance (i.e. σ^2_P equals phenotypes after accounting for the effect of the covariates rather than the total phenotypic variance). The contribution of the genetic factors in these models is estimated by multiplying the heritability estimate (h^2) by the total variance not explained by the covariates [1- Variance explained by covariates]. This to rescale the phenotypic variance to exclude the variance explained by covariates.

A rank-based inverse-normal transformation was performed to the BP traits values before estimating heritability to obtain a normal distribution of the

residuals. The normalization feature is included in SOLAR to improve the normality of the traits while retaining much of the original information, as the inverse normalized variable will have a mean near zero and a SD near 1.0, with a distribution that approximates normal.

5.3.3 Calculating the bivariate genetic correlation for BP traits

An extension to the univariate heritability analysis to encompass the multivariate state is possible by SOLAR. This analysis identifies the origin of correlations between each pair of traits (e.g. SBP and DBP), and partitions the correlation into additive genetic correlation (ρ_G), and environmental correlation (ρ_E). A value of 1 or -1 for ρ_G between two traits imply that all the additive genetic factors that influence trait X will also influence Y (i.e. complete pleiotropy). Similarly, a significant value of ρ_G that is different from both 0 and 1 or -1 is evidence of incomplete pleiotropy (i.e. some of the phenotypic similarity between two traits are due to the same gene or genes). Hence, the phenotypic correlation (ρ_p) between the two traits can be dissected into genetic and environmental constituents. For example, ρ_p for SBP and DBP can be calculated as by this equation;

$$\rho_{p(SBP*DBP)} = \left(\sqrt{h^2_{SBP}}\,\sqrt{h^2_{DBP}}\,\rho_G\right) + \left(\sqrt{(1 - h^2_{SBP})}\,\sqrt{(1 - h^2_{DBP})}\,\rho_E\right)$$

The contribution of the shared genetic factors to the phenotypic correlation between two traits is called the bivariate heritability, and can be calculated by taking the square root of both univariate heritabilities multiplied by the genetic correlation (i.e. the first part of the above equation).

5.3.4 Estimate of sibling recurrence risk of hypertension and treated hypertension

Sibling recurrence risk (κ_S) is the probability that a sibling of an affected individual is also affected. λ_S is calculated by dividing κ_S by the population prevalence of the outcome (i.e. the outcomes for this analysis are hypertension and treated hypertension). In this study we adopted the method proposed by Olson and Cordell to estimate λ_S in siblings of hypertensive patients compared

with the background population.[243] This method is appropriate for a population cohort in which participants are not ascertained via any particular phenotypes and under a complete ascertainment sampling method, in which λ_S is assumed to be unbiased and consistent.[243] This method is applicable here as the original pedigrees were not ascertained via any specific BP phenotype, leading to the use of the equation proposed by Olson and Cordell when proband status is unknown;

$$\kappa_s = \frac{\sum_{s=1}^{\infty}\sum_{a=1}^{s} W_a(a-1)n_{s(a)}}{\sum_{s=1}^{\infty}\sum_{a=1}^{s} W_a(s-1)n_{s(a)}} \quad \text{---} \quad \text{Equation (1)}$$

where w_a is the weight given to a sibship with (a) affecteds, and $n_{s(a)}$ is the number of sibships in the population of size (s) with (a) affected. For randomly or completely sampled siblings such as GS:SFHS, w_a is equal to (a), which provide an estimate of the proportion of affected among all sibling of affected in the sample. The denominators in the λ_S (i.e. population prevalence) for these outcomes were based on the reported prevalence in the SHeS 2010/2011 (reported in Section 1.3.4 – p 38).

5.4 Results

The number of extended-families (i.e. pedigrees) was 7,025 with an average size of 2.6 members per family (SD =2.02), and range from a singleton family to a family of 29 members. The number of available sibships was 4,616 sibships, which ranged from two to eight members per sibships. The full description of family structures was previously described in Section 3.3.5 -p122. The distribution of BP traits per family is showed in Figure 5-1, with mean values that ranged from 102–177 mmHg for SBP, 60–111 mmHg for DBP, 76–134 mmHg for MAP, and 31–83 mmHg for PP.

5.4.1 Familial correlation of BP traits

The familial correlation coefficients (r^2) for the different types of family pairs are presented in Figure 5-2. The familial correlation coefficients are shown for the two types (as it presented by S.A.G.E output); that is the subtypes in Figure 5-2A, and main types in Figure 5-2B. The correlation coefficients for all the possible family relative pairs including second-degree relative are listed in Table 5-1.

For SBP, the highest familial correlation was observed for the sister-sister pair (0.38), with very close correlation coefficients for the brother-brother pair (0.37). After adjusting for the covariates, the correlation coefficients slightly decreased and brother-brother pairs had the highest correlation (0.29). For DBP, the highest correlation coefficient was observed for brother-brother pairs in both models (0.36 and 0.23, for model 1 and 2, respectively). The change in correlation values between spouses in the two models was very small, from 0.1 in model 1 to 0.09 in model 2. Spouse correlation was higher for both SBP and DBP in the unadjusted model, but reduced by half for SBP and by only 0.01 for DBP in the adjusted model to have a same value of 0.09.

For MAP, the brother-brother pairs had the highest correlation coefficients in both models (0.38 and 0.27, for model 1 and 2, respectively). The correlation of spouses has slightly decreased from 0.13 in the first model to 0.09 in the second model. For PP, the sister-sister pairs had the highest correlation value in the first model (0.37), but brother-brother pair had the highest correlation value

after adjusting for the covariates in the second model (0.25). The correlation between son and both father and mother were very small and not significant in the first model, but after adjusting for the covariates the correlation values increased slightly and became significant. The highest observed correlation among the four BP traits for spouse pairs was for PP in the first model (0.23). Yet, adjusting for the covariates in the second model reduced the correlation.

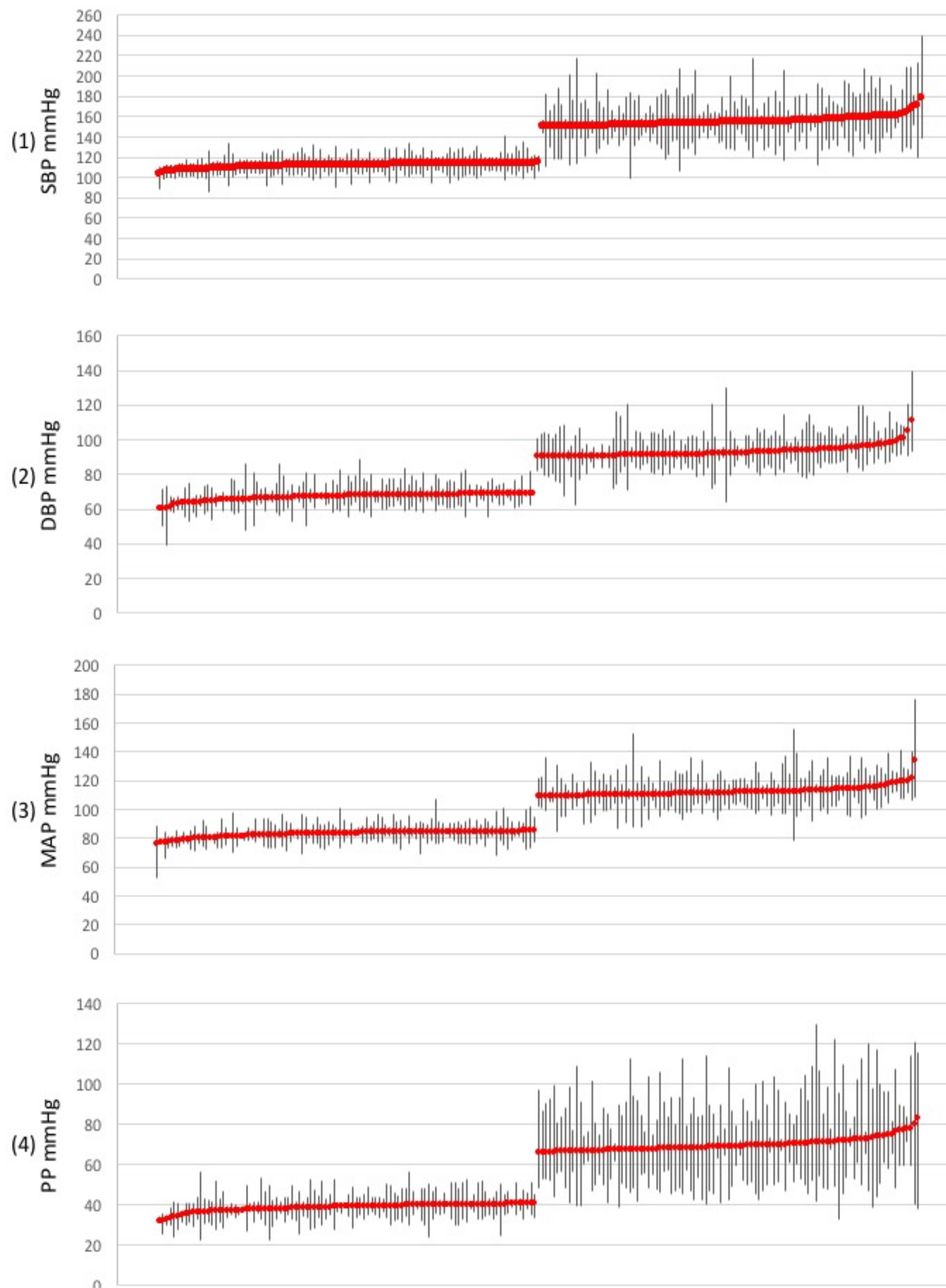


Figure 5-1 Mean values of BP traits per family.
Red points represent the mean values of (1) SBP, (2) DBP, (3) MAP, and (4) PP for 100 families from the lower and 100 families from the upper extreme tails of the distributions. Only families with at least three members per family were considered in the distributions. The vertical lines show minimum and maximum values within each family.

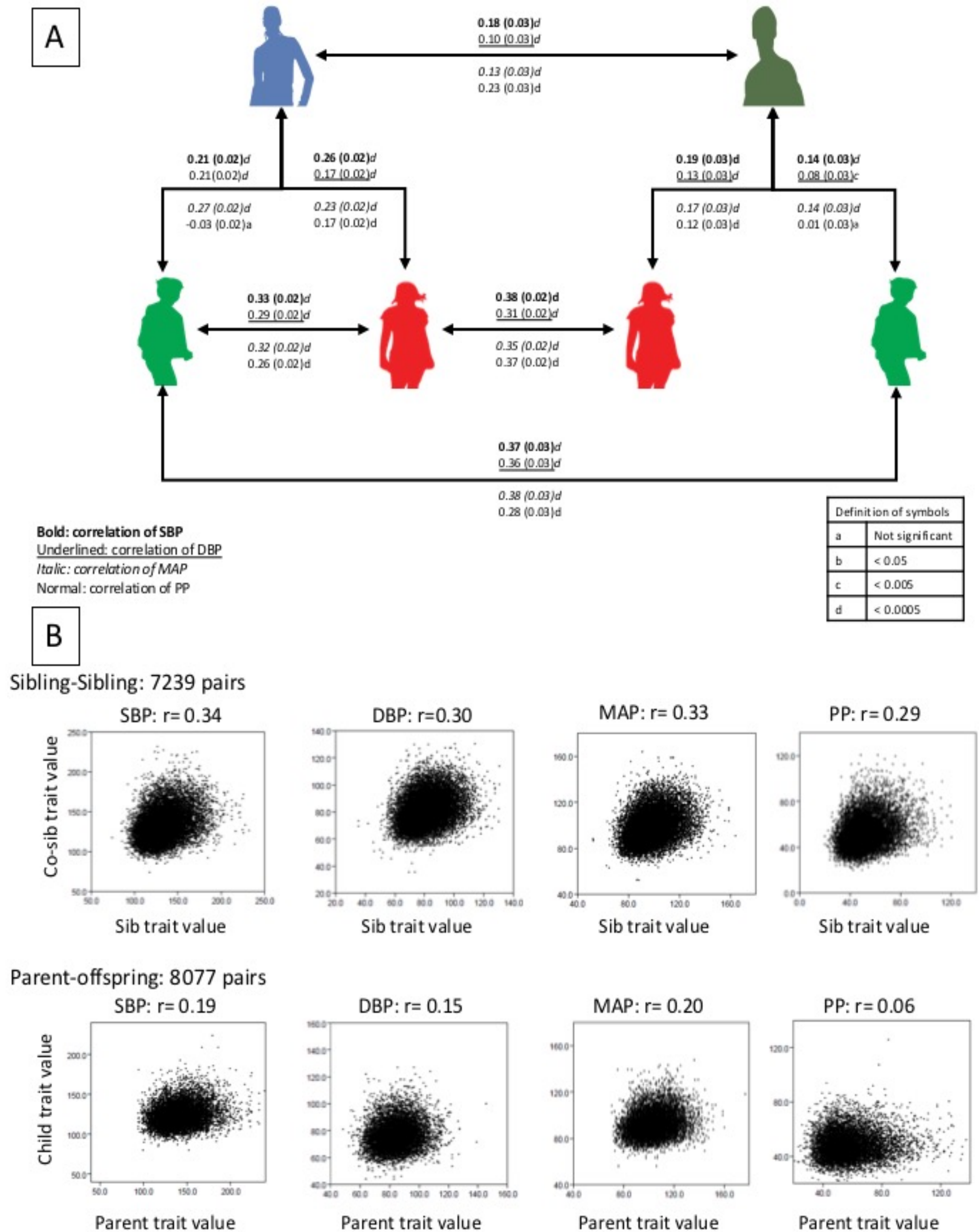


Figure 5-2 Familial correlation coefficients for all the four BP traits.
(A) The correlation coefficients of the subtype pairs as represented by a pedigree of a nuclear family. The correlation coefficients of BP traits are written in the middle of the arrows between each pairs, in the forms of r (se) with a symbol that indicates the p-value for the correlation. The BP traits are SBP, DBP, MAP, PP ordered from top to bottom. (B) Relative pair plots of sibling-sibling, and parent-offspring with correlation coefficients for BP traits (SBP, DBP, MAP, and PP, as ordered from left to right). Values for pair types are represented by a single data point formed by plotting the first sibling or the parent along the x-axis, and the second sibling or the offspring along the y-axis. For all plots, the correlation coefficients are shown above the plots.

Table 5-1 Familial correlation coefficients for different relative pairs

Pairs	#of pairs	Model	SBP			DBP			MAP			PP		
			r	se	*p*	r	se	*p*	r	se	*p*	r	se	*p*
Parent-Offspring	8077	1	0.19	0.01	<0.0001	0.15	0.01	<0.0001	0.20	0.01	<0.0001	0.06	0.01	<0.0001
		2	0.11	0.01	<0.0001	0.13	0.01	<0.0001	0.13	0.01	<0.0001	0.07	0.01	<0.0001
Father-Son	1419	1	0.14	0.03	<0.0001	0.08	0.03	0.003	0.14	0.03	<0.0001	0.01	0.03	0.83
		2	0.09	0.03	0.0021	0.07	0.03	0.01	0.07	0.03	0.0079	0.09	0.03	0.0012
Mother-Son	2042	1	0.21	0.02	<0.0001	0.21	0.02	<0.0001	0.27	0.02	<0.0001	-0.03	0.02	0.17
		2	0.13	0.02	<0.0001	0.17	0.02	<0.0001	0.17	0.02	<0.0001	0.05	0.02	0.033
Father-Daughter	1718	1	0.19	0.03	<0.0001	0.13	0.03	<0.0001	0.17	0.03	<0.0001	0.12	0.03	<0.0001
		2	0.12	0.03	<0.0001	0.14	0.03	<0.0001	0.14	0.03	<0.0001	0.08	0.03	0.0037
Mother-Daughter	2898	1	0.26	0.02	<0.0001	0.17	0.02	<0.0001	0.23	0.02	<0.0001	0.17	0.02	<0.0001
		2	0.13	0.02	<0.0001	0.12	0.02	<0.0001	0.13	0.02	<0.0001	0.11	0.02	<0.0001
Sibling-Sibling	7239	1	0.34	0.01	<0.0001	0.30	0.01	<0.0001	0.33	0.01	<0.0001	0.29	0.01	<0.0001
		2	0.21	0.01	<0.0001	0.20	0.01	<0.0001	0.21	0.01	<0.0001	0.18	0.01	<0.0001
Brother-Brother	1256	1	0.37	0.03	<0.0001	0.36	0.03	<0.0001	0.38	0.03	<0.0001	0.28	0.03	<0.0001
		2	0.29	0.03	<0.0001	0.23	0.03	<0.0001	0.27	0.03	<0.0001	0.25	0.03	<0.0001
Brother-Sister	3256	1	0.33	0.02	<0.0001	0.29	0.02	<0.0001	0.32	0.02	<0.0001	0.26	0.02	<0.0001
		2	0.17	0.02	<0.0001	0.18	0.02	<0.0001	0.18	0.02	<0.0001	0.13	0.02	<0.0001
Sister-Sister	2727	1	0.38	0.02	<0.0001	0.31	0.02	<0.0001	0.35	0.02	<0.0001	0.37	0.02	<0.0001
		2	0.20	0.02	<0.0001	0.22	0.02	<0.0001	0.21	0.02	<0.0001	0.20	0.02	<0.0001
Half Sib.	134	1	0.24	0.09	0.01	0.19	0.09	0.04	0.23	0.09	0.02	0.20	0.09	0.03
		2	0.16	0.09	0.08	0.07	0.09	0.43	0.09	0.09	0.34	0.21	0.09	0.02
Grandp.	654	1	-0.05	0.04	0.27	-0.03	0.04	0.50	-0.03	0.04	0.50	-0.07	0.04	0.09
		2	0.02	0.05	0.72	0.03	0.04	0.41	0.03	0.04	0.47	-0.04	0.05	0.45
Avunc.	5288	1	0.09	0.02	<0.0001	0.08	0.02	<0.0001	0.11	0.02	<0.0001	0.00	0.02	0.82
		2	0.05	0.02	0.0014	0.06	0.02	0.0001	0.06	0.02	<0.0001	0.02	0.02	0.32
Cousin	1729	1	0.08	0.03	0.01	0.14	0.03	<0.0001	0.12	0.03	0.0001	0.03	0.03	0.37
		2	0.00	0.04	0.94	0.03	0.03	0.26	0.03	0.03	0.33	0.01	0.03	0.77
Spouse	1189	1	0.18	0.03	<0.0001	0.10	0.03	0.0005	0.13	0.03	<0.0001	0.23	0.03	<0.0001
		2	0.09	0.03	0.0034	0.09	0.03	0.0023	0.09	0.03	0.0014	0.06	0.03	0.0309

Model 1 is the basic model that includes BP without any covariates; model 2 was adjusted for age, age2, sex, and BMI, (Avunc.: avuncular),

5.4.2 Univariate heritability of BP traits

The estimated univariate heritabilities for the four BP traits in the population of GS:SFHS are presented in Table 5-2. The heritability estimates (h^2) ranges from 12% for PP for the unadjusted model to 35% for MAP for the second model, and all heritability were significant (p <0.0001). In the two models (model 2 and 3) that were adjusted for covariates, h^2 represents the total additive genetic effect for the correspondence trait, which is different from the total genetic variance for the trait. For SBP, the additive genetic effect of genes (h^2) accounts for 33% of the variation in SBP that is not explained by the covariates, with 29% of the variance being explained by the covariates. Hence, genetic factors account for 23% of the total variance in the SBP trait. After further adjusting for BMI, the estimate of heritability reduced to 30% and the variance explained by adding BMI to covariates increased to 37%, indicating that only 19% of the total variance of SBP is explained by genetic components. Since BMI also has genetic components, this indicates that 19% of the total variance of SBP in GS:SFHS is explained by unknown genetic factors that contribute to SBP variance independent of the BMI.

For DBP, the heritability in the second model was 34%, and in the third model was 31%, suggesting that genetic factors explain 28% and 22% of the total variance of DBP in the second and third models, respectively. Estimates for MAP and PP are shown in Table 5-2, and can be interpreted in the same way.

5.4.3 Bivariate genetic correlation

Table 5-3 shows the bivariate genetic correlation and heritability of BP traits. All correlations were positive and significant. The correlation between DBP and MAP showed the highest phenotypic correlation (94%), genetic correlation (93%), and environmental correlation (93%), with a bivariate heritability equals to 31%; 33% of the total phenotypic correlation between DBP and MAP is due to shared genetic factors. Although the phenotypic correlation between DBP and PP was only 10%, almost 60% of this correlation can be explained by shared genetic factors. All the genetic correlations were significantly different than both 0 and 1 or -1, suggesting incomplete pleiotropy between these traits (i.e. some of the phenotypic similarity between two traits are due to the same gene or genes).

The proportions of phenotypic correlation that is explained by genetic factors between BMI and all the four BP traits were exactly the same (54%), despite the variation in the value of phenotypic correlations, which range from 36% to 10%.

5.4.4 Sibling Recurrence Risk ratio (λ_S)

The number of individuals with at least one sibling presented in the analysis was 10,376, making a total of 4,616 sibships. Calculating the K_S for hypertension was based on a total of 2,548 informative sibships (55%; i.e. contain at least one hypertensive sibling) (Table 5-4). The proportion of sibships with at least two hypertensive siblings was 22%. The number of hypertensive individuals was 2,938 among a total of 5,461 siblings, resulting with an estimate of K_S that equals 54% (2,938/5,461), and λ_S for hypertension was then calculated as 1.6 (0.54/0.32), based on the hypertensive prevalence reported in SHeS (32%). The total number of informative sibships for hypertension treatment was 1,030 (22%), and 78% of the sibships did not have any sibling who was taking any BP-lowering medication (Table 5-5). K_S for hypertension treatment was estimated by dividing the total number of treated siblings 616 by the total number of informative siblings 1,883, resulting a K_S estimate of 33%, and an estimate of λ_S for hypertension treatment was 2.04 (0.33/0.16).

Table 5-2 Univariate heritability of BP traits

Model	Trait	h^2	SE	*P*	Var. by Cov.[a]	Genetic factors[b]	Covariates
1	SBP	0.17	0.02	6.95E-30	-	-	None
2		0.33	0.02	2.44E-90	0.29	0.23	Sex, age, age2
3		0.30	0.02	2.39E-78	0.37	0.19	Sex, age, age2, BMI
1	DBP	0.24	0.02	1.74E-54	-	-	None
2		0.34	0.02	1.57E-99	0.18	0.28	Sex, age, age2
3		0.31	0.02	8.61E-85	0.28	0.22	Sex, age, age2, BMI
1	MAP	0.22	0.02	1.09E-45	-	-	None
2		0.35	0.02	1.49E-104	0.24	0.27	Sex, age, age2
3		0.32	0.02	9.62E-90	0.34	0.21	Sex, age, age2, BMI
1	PP	0.12	0.02	1.11E-14	-	-	None
2		0.24	0.02	7.35E-51	0.26	0.18	Sex, age, age2
3		0.24	0.02	3.02E-49	0.27	0.17	Sex, age, age2, BMI

a. Var. by Cov is the proportion of the total variance that is explained by the covariates in the model.
b. The contribution of the genetic factors to the total variance of the trait, which is calculated as: [(1 – Var. by Cov.) x h^2].

Table 5-3 Phenotypic, genetic, and environmental correlations for BMI and BP traits

Trait pairs	ρ_p	ρ_G (se)	ρ_E (se)	Bivariate h^{2*}	% Explained by shared genetics**
SBP, DBP	0.72	0.81 (0.02)	0.68 (0.02)	0.25	0.35
SBP, MAP	0.91	0.94 (0.01)	0.90 (0.01)	0.30	0.32
SBP, PP	0.77	0.78 (0.02)	0.78 (0.01)	0.21	0.27
SBP, BMI**	0.32	0.41 (0.03)	0.27 (0.02)	0.17	0.54
DBP, MAP	0.94	0.97 (0.01)	0.93 (0.01)	0.31	0.33
DBP, PP	0.10	0.22 (0.01)	0.06 (0.02)	0.06	0.59
DBP, BMI**	0.35	0.44 (0.03)	0.30 (0.02)	0.19	0.54
MAP, PP	0.41	0.46 (0.04)	0.40 (0.01)	0.13	0.31
MAP, BMI**	0.36	0.44 (0.03)	0.31 (0.02)	0.20	0.54
PP, BMI**	0.13	0.19 (0.03)	0.11 (0.02)	0.07	0.54

* Bivariate h^2 is the contribution of shared genetic factors to the phenotypic correlation between two traits, calculated by this formula [$\sqrt{h_A^2} \times p_{G\ (A.B)} \times \sqrt{h_B^2}$), where A and B are trait one and trait two.
**This column shows the proportion of the phenotypic correlation explained by shared genetics factors, calculated by this formula [Bivariate h^2/ ρ_p].
**The model included sex, age, age^2 as covariates. For the remaining, models included BMI in the covariates in addition to sex, age, and its square.
Abbreviations: ρ_p: phenotypic correlation, ρ_G: genetic correlation, ρ_E: environmental correlation.

Table 5-4 Number of siblings with hypertension per sibship

Size of sibship (s)	Number of hypertensive siblings in the sibship (a)								Total
	0	1	2	3	4	5	6	7	
2	1799	1257		-	-	-	-	-	3724
3	225	221	156	99	-	-	-	-	701
4	37	36	36	21	17	-	-	-	147
5	7	3	12	3	2	4	-	-	31
6	0	2	1	3	3	1	0	-	10
7	0	0	0	0	0	0	1	1	2
8	0	1	0	0	0	0	0	0	1
Total	2068	1520	873	126	22	5	1	1	4616
Affected siblings[1]	0	0	1746	756	264	100	30	42	2938
Total sibling[2]	0	1836	2282	864	296	105	36	42	5461

[1] This row is calculated by multiplying the number in total row by number of affected sibling (a) and number of affected sibling minus one (a-1), to represents the numerator K_S (equation 1).
[2] Total sibling is the sum of number of affected sibling (a) multiplied by the size of sibship minum one (s-1) and number of affected sibling for each sibship size.

Table 5-5 Number of siblings with treated hypertension per sibship.

Size of sibship	Number of sibling taking a BP-lowering medication in the sibship								Total
	0	1	2	3	4	5	6	7	
2	2972	613	139	-	-	-	-	-	3724
3	494	140	49	18	-	-	-	-	701
4	95	34	11	6	1	-	-	-	147
5	19	3	3	5	1	0	-	-	31
6	5	2	1	2	0	0	0	-	10
7	0	2	0	0	0	0	0	0	2
8	1	0	0	0	0	0	0	0	1
Total	3586	794	203	31	2	0	0	0	4616
Affected siblings[1]	0	0	406	186	24	0	0	0	616
Total sibling[2]	0	1029	574	252	28	0	0	0	1883

[1] This row is calculated by multiplying the number in total row by number of affected sibling (a) and number of affected sibling minus one (a-1), to represents the numerator K_S (equation 1).
[2] Total sibling is the sum of number of affected sibling (a) multiplied by the size of sibship minum one (s-1) and number of affected sibling for each sibship size.

5.5 Discussion

In this chapter, I have demonstrated the familial aggregation of BP and hypertension traits in GS:SFHS by using multiple measurements that assess the genetic components of BP. The presence of nuclear family structure with different relative pairs types has allowed me to carried out these analyses. The main findings of this study can be summarised as follows; 1) the familial correlation of BP traits showed higher correlation among first-degree relatives than other types of relative pairs; 2) heritability of BP traits ranged from 24% to 32% with PP having the lowest heritability estimate; 3) bivariate genetic correlation between BP traits showed a high correlation between SBP, DBP and MAP (ρ_G: 81% to 94%), but lower correlations with PP (ρ_G: 22% to 78%); 4) the familial aggregation of hypertension and treated hypertension as a binary trait were 1.6 and 2.04 respectively. In addition, despite the wide variation in the extent of phenotypic and genotypic correlation between BMI and the four BP traits, the proportion of variance that is explained by shared genetic factors was exactly the same (54%).

Studying familial aggregations of BP and hypertension represents an important aspect of genetic epidemiological studies. As heritability estimates are specific to the population studied, this is the first study of the Scottish population. The large number of families with an extended family structure is another advantage that provides more reliability to the estimate reported here. It must be noted that the direct comparison of the results of this study with other studies should be considered carefully due to several factors that are related to differences in study design, sample size, and methods of analysis (i.e. including using of different covariates in the adjusted model). For instance, heritability estimates from twin studies tend to show higher estimates of heritability than studies that include nuclear families. The heritability estimate *per se* is a population-specific measurement, and there were no other studies that looked at the Scottish population particularly, to compare my results against.

Several studies have reported the familial correlation of BP traits, under the theory that relatives share genetic and environmental factors, leading to greater phenotypic similarity between relatives than unrelated individuals. For instance, siblings and parent-offspring have a kinship coefficient of 25%, but spouse pairs

under the assortative mating assumption share the environmental factors only. This was also observed in our results in which the correlation between spouse pairs was greater than zero and statistically significant, suggesting the influence of shared environmental factors. The correlations reported here are within the range of correlation coefficients reported in other populations, such as the Finnish [246], the Portuguese [247], the English [248], and the Chinese Han [249]. Unlike most of these studies, our estimates in the different pairs of nuclear family (i.e. father-son, mother-son, father-daughter, and mother-daughter) were all statistically significant ($p < 0.0001$), except for DBP in father-son pairs, and for PP in both father-son and mother-son. Adjusting for the effect of BMI has reduced the crude familial correlation in all of the possible relatives pairs, but remained significant in most relative-pairs.

Correlation coefficients between parent-offspring pairs were generally smaller compared to sibling pairs, which may suggest a possible age-specific factors or the presence of non-additive genetic effects. This was also observed in other studies that calculated the narrow-sense and broad-sense heritability.[250,251] In which, an important proportion of the trait heritability was attributed to non-additive genetic effects, particularly for older groups. This proportion of non-additive genetics variance may include genetic dominance effects, gene-gene interaction (epistasis), or gene-environment interaction factors that act in non-additive way to produce a different correlation between parent-offspring and siblings pairs. Furthermore, van Dongen et al. suggested that age factor is unlikely to explain the non-additive effects, as they could not find evidence for the presence of genes that act differently in different ages.[250] Alternately, this can be attributed to the fact that siblings are more likely to share more similar environments with each other than they do with their parents.

The spouse resemblance was significant in all the BP traits in the crude model and before adjusting for BMI. The adjusted model showed lower correlation coefficients and was only significant at the nominal p value level for PP, which showed the maximum reduction after adjusting for BMI (from 23% to 6%). This may be an indication of some sort of assortative mating. Indeed, Allison et al. reported evidence of assortative mating for relative weight, which may overestimate the correlation between spouse.[252] Theoretically, adjusting for BMI

should eliminate the effect of assortative mating based on weight, and produce lower correlation coefficients after adjustment, which was observed in our study. The significant decreases observed in PP after adjusting for BMI may be indicative of the important role of BMI on PP. It has been shown that BMI is an independent factor that is associated with decreasing arterial compliance and elevated PP.[253]

The univariate heritability estimates reported here for BP traits were statistically significant and ranged from 24–32%, with PP having the lowest estimate and MAP having the highest estimates, which is in line with two other studies.[233,254] These estimates were in the range of 19–45% for SBP, and 6–43% for DBP which are similar to other reported heritability from family-based studies.[233,234,236,248,255] Our findings confirm the previous reported results of moderate influence of additive genetic factors on the variation of BP traits.

The covariates included in the model accounted for a large proportion of the phenotypic variance (up to 37% in the SBP model that included BMI); hence, the heritability estimates were not inflated by their effects. In line with previous studies, the heritability estimates of BP decreased after adjustment for BMI, which along with the findings from the familial correlation analyses suggest the importance role of BMI, and that the variance in BMI explained large proportion of total phenotypic variation in BP. Wu et al. have suggested that this reduction in heritability estimates after adjusting for BMI is due to a higher influence of common and unique environmental factors with increasing levels of BMI.[256] For instance, several behavioural factors that predict BP level such as unhealthy diet and sedentary lifestyle are more prevalent among groups of higher BMIs. In addition, other factors that can be considered as pure environmental factors such as socio-economic status may also have some genetic basis. Indeed, another study that also used the GS:SFHS cohort has reported a heritability of 41% for education, 54% for intelligence, and 71% for socio-economic factors as represented by SIMD.[257] This can also be attributed to the fact that children share the same environment as their parents, such as they grow up together in the same household. The extent to which the genetic factors of socio-economic might also contribute to the familial correlation of BP has not been studied in twin or family studies.

The bivariate genetic correlation analysis between BP traits indicate that they partly share common genetics factors, but the observed incomplete correlation as represented by findings that ($\rho_G \neq 0$, 1, or -1; p <0.0001), suggest the presence of trait-specific set of genes (i.e. incomplete pleiotropy). Similarly, BP traits were greatly influenced by similar environmental factors, but each trait was still influenced by specific-environmental factors ($\rho_E \neq 0$, 1, or -1; p <0.0001). A high genetic correlation between SBP, DBP and MAP has been observed in two previous studies, which reported similar findings to this thesis.[233,254] However, this thesis found a higher genetic correlation between DBP and PP than one of the studies, which reported an absence of this correlation [233], and a higher phenotypic correlation than the other study, which also reported an absence of this correlation.[254] Nevertheless, the correlation between PP and the remaining BP traits were the lowest in this thesis, suggesting the presence of an independent set of genes that influence PP. This is concordant with animal studies that shows that different loci are involved in regulation of steady-state (DBP) and pulsatile (PP) [258], and the absence of phenotypic correlation between DBP and PP in rats.[259]

The bivariate correlation between BP traits and BMI has showed that the extent of phenotypic correlation that is explained by shared genetic factors is exactly the same (54%). These interesting findings may emphasize the presence of gene(s) that influence both BMI and BP traits by the same way. The genetic correlation between BP traits and BMI in our study ranged from 19–44%, with PP having the lowest correlation and very close estimates for SBP, DBP, and MAP. Our genetic correlation estimates are higher than those reported in Spanish [254], Atherosclerosis Risk in Communities,[160] Victorian Family Heart Study,[260] and Chinese Han population.[256] Recently, two twins studies have examined the genetic and environmental determinates of BP and BMI in Italy and Chinese Han populations.[256,261] The two studies have reported a significant genetic correlation of BMI with SBP (0.29–0.38) and DBP (0.15–0.48). The phenotypic correlations were largely explained by genetic factors, which were 0.82 and 0.86 in the Chinese Han population, and 0.74 and 0.65 in the Italian population for SBP and DBP, respectively. These results are in line with our findings that these traits share some common genetic factors. However, both studies reported that

the proportion of genetic factors that influence both of BMI and BP traits were less than 8%.[256,261]

Analysis of hypertension as a binary trait was performed using λ_S, which is the relative risk of the siblings of hypertensive individuals. GS:SFHS was not ascertained for any particular BP traits, and thus there was not any requirement to correct for ascertainment bias. For hypertension, the λ_S in GS:SFHS was 1.6, which is lower than the λ_S reported by the WTCCC for hypertension (2.5 to 3.5) [131], but within the range reported by Caulfield et al. (1.2 to 1.7)[241]. Calculating λ_S based on the treatment status may provide a more accurate estimate of the clinically diagnosed hypertension, since individuals taking BP-lowering medication are more likely to receive a clinical diagnosis of hypertension than people defined as hypertension based on the study criteria (BP ≥140/90 mmHg, or taking BP-lowering medications). Especially as only 25% of the hypertensive participants in GS:SFHS were aware of their disease (discussed in Section 4.4.6 - p153). The λ_S for treated hypertension (2.04) was higher than hypertension (1.60), suggesting that having a treated hypertensive sibling confers a higher risk of receiving a clinical diagnosis of hypertension. The importance of family history in identify participants at higher risk of developing the disease, or initiating the pharmacological therapy for chronic diseases has not been appreciated.[262] For instance, Daniels et al. reported that having one controlled sibling increases the odds of the other sibling to be also controlled by 51% (95% CI: 25–83%). This led him to suggest that having a complete family history for the patients may enable the physician to switch the drug regimens in the uncontrolled patient to the regimen that is more effective in his sibling.[232]

5.5.1 Conclusion

In summary, this chapter has assessed the familial aggregation of BP and hypertension traits in the Scottish population using one of the largest family-based studies, which include 18,471 individuals and 7,025 extended families. The results are generally in line with similar studies in other populations and confirm that genetic factors have moderate effect in explaining BP variability in the Scottish population. The importance of other factors such as BMI was also highlighted in our study, which in combination with environmental factors that

are usually shared within the same family or household, such as socio-economic factors, can explain part of the remaining variability.

6 SNP association analysis

6.1 Introduction

GWASs have identified several genetic loci that are associated with different BP traits including SBP, DBP, MAP, and PP as quantitative traits, and hypertension as a binary trait. The ICBP is the largest ever GWAS for BP and hypertension, and reported 29 SNPs in 28 loci with a combined genetic effect explaining 0.9% of the phenotypic variance in BP (details of all the GWAS for hypertension is discussed previously in Section 1.6.2 -p77).[107] Considering the heterogeneity among different populations, it is essential to replicate the reported genetic variants in the large GWAS in independent populations. For instance, the association of these variants in specific population has been examined in two separate studies of Finnish and Chinese populations, where both studies have also assessed the association of genetic risk score with BP traits in their populations.[263,264] Comparing the results of these SNPs between populations may provide insight about the genetic contribution in BP variability between populations.

6.1.1 Aims

The primary aims of this chapter are: (1) to validate SNPs that have been previously reported to be associated with BP at a genome-wide significance level in the GS:SFHS Scottish population, using SBP and DBP as the primary phenotypes; and (2) to assess the association of a genetic risk scores that is constructed from these variants with SBP and DBP.

The secondary aims include: (1) evaluate the association of these SNPs with other BP traits including MAP and PP, and (2) to evaluate the effect of the association signals and variants effect size using different BP-lowering medication adjustment models, and (3) to compare SNPs effect sizes to those reported in other population.

6.2 Methods

6.2.1 DNA extraction, genotyping, and SNP selection

Chapter 2 explained in details the methods of DNA extraction (Section 2.4.1 -p106), the genotyping procedure (Section 2.4.2 -p20), and SNP selection

procedure (Section 2.5 -p116). In summary, DNA samples were extracted from blood samples for the vast majority of the samples, and from saliva from the remaining postal samples. The extracted DNA was transferred to TaqMan® OpenArray® plate and mixed with TaqMan® OpenArray® Master Mix, following the manufacture-recommended protocol. The genotyping procedure was performed using 64-plex OpenArray® Genotyping plates, which genotypes up to 48 samples on a single OpenArray® Genotyping plate. The plate contained 64 SNPs, of which 44 SNPs were related to this project and were selected based on two approaches. Firstly, by consulting the NHGRI catalogue of Published GWAS as of July 2012. Second, snowballing from the reference list of the identified studies and candidate-gene studies to expand the search results. Only SNPs reported in the large GWAS of European populations for association with BP traits or hypertension, and achieving a genome-significant level were selected for this study.

6.2.2 BP phenotypes

BP phenotypes and definition of hypertension were described previously in the Methods and Materials chapter (Section 2.3.1 -p98). For this chapter, the BP values used are the average of two BP readings, plus a fixed value of 15/10 mmHg for individuals taking BP-lowering medication (Section 2.3.1.1 -p98). This is in accordance with the recommendation of Tobin et al. to add a fixed value for adjustment of the treatment effect.[87] Hypertension was defined as having an average SBP of 140 mmHg or higher, an average of DBP of 90 mmHg or higher, or taking BP-lowering medications. For the overall group (n =18,470), definition of hypertension and BP adjustments were based on the information retrieved from EPRs primarily, and then from the SRMs for those individuals with missing EPRs, following the findings from the assessment of self-reported reliability in Chapter 4 (Section 4.3.2 -p136). These BP traits for the overall group were given a superscript of the letter “O”. Another two classes of adjustments were developed from information retrieved solely from EPRs and SRMs for the group of individuals with these two drug exposure sources (n =12,347). In the first class, BP traits were based on treatment-adjustments from EPRs (represented by a superscript “P”). In the second class, BP traits were based on treatment-adjustment from SRMs (represented by a superscript “Q”). The last BP traits class was based on a random treatment-adjustment using the same sample size

as group 2 (n =12,347), but a similar treatment-adjustment structure as the overall group (n =18,470). In this traits class, treatment-adjustment of 8,231 individuals was based on EPRs, and the remaining 4,116 subjects were based on SRMs (represented by a superscript "R").

6.2.3 Statistical analyses

The population sample in this study included extended-pedigree families with complex familial relationships and multi-generational pedigrees. In addition, as described in Section 3.3.5 -p122, the sample included 2,396 singleton-families in which no other relatives are presented in the study. Hence, this mixture of related and unrelated individuals in the sample requires a statistical test that account for both within- and between-family information. There are several widely-used tests for family-based studies that test for within-family information only, by assessing transmission of alleles within a family without incorporating information of allelic association observed across families, leading to loss of information. Even for studies that include family data only, simulation studies have shown that methods that assess the total association outperformed the methods that assess within-family information only.[265]

Several statistical methods were recently developed to account for family and structural population jointly using a linear mixed models (LMMs) approach, in which a genetic relationship matrix (GRM) is incorporated as a random-effect variable in the statistical model, which may also include fixed-effect variables such as age and sex. The contribution of the GRM to the total phenotypic variance of the trait is then assessed by the model, along with computing statistics that account for the remaining phenotypic variance. Basically, this approach calculates to what extent phenotypic similarity between a pair of individuals can be attributed to their similarity in the GRM, allowing an estimate of the extent to which phenotypic variance can be explained by the tested genetic variance. This approach was firstly proposed for samples of related individuals, in which the GRM can be easily constructed from the pair-wise kinship coefficients estimated from the pedigree.[125,126] This approach was then extended to use a GRM that is based on information from the genomic SNP data, by using a marker-based kinship in the GRM instead of the pedigree-based kinship matrix.[266-268] This approach of marker-based kinship is appropriate for

samples with a complex unknown family structure, in which the study has genotyped a large number of genomic markers to establish a reliable marker-based kinship. Several software packages are available for fitting LMM for GWAS, but one example is the genome-wide efficient mixed-model association (GEMMA), which provides an exact estimate of the test statistics, unlike other approximate methods.[268]

In this study, SNP association tests were performed using LMM methods to account for relatedness in GS:SFHS and dependent observations between family members. This was carried out by fitting a LMM, as implemented in GEMMA (http://www.xzlab.org/software.html, version 0.94). The included covariates in the model were sex, BMI, age, and age-squared. Adjusting for both age and age-squared allows for the non-linear effect of age on BP. The GRM in the model was provided by constructing a kinship coefficient matrix based on the pedigree information, using the "kinship 2" package in R.[269] The kinship coefficient for any pair of individuals is the probability that an allele chosen at random at a given locus is identical-by-descent. The quantitative trait association test (-robustAssoc) function that is implemented in MERLIN software was used to identify the variance explained by each tested SNP.[99,270]

6.2.3.1 Genetic risk score

The effect of GRS on the different BP traits was estimated as follows: (1) individual GRS was calculated for each individual by multiplying the effect size reported in the reference study by the number of copies of the coded allele for each individual SNP. The coded allele is the allele coded 0, 1, or 2 according to the number of copies of the allele. This is opposed to the alternative approach, in which no weighting of effect is used, and each SNP allele counts equally in the score. Five GRS were constructed for SBP, DBP, MAP, PP, and for hypertension (included SNPs in each GRS and their estimated effect size as reported by the reference study is shown in Appendix 5 -p254). (2) The association test between each GRS and BP trait was performed using a two level linear mixed model, in which fixed effects were the adjusted covariates (sex, BMI, age, and its square), and the random effect was the family unit. (3) to visualize the relationship between each GRS and its corresponding trait, GRSs were categorized into six groups based on the cutpoints (± 1&2 SD), and the

means of BP within each category were plotted. The mean difference in BP levels between the GRS groups was tested using a one-way analysis of variance (ANOVA) - linear trend analysis. Statistical analysis was performed using SPSS (version 22.0; SPSS Inc. Chicago, Illinois, USA).

6.3 Results

6.3.1 SNPs association test results for SBP

The association test revealed 9 significant SNPs associated with SBP ($p < 0.0012$), and another 14 SNPs were significant at nominal p value level ($p < 0.05$) (Table 6-1). All of the tested SNPs showed concordant direction of effect as reported by the reference studies, except for two SNPs that were not statistically significant in *MECOM* and CACNB2(5') (Figure 6-1). The coded allele effect sizes were also highly correlated with original estimated effects reported in the reference studies (r =0.86). Eight of the nine significant SNPs were originally reported in the ICBP studies, which are located in loci near to *MTHFR-NPPB*, *FGF5, NPR3-C5orf23, CACNB2(3')*, *CYP1A1-ULK3*, *FURIN-FES*, *JAG1*, and *GNAS-EDN3*. The last significant SNP is located in the *UMOD* locus, which was originally reported for association with hypertension in the BP-extreme study.

Examining the effect of the BP-treatment adjustment source by SNP association test has showed that only rs2521501 in *FURIN-FES* remained significant in all the adjustment models (even for the unadjusted model), and rs1458038 (*FGF5*) was significant with SBP^{Q} and SBP^{R}, which then became the top significant SNP in SBP^{o} (Figure 6-2). The unadjusted model (SBP^{U}) was the least correlated model with the reference studies in term of the direction and estimated effect size (r =82), compared to the other three adjusted models, which had a better correlation values (r =86; Figure 6-3)

The total phenotypic variance explained by all the genotyped SNPs was 1.4%, with the most variance explained by SNPs reported in the ICBP study (1.1%). The top significant SNP with SBP (rs1458038; *FGF5*) explained the largest extent of phenotypic variance (0.1%), with a range between 0.06% to 0.09% for the remaining significant SNPs.

Table 6-1 SNPs association results for SBP

SN	SNPs	locus	Chr	A1	A2	GS:SFHS					Reference study					Finnish cohort		
						RAF	Beta	se	P value	h^2%	Ref.	RAF	Beta	se	P value	RAF	beta	P value
1	**rs17367504**	**MTHFR-NPPB**	**1**	**G**	**A**	**0.164**	**-0.80**	**0.23**	**0.0006**	**0.06**	I	0.15	-0.90	0.09	8.72E-22	**0.85**	**-0.90**	**1.0E-05**
2	rs5068	NPPA	1	G	A	0.052	-0.96	0.39	0.0137	0.03	N	0.06	-0.80	0.02	2.00E-06	-	-	-
3	rs17030613	ST7L	1	C	A	0.2	0.30	0.22	0.1617	0.01	-	-	-	-	-	-	-	-
4	rs2932538	MOV10	1	G	A	0.756	0.29	0.20	0.1514	0.01	I	0.75	0.39	0.06	1.17E-09	**0.77**	**0.60**	**0.0005**
5	rs2004776	AGT	1	T	C	0.258	0.45	0.20	0.0223	0.03	C2	0.23	0.42	0.09	3.80E-06	0.19	0.04	0.8200
6	rs1446468	FIGN	2	A	G	0.463	-0.27	0.17	0.1236	0.02	-	-	-	-	-	-	-	-
7	rs13082711	SLC4A7	3	T	C	0.744	-0.20	0.20	0.3154	0.01	I	0.78	-0.32	0.07	1.51E-06	**0.18**	**-0.61**	**0.0010**
8	rs3774372	ULK4	3	T	C	0.851	-0.09	0.24	0.7221	0.00	I	0.83	-0.07	0.08	3.90E-01	0.21	0.02	0.9100
9	rs319690	MAP4	3	A	G	0.684	0.19	0.19	0.3150	0.01	-	-	-	-	-	-	-	-
10	rs419076	MECOM	3	T	C	0.474	-0.02	0.17	0.8956	0.00	I	0.47	0.41	0.06	1.78E-13	-	-	-
11	rs871606	CHIC2	4	T	C	0.893	0.08	0.28	0.7654	0.00	-	-	-	-	-	-	-	-
12	**rs1458038**	**FGF5**	**4**	**T**	**C**	**0.295**	**0.79**	**0.19**	**0.00005**	**0.10**	I	0.29	0.71	0.07	1.47E-23	**0.33**	**0.91**	**5.0E-09**
13	rs13107325	SLC39A8	4	T	C	0.058	-0.62	0.38	0.0967	0.01	I	0.05	-0.98	0.13	3.27E-14	**0.99**	**-2.36**	**0.0002**
14	rs13139571	GUCY1A3-GUCY1B3	4	C	A	0.751	0.37	0.20	0.0644	0.02	I	0.76	0.32	0.07	1.16E-06	**0.76**	**0.65**	**0.0001**
15	**rs1173771**	**NPR3-C5orf23**	**5**	**G**	**A**	**0.604**	**0.61**	**0.18**	**0.0007**	**0.07**	I	0.6	0.50	0.06	1.79E-16	**0.59**	**1.02**	**6.0E-12**
16	rs11953630	EBF1	5	T	C	0.365	-0.48	0.18	0.0071	0.04	I	0.37	-0.41	0.06	3.02E-11	0.66	-0.28	0.0680
17	rs1799945	HFE	6	G	C	0.158	0.28	0.24	0.2464	0.01	I	0.14	0.63	0.09	7.69E-12	0.1	0.50	0.0380
18	rs805303	BAT2-BAT5	6	G	A	0.622	0.06	0.18	0.7313	0.00	I	0.61	0.38	0.06	1.49E-11	0.52	0.29	0.0470
19	rs12705390	PIK3CG	7	A	G	0.198	0.47	0.22	0.0299	0.03	-	-	-	-	-	-	-	-
20	rs3918226	NOS3	7	T	C	0.085	0.61	0.31	0.0528	0.02	-	-	-	-	-	-	-	-
21	rs2071518	NOV	8	T	C	0.281	0.03	0.19	0.8779	0.00	-	-	-	-	-	-	-	-
22	rs4373814	CACNB2(5')	10	G	C	0.579	0.06	0.18	0.7435	0.00	I	0.55	-0.37	0.06	4.81E-11	-	-	-
23	**rs1813353**	**CACNB2(3')**	**10**	**A**	**G**	**0.661**	**0.67**	**0.18**	**0.0002**	**0.08**	I	0.68	0.57	0.08	2.56E-12	**0.69**	**0.42**	**0.0080**
24	rs4590817	C10orf107	10	G	C	0.825	0.52	0.23	0.0239	0.03	I	0.84	0.65	0.09	3.97E-12	**0.88**	**0.82**	**0.0006**

25	rs1530440	c10orf107	10	T	C	0.191	-0.69	0.22	0.0017	0.06	-	-	-	-	-	-	-	-
26	rs932764	PLCE1	10	G	A	0.431	0.36	0.18	0.0411	0.03	I	0.44	0.48	0.06	7.10E-16	**0.57**	**0.93**	**2.0E-10**
27	rs11191548	CYP17A1-NT5C2	10	T	C	0.923	0.72	0.32	0.0261	0.03	I	0.91	1.10	0.10	6.90E-26	**0.92**	**1.52**	**2.0E-08**
28	rs2782980	ADRB1	10	T	C	0.267	-0.39	0.20	0.0449	0.02	-	-	-	-	-	-	-	-
29	rs661348	LSP1/TNNT3	11	C	T	0.412	0.39	0.18	0.0268	0.03	-	-	-	-	-	-	-	-
30	rs7129220	ADM	11	G	A	0.874	-0.14	0.26	0.5798	0.00	I	0.89	-0.62	0.09	2.97E-12	**0.22**	**-1.05**	**3.0E-09**
31	rs381815	PLEKHA7	11	T	C	0.289	0.38	0.19	0.0453	0.02	I	0.26	0.57	0.09	5.27E-11	0.25	-0.22	0.1800
32	rs633185	FLJ32810-TMEM133	11	G	C	0.271	-0.60	0.19	0.0020	0.05	I	0.28	-0.56	0.07	1.21E-17	**0.71**	**-0.81**	**4.0E-06**
33	rs11222084	ADAMTS-8	11	T	A	0.37	0.08	0.18	0.6754	0.00	-	-	-	-	-	-	-	-
34	rs17249754	ATP2B1	12	G	A	0.823	0.72	0.23	0.0016	0.06	I	0.84	0.93	0.11	1.82E-18	-	-	-
35	rs3184504	SH2B3, ATXN2	12	T	C	0.502	0.28	0.17	0.1055	0.02	I	0.47	0.60	0.07	3.83E-18	-	-	-
36	rs653178	SH2B3	12	T	C	0.497	-0.31	0.17	0.0753	0.02	C1	0.53	-0.74	0.15	8.50E-07	-	-	-
37	rs10850411	TBX5-TBX3	12	T	C	0.68	0.58	0.18	0.0015	0.06	I	0.7	0.35	0.07	5.38E-08	0.67	-0.05	0.7400
38	**rs1378942**	**CYP1A1-ULK3**	**15**	**C**	**A**	**0.315**	**0.72**	**0.19**	**0.0001**	**0.08**	I	0.35	0.61	0.06	5.69E-23	-	-	-
39	**rs2521501**	**FURIN-FES**	**15**	**T**	**A**	**0.316**	**0.73**	**0.19**	**0.0001**	**0.09**	I	0.31	0.65	0.07	5.20E-19	**0.26**	**0.83**	**5.0E-07**
40	**rs13333226**	**UMOD**	**16**	**G**	**A**	**0.185**	**-0.84**	**0.22**	**0.0002**	**0.08**	S	0.19	-0.49	-	2.60E-05	0.78	-0.21	0.2200
41	rs12940887	ZNF652	17	T	C	0.361	0.23	0.18	0.2127	0.01	I	0.38	0.36	0.06	1.79E-10	0.42	0.21	0.1600
42	**rs1327235**	**JAG1**	**20**	**G**	**A**	**0.489**	**0.68**	**0.17**	**0.0001**	**0.09**	I	0.46	0.34	0.06	1.87E-08	**0.42**	**0.53**	**0.0003**
43	**rs6015450**	**GNAS-EDN3**	**20**	**G**	**A**	**0.126**	**1.02**	**0.26**	**0.0001**	**0.08**	I	0.12	0.90	0.09	3.87E-23	**0.17**	**1.41**	**4.0E-13**

SNPs in bold are statistically significant (p value <0.0012, for GS:SFHS, and p value <0.0016 for Finnish Study).
h^2 %: the extent of phenotypic variance explained by the SNP.
Ref. I: ICBP [107], S: BP-extreme[82], C1: CHARGE [115], C2: CHARGE drug target [149], N: Newton-Cheh et al.[134], Finish Cohort [263]

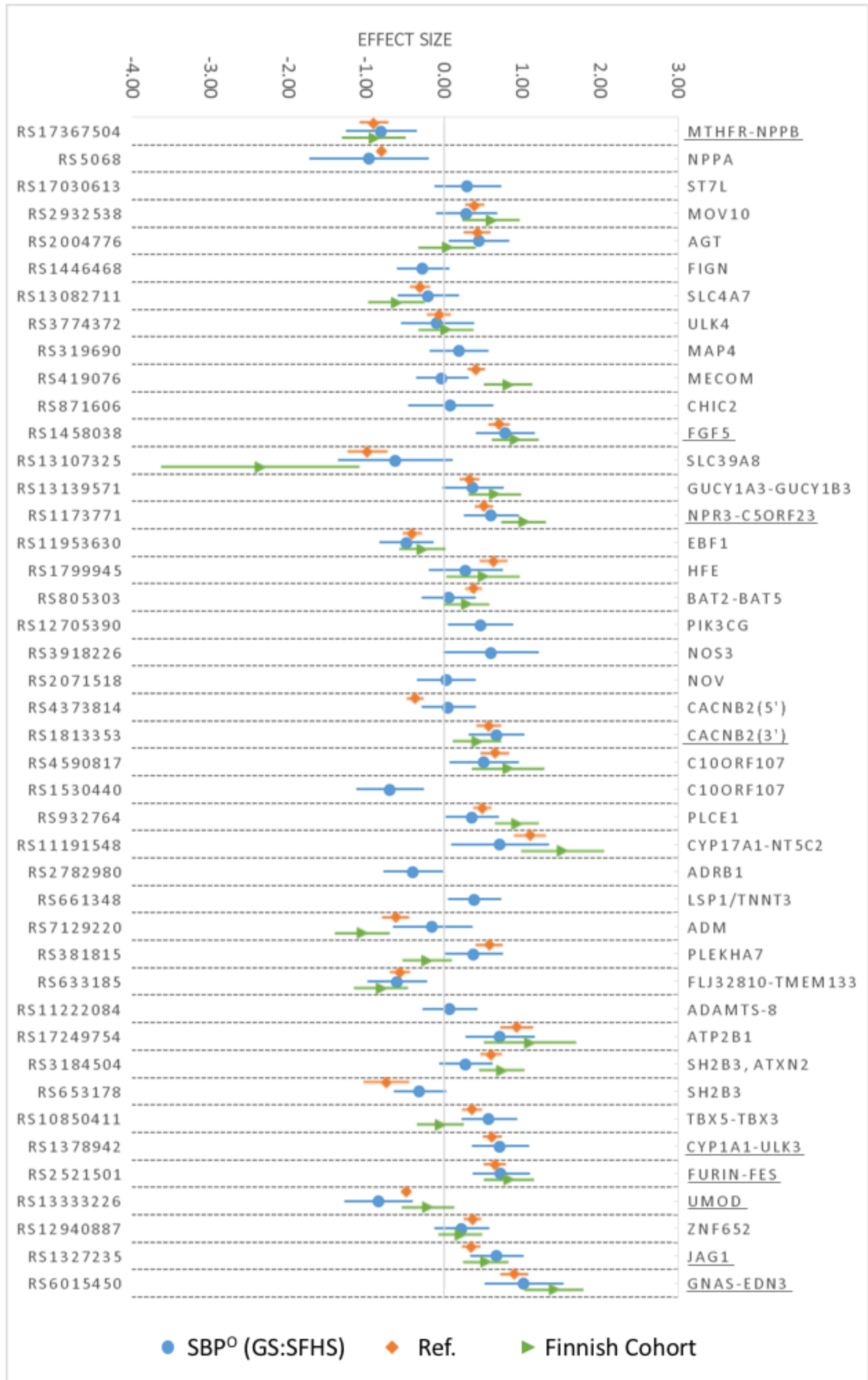


Figure 6-1 Effect size of the coded allele for SBP in GS:SFHS.
The effect sizes of coded alleles from the SBP^{O} (GS:SFHS), Reference and Finnish cohort studies are represented by different coloured markers, as indicated, with horizontal bars representing the 95% CI. SNP names are shown to the left, and loci to the right, with statistically significant loci for SBP^{O} underlined.

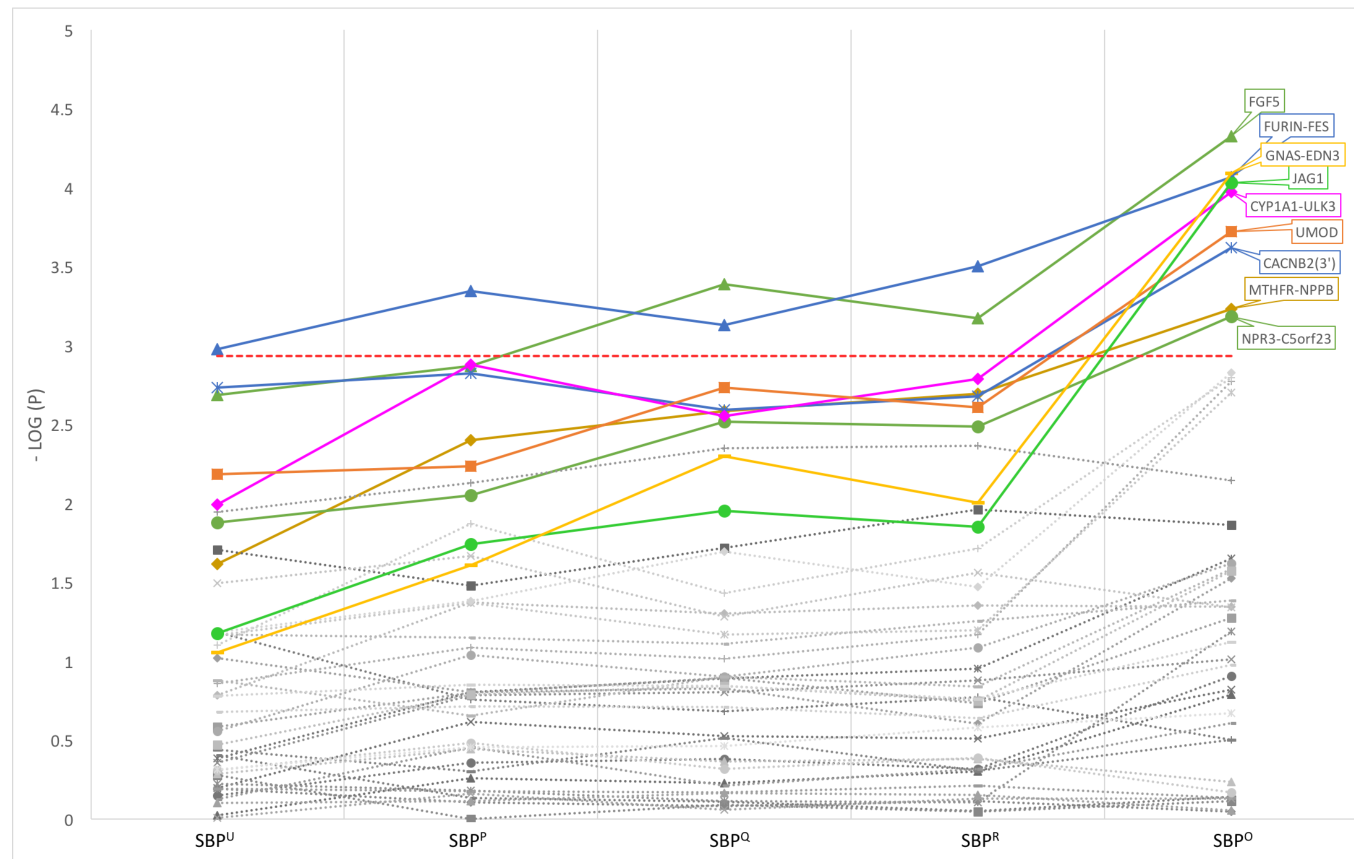


Figure 6-2 SNPs association with SBP for the different adjustment models.
SNPs that exceed the statistic threshold (p<0.0012), represented as red dotted line in the figure, for SBP are in colour and labelled with the locus name, the remaining non significant SNPs are in greyscale with dotted line. Abbreviations, SBP^U: Un-adjusted (measured) SBP, SBP^P: BP-adjustment was based on EPRs only, SBP^Q: BP-adjustment was based on SRMs only, SBP^R: BP-adjustment was based on both sources, SBP^O: BP-adjustment was based on both sources, and using the total sample size in GS:SFHS. The red dotted line shows the GS:SFHS statistical threshold.

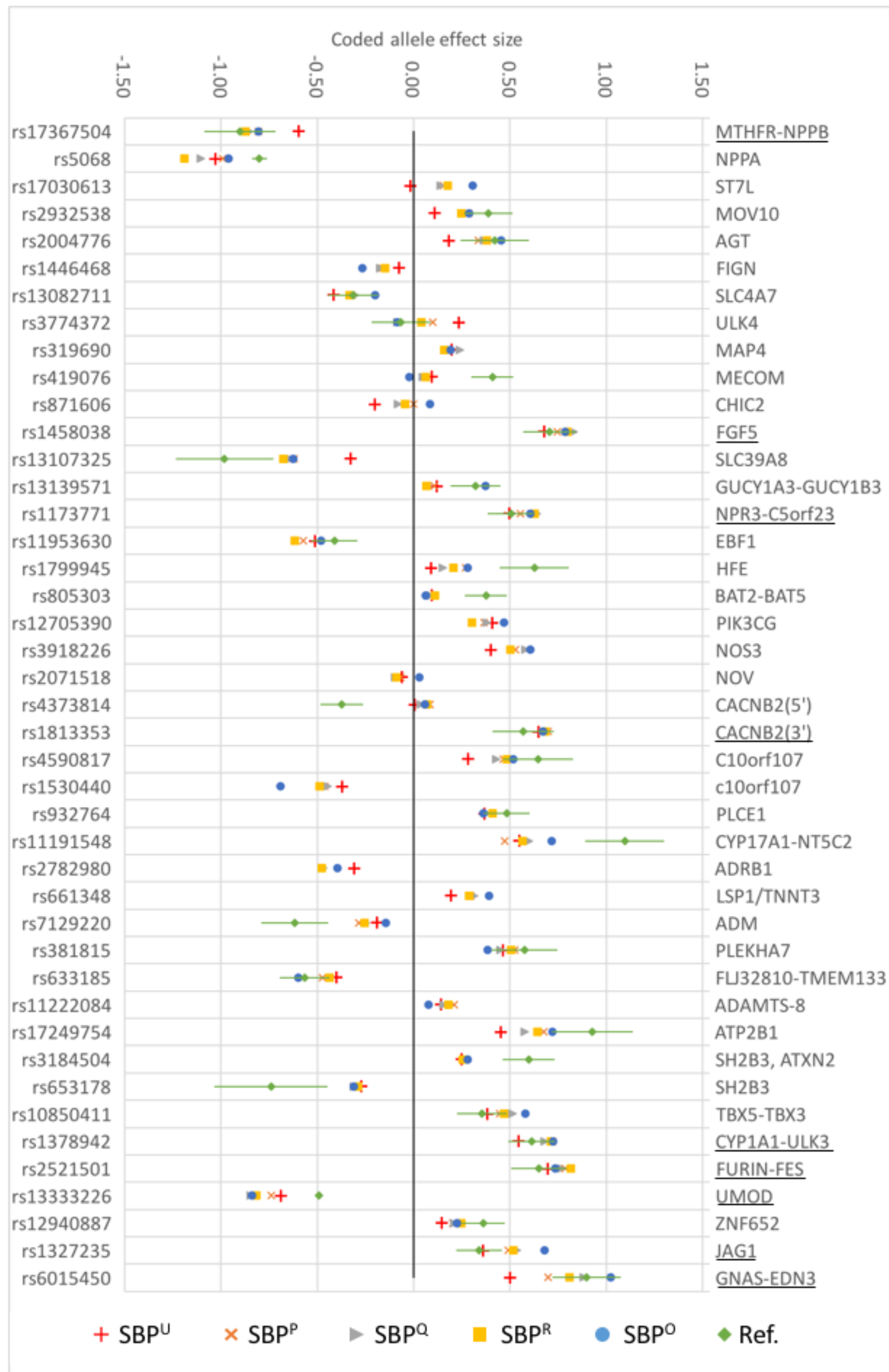


Figure 6-3 Coded allele effect size for SBP in the different adjustment models.
SNP names are shown to the left, and loci to the right, with statistically significant loci for SBP^O underlined.
Abbreviations, SBP^U: Un-adjusted (measured) SBP, SBP^P: BP-adjustment was based on EPRs only, SBP^Q: BP-adjustment was based on SRMs only, SBP^R: BP-adjustment was based on both sources, SBP^O: BP-adjustment was based on both sources, and using the total sample size in GS:SFHS.

6.3.2 SNPs association test results for DBP

In the overall model, DBP^{O} was found to be significantly associated with six loci at the Bonferroni-corrected level ($p<0.0012$), and with another 23 SNPs at the nominal p value level (Table 6-2). All the association signals were in the same direction as reported by the reference study, except for rs7129220 near to *ADM* (Figure 6-4). The correlation for the estimated effect size of the coded allele between our results and the reference study was also high (r =0.92). Of the six significant SNPs, five loci were reported originally by the ICBP study (*FGF5*, *NPR3-C5orf23*, *CYP1A1-ULK3*, *FURIN-FES*, and *GNAS-EDN3*), and the last one was originally reported by the Global BPgen study in the locus near to *c10orf107*.

Studying the association signals across the different DBP adjustment models shows that the top significant SNP (rs1378942; *CYP1A1-ULK3*) has remained significant in all the models, only rs1173771 near to *NPR3-C5orf23* has just reached the statistical significant threshold in the DBP^{Q} model only (Figure 6-5). The correlation of the coded allele effect size and direction was the lowest in the unadjusted model (r =0.89), and the remaining adjustment models had almost similar correlation values (r =0.93) (Figure 6-6).

The 43 genotyped SNPs explained 1.52% of the total phenotypic variance in DBP^{O}, with each individual significant SNP contributing between 0.06-0.14% to the phenotypic variance.

Table 6-2 SNPs association results for DBP

SN	SNP	locus	Chr	A1	A2	GS:SFHS					Reference Study					Finnish study		
						A1F	Beta	se	P value	H2	Ref	A1F	Beta	se	P value	A1F	Beta	P value
1	rs17367504	MTHFR-NPPB	1	G	A	0.164	-0.44	0.14	0.0017	0.05	I	0.15	-0.55	0.06	3.55E-19	**0.85**	**-0.46**	**0.0002**
2	rs5068	NPPA	1	G	A	0.052	-0.38	0.24	0.1080	0.01	N	0.06	-0.08	0.02	1.00E-06	-	-	-
3	rs17030613	ST7L	1	C	A	0.2	0.26	0.13	0.0505	0.02	-	-	-	-	-	-	-	-
4	rs2932538	MOV10	1	G	A	0.756	0.22	0.12	0.0735	0.02	I	0.75	0.24	0.04	9.88E-10	**0.77**	**0.32**	**0.0020**
5	rs2004776	AGT	1	T	C	0.258	0.35	0.12	0.0038	0.05	C2	0.23	0.32	0.06	5.00E-08	0.19	0.09	0.4100
6	rs1446468	FIGN	2	A	G	0.463	-0.24	0.11	0.0204	0.03	-	-	-	-	-	-	-	-
7	rs13082711	SLC4A7	3	T	C	0.744	-0.03	0.12	0.8172	0.00	I	0.78	-0.24	0.04	3.77E-09	**0.18**	**-0.40**	**0.0004**
8	rs3774372	ULK4	3	T	C	0.851	-0.35	0.15	0.0176	0.03	I	0.83	-0.37	0.05	9.02E-14	**0.21**	**-0.41**	**9.0E-0.5**
9	rs319690	MAP4	3	A	G	0.684	-0.01	0.12	0.9579	0.00	-	-	-	-	-	-	-	-
10	rs419076	MECOM	3	T	C	0.474	0.07	0.11	0.5302	0.00	I	0.47	0.24	0.03	2.12E-12	-	-	-
11	rs871606	CHIC2	4	T	C	0.893	-0.33	0.17	0.0483	0.02	-	-	-	-	-	-	-	-
12	**rs1458038**	**FGF5**	4	**T**	**C**	**0.295**	**0.38**	**0.12**	**0.0010**	**0.06**	I	0.29	0.46	0.04	8.46E-25	**0.33**	**0.63**	**6.0E-12**
13	rs13107325	SLC39A8	4	T	C	0.058	-0.57	0.23	0.0116	0.03	I	0.05	-0.68	0.08	2.28E-17	**0.99**	**-2.01**	**2.0E-07**
14	rs13139571	GUCY1A3-GUCY1B3	4	C	A	0.751	0.38	0.12	0.0018	0.05	I	0.76	0.26	0.04	2.17E-10	**0.76**	**0.43**	**2.0E-05**
15	**rs1173771**	**NPR3-C5orf23**	**5**	**G**	**A**	**0.604**	**0.37**	**0.11**	**0.0005**	**0.07**	I	0.6	0.26	0.04	9.11E-12	**0.59**	**0.37**	**3.0E-05**
16	rs11953630	EBF1	5	T	C	0.365	-0.23	0.11	0.0326	0.03	I	0.37	-0.28	0.04	3.81E-13	0.66	-0.16	0.0720
17	rs1799945	HFE	6	G	C	0.158	0.30	0.15	0.0432	0.02	I	0.14	0.46	0.06	1.45E-15	0.1	0.42	0.0040
18	rs805303	BAT2-BAT5	6	G	A	0.622	0.09	0.11	0.3821	0.01	I	0.61	0.23	0.03	2.98E-11	0.52	0.07	0.4400
19	rs12705390	PIK3CG	7	A	G	0.198	-0.05	0.13	0.6880	0.00	-	-	-	-	-	-	-	-
20	rs3918226	NOS3	7	T	C	0.085	0.60	0.19	0.0016	0.05	C1	0.08	0.78	0.18	2.20E-09	-	-	-
21	rs2071518	NOV	8	T	C	0.281	-0.25	0.12	0.0316	0.02	-	-	-	-	-	-	-	-
22	rs4373814	CACNB2(5')	10	G	C	0.579	-0.17	0.11	0.1106	0.02	I	0.55	-0.22	0.04	4.36E-10	-	-	-
23	rs1813353	CACNB2(3')	10	A	G	0.661	0.33	0.11	0.0028	0.05	I	0.68	0.41	0.05	2.30E-15	**0.69**	**0.36**	**0.0001**
24	rs4590817	C10orf107	10	G	C	0.825	0.37	0.14	0.0084	0.04	I	0.84	0.42	0.06	1.29E-12	0.88	0.42	0.0030
25	**rs1530440**	**c10orf107**	**10**	**T**	**C**	**0.191**	**-0.48**	**0.13**	**0.0003**	**0.08**	G	0.19	-0.39	0.06	1.00E-09	-	-	-
26	rs932764	PLCE1	10	G	A	0.431	0.10	0.11	0.3531	0.01	I	0.44	0.18	0.04	8.06E-07	**0.57**	**0.24**	**0.0070**

27	rs11191548	CYP17A1-NT5C2	10	T	C	0.923	0.44	0.20	0.0254	0.03	I	0.91	0.46	0.07	9.44E-13	**0.92**	**0.52**	**0.0010**
28	rs2782980	ADRB1	10	T	C	0.267	-0.31	0.12	0.0096	0.04	-	-	-	-	-	-	-	-
29	rs661348	LSP1/TNNT3	11	C	T	0.412	0.12	0.11	0.2769	0.01	-	-	-	-	-	-	-	-
30	rs7129220	ADM	11	G	A	0.874	0.12	0.16	0.4520	0.00	I	0.89	-0.30	0.06	6.44E-08	**0.22**	**0.36**	**0.0006**
31	rs381815	PLEKHA7	11	T	C	0.289	0.04	0.12	0.7054	0.00	I	0.26	0.35	0.06	5.34E-10	0.25	-0.01	0.9300
32	rs633185	FLJ32810-TMEM133	11	G	C	0.271	-0.35	0.12	0.0027	0.05	I	0.28	-0.33	0.04	1.95E-15	**0.71**	**-0.46**	**1.0E-05**
33	rs11222084	ADAMTS-8	11	T	A	0.37	-0.29	0.11	0.0087	0.04	-	-	-	-	-	-	-	-
34	rs17249754	ATP2B1	12	G	A	0.823	0.30	0.14	0.0325	0.03	I	0.84	0.52	0.07	1.16E-14	-	-	-
35	rs3184504	SH2B3, ATXN2	12	T	C	0.502	0.28	0.11	0.0076	0.04	I	0.47	0.45	0.04	3.59E-25	-	-	-
36	rs653178	SH2B3	12	T	C	0.497	-0.30	0.11	0.0041	0.05	G	0.52	-0.46	0.05	3.00E-18	-	-	-
37	rs10850411	TBX5-TBX3	12	T	C	0.68	0.27	0.11	0.0156	0.04	I	0.7	0.25	0.04	5.43E-10	0.67	-0.02	0.8500
38	**rs1378942**	**CYP1A1-ULK3**	**15**	**C**	**A**	**0.315**	**0.58**	**0.11**	**2.32E-07**	**0.14**	I	0.35	0.42	0.04	2.69E-26	-	-	-
39	**rs2521501**	**FURIN-FES**	**15**	**T**	**A**	**0.316**	**0.39**	**0.11**	**0.0006**	**0.07**	I	0.31	0.36	0.05	1.89E-15	0.26	0.30	0.0020
40	rs13333226	UMOD	16	G	A	0.185	-0.41	0.14	0.0029	0.05		0.19	-0.23	-	1.50E-05	0.78	-0.18	0.0790
41	rs12940887	ZNF652	17	T	C	0.361	0.17	0.11	0.1121	0.02	I	0.38	0.27	0.04	2.29E-14	**0.42**	**0.30**	**0.0006**
42	rs1327235	JAG1	20	G	A	0.489	0.33	0.11	0.0019	0.06	I	0.46	0.30	0.04	1.41E-15	**0.42**	**0.38**	**1.0E-05**
43	**rs6015450**	**GNAS-EDN3**	**20**	**G**	**A**	**0.126**	**0.61**	**0.16**	**0.0001**	**0.08**	I	0.12	0.56	0.06	5.63E-23	**0.17**	**0.80**	**5.0E-12**

SNPs in bold are statistically significant (p value <0.0012, for GS:SFHS, and p value <0.0016 for Finnish Study).

h^2% shows the extent of phenotypic variance explained by the SNP.

Ref. I: ICBP [107], S: BP-extreme[82], C1: CHARGE [115], C2: CHARGE drug target [149], N: Newton-Cheh et al.[134], Finish Cohort [263], G: Global BPgen [133]

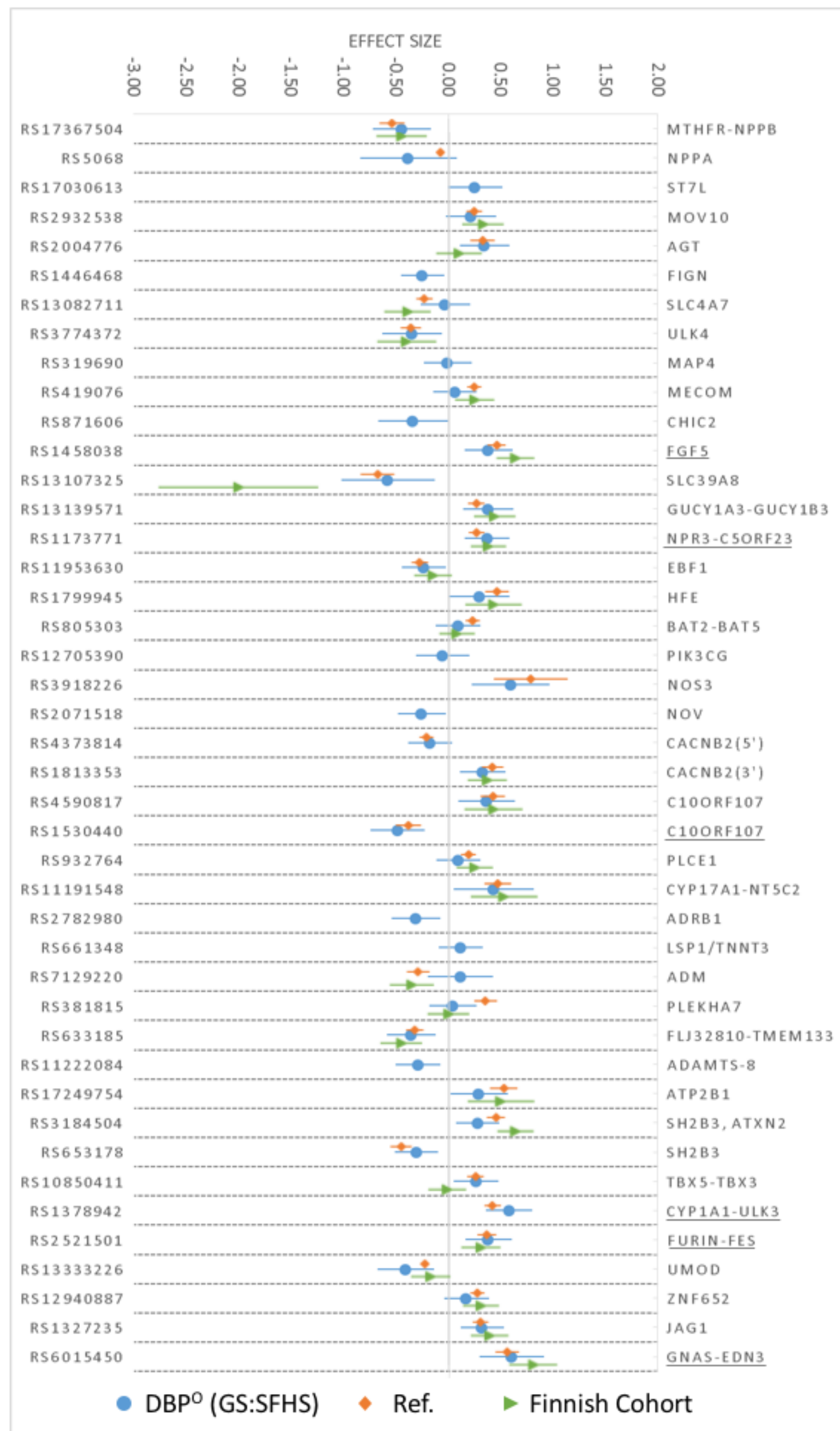


Figure 6-4 Effect size of the coded allele for DBP in GS:SFHS
The effect sizes of coded alleles from the DBP^{O}(GS:SFHS), Reference and Finnish cohort studies are represented by different coloured markers, as indicated, with horizontal bars representing the 95% CI confidence intervals. SNP names are shown in the left, and loci in the right, with statistically significant loci for DBP^{O} underlined.

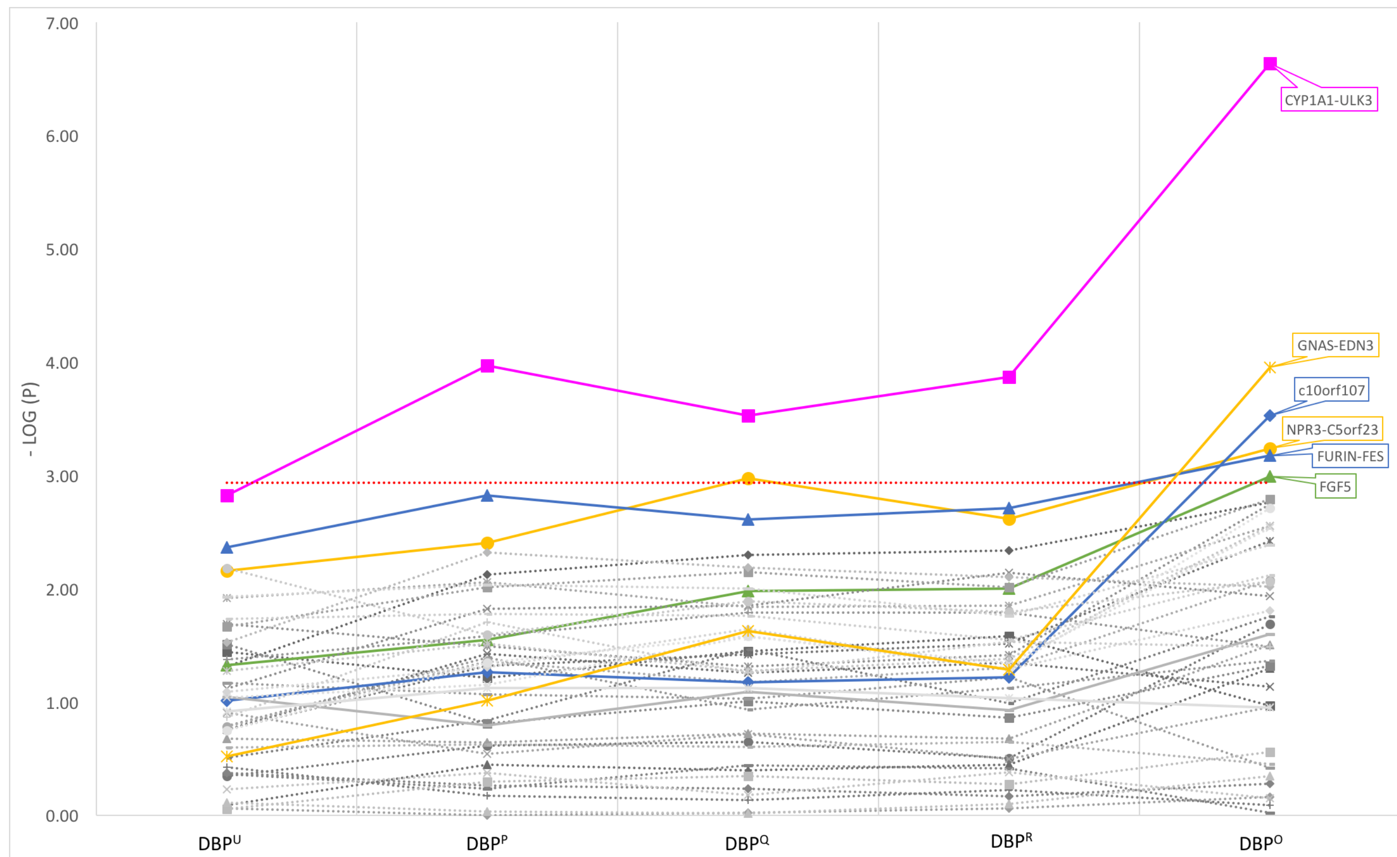


Figure 6-5 SNPs association with DBP for the different adjustment models.
SNPs that exceed the GS:SFHS statistic threshold ($p<0.0012$), represented as red dotted line in the figure, for DBP are in colour and labelled with the locus name, the remaining non significant SNPs are in greyscale. Abbreviations, DBP^U: Unadjusted (i.e. measured) DBP, DBP^P: BP-adjustment was based on EPRs only, DBP^Q: BP-adjustment was based on SRMs only, DBP^R: BP-adjustment was based on both sources, DBP^O: BP-adjustment was based on both sources, and using the total sample size in GS:SFHS.

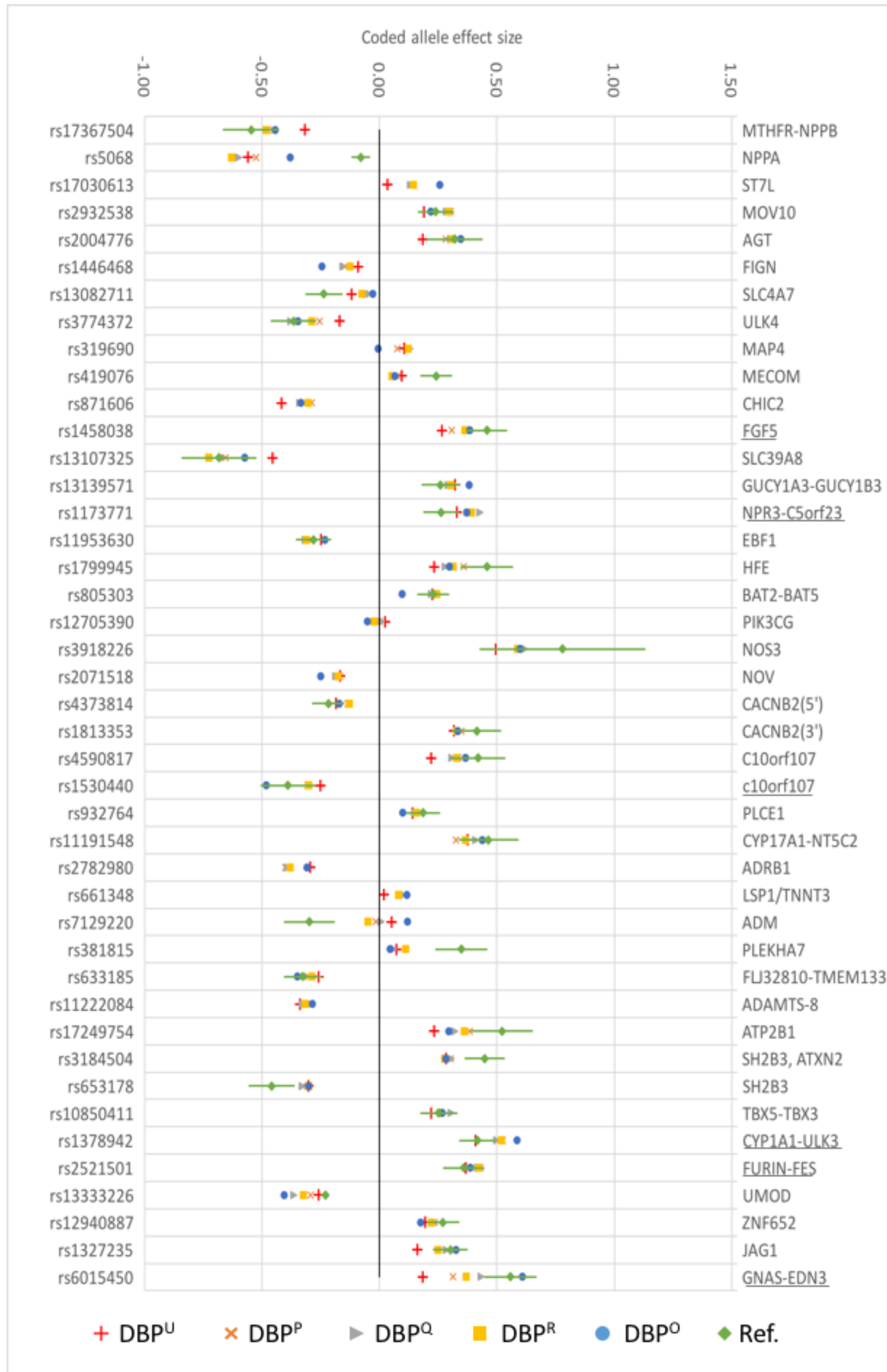


Figure 6-6 Coded allele effect size for DBP in the different adjustment models.
SNP names are shown to the left, and loci to the right, with statistically significant loci for DBP^{O} underlined.
Abbreviations, DBP^{U}: Unadjusted (i.e. measured) DBP, DBP^{P}: BP-adjustment was based on EPRs only, DBP^{Q}: BP-adjustment was based on SRMs only, DBP^{R}: BP-adjustment was based on both sources, DBP^{O}: BP-adjustment was based on both sources, and using the total sample size in GS:SFHS.

6.3.3 SNPs association test results for MAP

Testing the association of the reported GWAS SNPs for BP with MAP^{O} has revealed nine significant loci at the Bonferroni corrected significance level (P<0.0012); all of them except rs13333226 (*UMOD*) were also reported by Wain et al. for association with MAP, plus an association with another 15 SNPs at the nominal p value (Table 6-3).These SNPs were located in loci near to *MTHFR-NPPB*, *FGF5*, *NPR3-C5orf23*, *CYP1A1-ULK3*, *FURIN-FES*, and *GNAS-EDN3*. The remaining three loci are near to *CACNB2(3'), c10orf107,* and *UMOD; which* were originally reported for other BP and hypertension traits. All the tested SNPs were in the same direction of effect as reported by the reference study (Figure 6-7). The correlation of the coded allele effect size between our estimates and the reported effect size in the reference study was very high (*r* =0.93).

The number of significant SNPs remained the same across the different adjustment models, as the top two SNPs in the full sample size model (MAP^{O}) near to *CYP1A1-ULK3* and *FURIN-FES* were also significant in the three smaller sample size models (Figure 6-8). The correlation coefficients of the estimated effect size from the different adjustment models with the reference study was similar (*r* =0.93) (Figure 6-9).

The total of phenotypic variance explained by all the genotyped SNPs was 1.6%, and the phenotypic variance explained by the significant SNPs only was 0.5%. Individually, rs1378942 near to *CYP1A1-ULK3* was the SNP that explained the largest phenotypic variance (0.13%), with other significant SNPs explaining 0.07-0.10% each.

Table 6-3 SNPs association results for MAP

SN	SNP	locus	Chr	A1	A2	GS:SFHS					Reference study			
						A1F	Beta	se	*P* value	H2	A1F	Beta	se	*P* value
1	**rs17367504**	**MTHFR-NPPB**	1	**G**	**A**	**0.16**	**-0.50**	**0.15**	**0.0010**	**0.07**	0.17	-0.53	0.07	2.00E-16
2	rs5068	NPPA	1	G	A	0.05	-0.58	0.26	0.0221	0.03	-	-	-	-
3	rs17030613	ST7L	1	C	A	0.20	0.24	0.14	0.0971	0.02	-	-	-	-
4	rs2932538	MOV10	1	G	A	0.76	0.19	0.13	0.1510	0.02	0.74	0.25	0.06	8.00E-06
5	rs2004776	AGT	1	T	C	0.26	0.31	0.13	0.0156	0.05	-	-	-	-
6	rs1446468	FIGN	2	A	G	0.46	-0.20	0.11	0.0792	0.03	0.53	-0.34	0.05	6.00E-12
7	rs13082711	SLC4A7	3	T	C	0.74	-0.09	0.13	0.4680	0.00	0.80	-0.34	0.06	5.00E-09
8	rs3774372	ULK4	3	T	C	0.85	-0.19	0.16	0.2300	0.01	-	-	-	-
9	rs319690	MAP4	3	A	G	0.68	0.05	0.13	0.6720	0.00	0.51	0.30	0.05	3.00E-08
10	rs419076	MECOM	3	T	C	0.47	0.05	0.11	0.6380	0.00	0.44	0.34	0.05	8.00E-13
11	rs871606	CHIC2	4	T	C	0.89	-0.24	0.18	0.1820	0.01	-	-	-	-
12	**rs1458038**	**FGF5**	**4**	**T**	**C**	**0.30**	**0.46**	**0.13**	**0.0003**	**0.09**	0.30	0.40	0.05	3.00E-14
13	rs13107325	SLC39A8	4	T	C	0.06	-0.52	0.25	0.0335	0.02	0.12	-0.63	0.10	1.00E-10
14	rs13139571	GUCY1A3-GUCY1B3	4	C	A	0.75	0.38	0.13	0.0036	0.04	0.74	0.29	0.06	3.00E-07
15	**rs1173771**	**NPR3-C5orf23**	**5**	**G**	**A**	**0.60**	**0.40**	**0.12**	**0.0006**	**0.09**	0.53	0.28	0.05	4.00E-09
16	rs11953630	EBF1	5	T	C	0.37	-0.27	0.12	0.0206	0.04	-	-	-	-
17	rs1799945	HFE	6	G	C	0.16	0.26	0.16	0.1030	0.02	-	-	-	-
18	rs805303	BAT2-BAT5	6	G	A	0.62	0.09	0.12	0.4210	0.00	-	-	-	-
19	rs12705390	PIK3CG	7	A	G	0.20	0.14	0.14	0.3280	0.00	-	-	-	-
20	rs3918226	NOS3	7	T	C	0.09	0.51	0.20	0.0123	0.04	-	-	-	-
21	rs2071518	NOV	8	T	C	0.28	-0.15	0.13	0.2230	0.01	-	-	-	-
22	rs4373814	CACNB2(5')	10	G	C	0.58	-0.11	0.12	0.3480	0.00	-	-	-	-
23	**rs1813353**	**CACNB2(3')**	**10**	**A**	**G**	**0.66**	**0.44**	**0.12**	**0.0003**	**0.07**	-	-	-	-
24	rs4590817	C10orf107	10	G	C	0.83	0.37	0.15	0.0131	0.04	0.83	0.58	0.07	2.00E-18

25	**rs1530440**	**c10orf107**	**10**	**T**	**C**	**0.19**	**-0.53**	**0.14**	**0.0003**	**0.08**	-	-	-	-
26	rs932764	PLCE1	10	G	A	0.43	0.18	0.12	0.1150	0.01	-	-	-	-
27	rs11191548	CYP17A1-NT5C2	10	T	C	0.92	0.51	0.21	0.0165	0.03	-	-	-	-
28	rs2782980	ADRB1	10	T	C	0.25	-0.30	0.13	0.0207	0.03	0.20	-0.34	0.06	2.00E-09
29	rs661348	LSP1/TNNT3	11	C	T	0.41	0.18	0.12	0.1240	0.02	-	-	-	-
30	rs7129220	ADM	11	G	A	0.87	0.03	0.17	0.8480	0.00	-	-	-	-
31	rs381815	PLEKHA7	11	T	C	0.30	0.15	0.13	0.2370	0.01	0.30	0.30	0.05	3.00E-08
32	rs633185	FLJ32810-TMEM133	11	G	C	0.27	-0.41	0.13	0.0014	0.06	0.32	-0.33	0.05	7.00E-10
33	rs11222084	ADAMTS-8	11	T	A	0.37	-0.17	0.12	0.1490	0.01	-	-	-	-
34	rs17249754	ATP2B1	12	G	A	0.82	0.42	0.15	0.0051	0.04	0.89	0.56	0.07	1.00E-17
35	rs3184504	SH2B3, ATXN2	12	T	C	0.50	0.27	0.11	0.0194	0.03	-	-	-	-
36	rs653178	SH2B3	12	T	C	0.50	-0.29	0.11	0.0115	0.04	0.59	-0.43	0.00	7.00E-20
37	rs10850411	TBX5-TBX3	12	T	C	0.68	0.33	0.12	0.0061	0.05	-	-	-	-
38	**rs1378942**	**CYP1A1-ULK3**	**15**	**C**	**A**	**0.32**	**0.59**	**0.12**	**1.16E-06**	**0.13**	0.33	0.39	0.05	2.00E-15
39	**rs2521501**	**FURIN-FES**	**15**	**T**	**A**	**0.32**	**0.50**	**0.12**	**4.66E-05**	**0.09**	0.37	0.34	0.06	3.00E-08
40	**rs13333226**	**UMOD**	**16**	**G**	**A**	**0.19**	**-0.49**	**0.15**	**0.0008**	**0.07**	-	-	-	-
41	rs12940887	ZNF652	17	T	C	0.36	0.17	0.12	0.1470	0.02	0.42	0.25	0.05	2.00E-07
42	rs1327235	JAG1	20	G	A	0.49	0.36	0.11	0.0014	0.08	0.58	0.26	0.05	4.00E-08
43	**rs6015450**	**GNAS-EDN3**	**20**	**G**	**A**	**0.13**	**0.63**	**0.17**	**0.0002**	**0.10**	0.07	0.52	0.07	2.00E-12

SNPs in bold are statistically significant (p value <0.0012).
h^2% shows the extent of phenotypic variance explained by the SNP.
Reference study for MAP is Wain et al.[135]

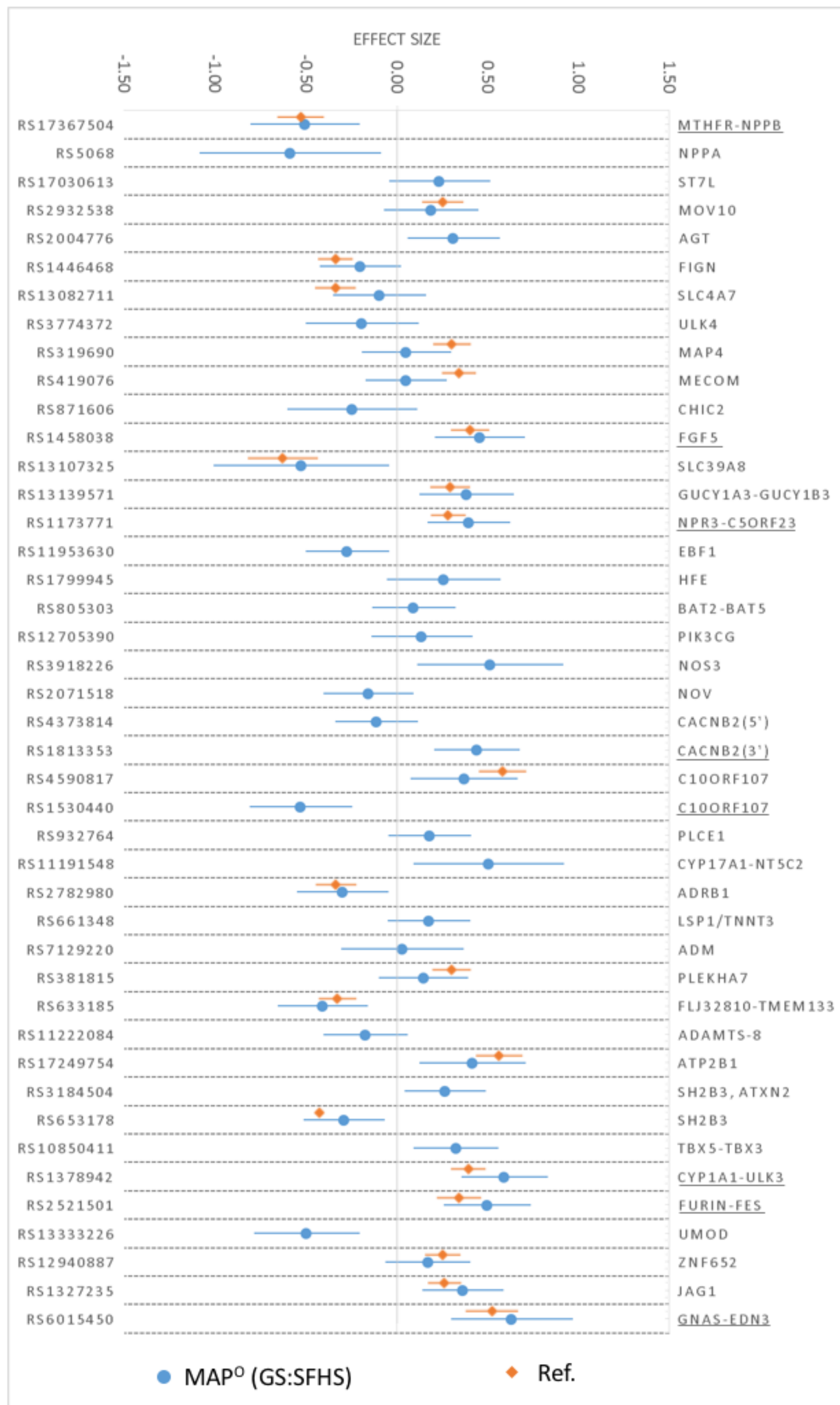


Figure 6-7 Effect size of the coded allele for MAP in GS:SFHS
Blue circles show the estimated effect size in GS:SFHS, and the estimated effect size of the reference study for the reported SNPs for association with MAP are shown as orange diamonds. Bars around the symbols show the 95% CI. SNP names are shown in the left, and loci in the right, with statistically significant loci for MAP^{O} underlined

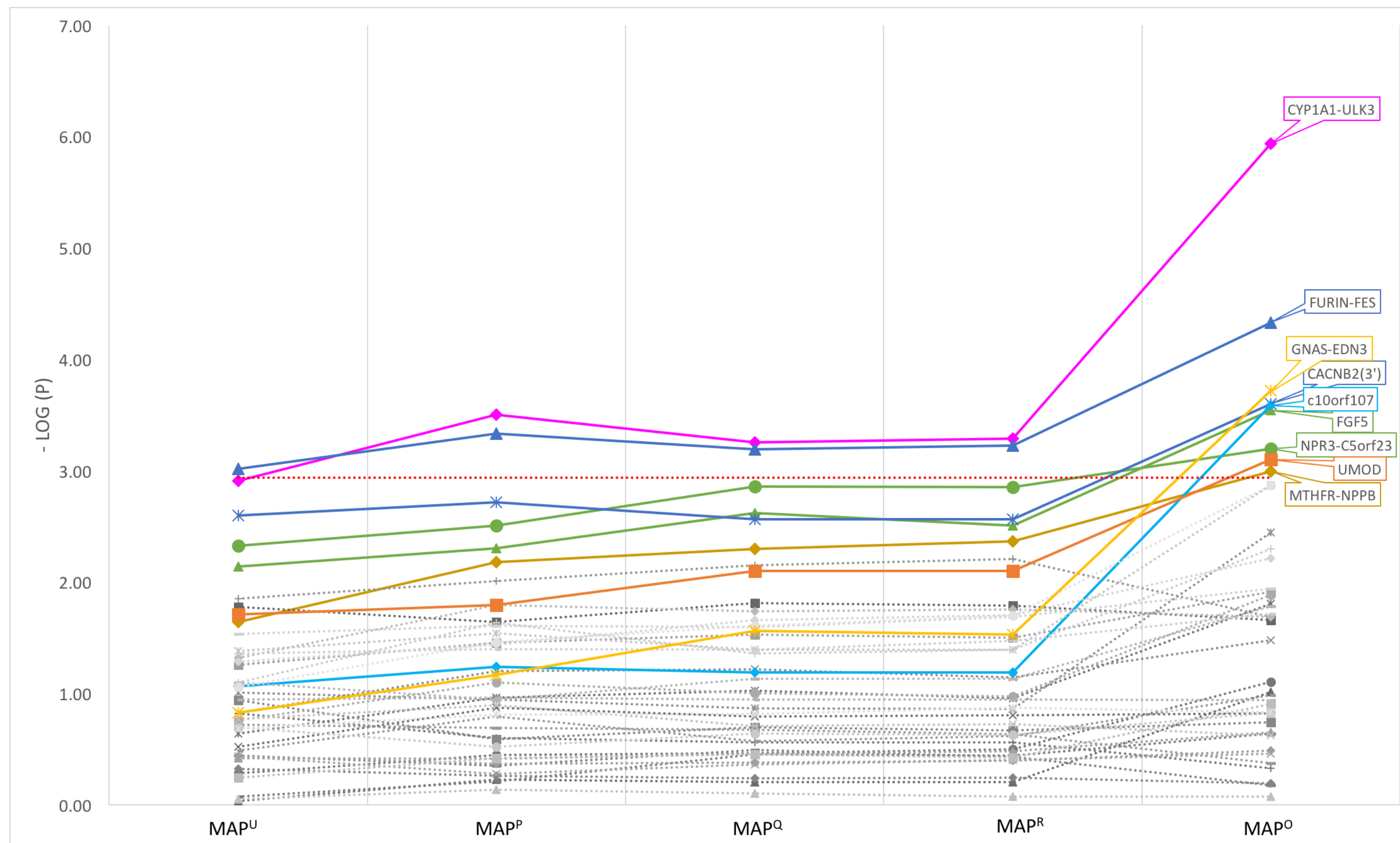


Figure 6-8 SNPs association results with MAP for the different adjustment models
SNPs that exceed the GS:SFHS statistic threshold ($p<0.0012$), represented as red dotted line in the figure, for MAP are in colour and labelled with the locus name, the remaining non significant SNPs are in greyscale with dotted lines. Abbreviations, Unadjusted (i.e. measured) MAP, MAP ^{P}: BP-adjustment was based on EPRs only, MAP ^{Q}: BP-adjustment was based on SRMs only, MAP ^{R}: BP-adjustment was based on both sources, MAP ^{O}: BP-adjustment was based on both sources, and using the total sample size in GS:SFHS.

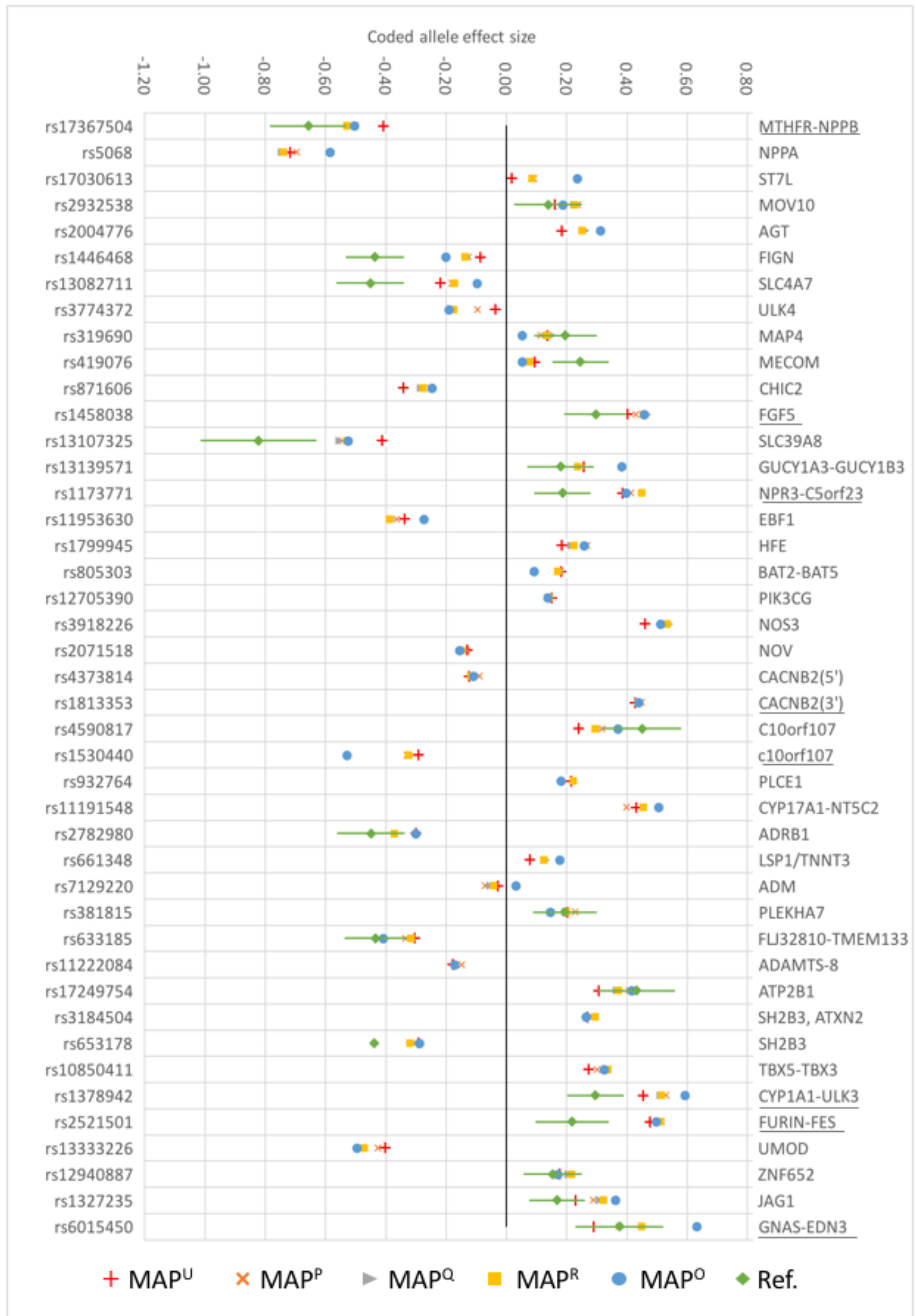


Figure 6-9 Coded allele effect size for MAP in the different adjustment models.
SNP names are shown to the left, and loci to the right, with statistically significant loci for MAP^O underlined.
Abbreviations, MAP^U: Unadjusted (i.e. measured) MAP, MAP^P: BP-adjustment was based on EPRs only, MAP^Q: BP-adjustment was based on SRMs only, MAP^R: BP-adjustment was based on both sources, MAP^O: BP-adjustment was based on both sources, and using the total sample size in GS:SFHS.

6.3.4 SNPs association test for PP

Only rs12705390 near to *PIK3CG* was significant with PP^O at the Bonferroni corrected significance level ($P<0.0012$); this SNP is in complete LD with rs17477177 that was reported by the reference study for association with PP (Table 6-4).The direction of effect for the coded allele was in the same direction as reported in the reference study, with a high correlation coefficient (r =0.9) for the size of the effect (Figure 6-10). In the other adjustment models, rs11222084 near to *ADAMTS-8* was statistically significant with PP^P but the association vanished after increasing the sample size in the overall model PP^O, as well as with the other two adjustment models (i.e. PP^R and PP^Q) (Figure 6-11). The estimated effect size across the different adjustment models with the reference study were very similar ($r \approx 0.73$), except for the overall model which was highly correlated with the reference study (r =0.9) (Figure 6-12) The genotyped SNPs explained only 0.8% of the total phenotypic variance in PP^O, with the only significant SNP near to *PIK3CG* (rs12705390) explained the highest proportion of the variance individually (0.07%).

Table 6-4 SNPs association results for PP

SN	SNP	locus	Chr	A1	A2	GS:SFHS					Reference study			
						A1F	Beta	se	*P* value	H2	A1F	Beta	se	*P* value
1	rs17367504	MTHFR-NPPB	1	G	A	0.16	-0.35	0.16	0.03	0.02	-	-	-	-
2	rs5068	NPPA	1	G	A	0.05	-0.57	0.27	0.03	0.02	-	-	-	-
3	rs17030613	ST7L	1	C	A	0.20	0.04	0.15	0.79	0.00	-	-	-	-
4	rs2932538	MOV10	1	G	A	0.76	0.07	0.14	0.59	0.00	-	-	-	-
5	rs2004776	AGT	1	T	C	0.26	0.10	0.14	0.48	0.00	-	-	-	-
6	rs1446468	FIGN	2	A	G	0.46	-0.02	0.12	0.88	0.00	-	-	-	-
7	rs13082711	SLC4A7	3	T	C	0.74	-0.18	0.14	0.20	0.01	-	-	-	-
8	rs3774372	ULK4	3	T	C	0.85	0.26	0.17	0.13	0.01	-	-	-	-
9	rs319690	MAP4	3	A	G	0.68	0.20	0.13	0.13	0.01	-	-	-	-
10	rs419076	MECOM	3	T	C	0.47	-0.08	0.12	0.49	0.00	-	-	-	-
11	rs871606	CHIC2	4	T	C	0.89	0.40	0.19	0.04	0.02	0.85	0.43	0.08	1.00E-08
12	rs1458038	FGF5	4	T	C	0.30	0.39	0.13	0.0034	0.05	-	-	-	-
13	rs13107325	SLC39A8	4	T	C	0.06	-0.04	0.26	0.88	0.00	-	-	-	-
14	rs13139571	GUCY1A3-GUCY1B3	4	C	A	0.75	-0.01	0.14	0.95	0.00	-	-	-	-
15	rs1173771	NPR3-C5orf23	5	G	A	0.60	0.23	0.12	0.07	0.02	0.53	0.28	0.05	5.00E-09
16	rs11953630	EBF1	5	T	C	0.37	-0.26	0.12	0.04	0.02	-	-	-	-
17	rs1799945	HFE	6	G	C	0.16	0.00	0.17	0.98	0.00	-	-	-	-
18	rs805303	BAT2-BAT5	6	G	A	0.62	-0.03	0.12	0.80	0.00	-	-	-	-
19	**rs12705390***	**PIK3CG**	7	**A**	**G**	**0.20**	**0.52**	**0.15**	**0.0005**	**0.07**	0.17	0.42	0.06	2.27E-13
20	rs3918226	NOS3	7	T	C	0.09	0.03	0.22	0.91	0.00	-	-	-	-
21	rs2071518	NOV	8	T	C	0.28	0.28	0.13	0.03	0.03	0.18	0.31	0.05	4.00E-09
22	rs4373814	CACNB2(5')	10	G	C	0.58	0.23	0.12	0.06	0.02	-	-	-	-
23	rs1813353	CACNB2(3')	10	A	G	0.66	0.34	0.13	0.01	0.04	-	-	-	-
24	rs4590817	C10orf107	10	G	C	0.83	0.15	0.16	0.36	0.01	-	-	-	-

25	rs1530440	c10orf107	10	T	C	0.19	-0.20	0.15	0.18	0.01	-	-	-	-
26	rs932764	PLCE1	10	G	A	0.43	0.26	0.12	0.03	0.03	-	-	-	-
27	rs11191548	CYP17A1-NT5C2	10	T	C	0.92	0.28	0.22	0.22	0.01	0.94	0.53	0.08	8.00E-11
28	rs2782980	ADRB1	10	T	C	0.27	-0.08	0.14	0.55	0.00	-	-	-	-
29	rs661348	LSP1/TNNT3	11	C	T	0.41	0.27	0.12	0.03	0.03	-	-	-	-
30	rs7129220	ADM	11	G	A	0.87	-0.26	0.18	0.14	0.01	0.87	-0.38	0.08	4.00E-07
31	rs381815	PLEKHA7	11	T	C	0.29	0.34	0.13	0.01	0.04	0.3	0.24	0.05	7.00E-06
32	rs633185	FLJ32810-TMEM133	11	G	C	0.27	-0.25	0.13	0.07	0.02	-	-	-	-
33	rs11222084	ADAMTS-8	11	T	A	0.37	0.36	0.12	0.0044	0.05	0.37	0.34	0.05	2.00E-11
34	rs17249754	ATP2B1	12	G	A	0.82	0.42	0.16	0.01	0.04	0.89	0.39	0.06	6.00E-10
35	rs3184504	SH2B3, ATXN2	12	T	C	0.50	0.00	0.12	0.98	0.00	-	-	-	-
36	rs653178	SH2B3	12	T	C	0.50	-0.01	0.12	0.94	0.00	-	-	-	-
37	rs10850411	TBX5-TBX3	12	T	C	0.68	0.31	0.13	0.01	0.03	-	-	-	-
38	rs1378942	CYP1A1-ULK3	15	C	A	0.32	0.13	0.13	0.30	0.01	-	-	-	-
39	rs2521501	FURIN-FES	15	T	A	0.32	0.34	0.13	0.01	0.04	-	-	-	-
40	rs13333226	UMOD	16	G	A	0.19	-0.43	0.16	0.01	0.04	-	-	-	-
41	rs12940887	ZNF652	17	T	C	0.36	0.05	0.13	0.67	0.00	-	-	-	-
42	rs1327235	JAG1	20	G	A	0.49	0.36	0.12	0.0028	0.05	-	-	-	-
43	rs6015450	GNAS-EDN3	20	G	A	0.13	0.40	0.18	0.03	0.03	0.07	0.35	0.07	2.00E-06

* SNP in bold is statistically significant (p value <0.0012), which is in a complete LD with SNP rs17477177 reported by Wain LV et al. for association with PP.
h^2% shows the extent of phenotypic variance explained by the SNP.
Reference study is [135]

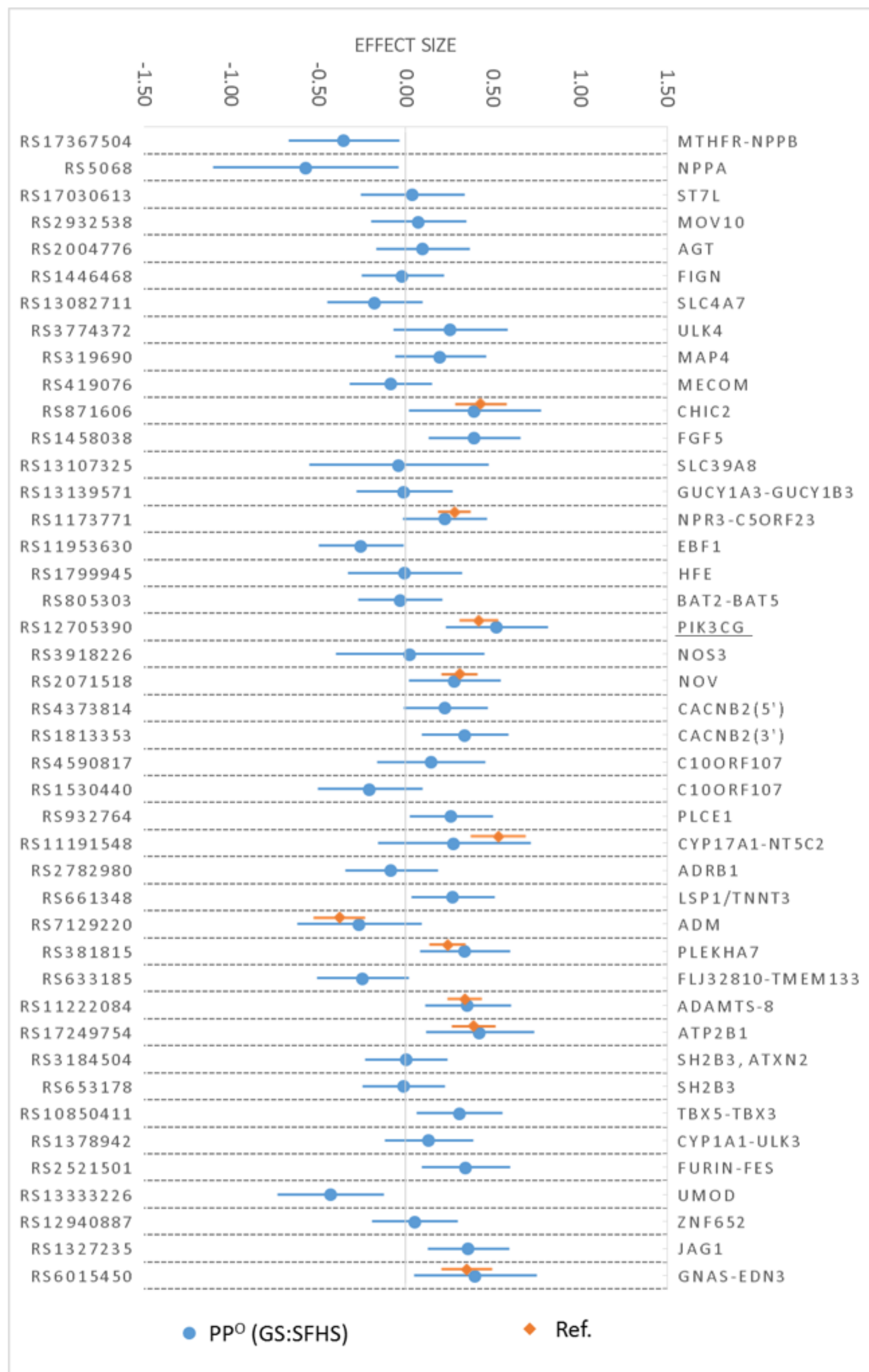


Figure 6-10 Effect size of the coded allele for PP in GS:SFHS
Blue circles show the estimated effect size in GS:SFHS, and the estimated effect size of the reference study for the reported SNPs for association with PP are shown as orange diamonds. Bars around the symbols show the 95% CI. SNP names are shown in the left, and loci in the right, with statistically significant loci for PP^{O} underlined.

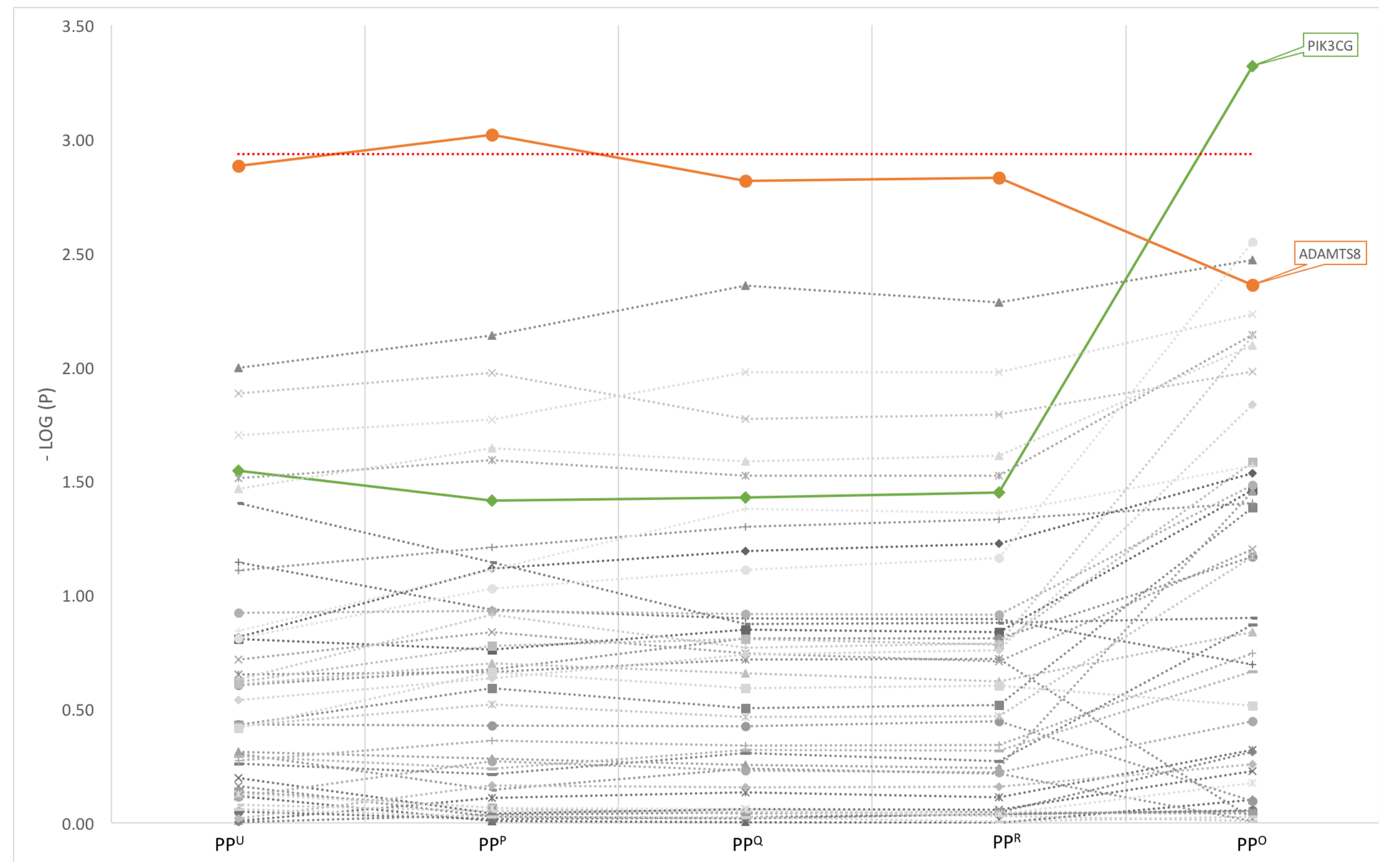


Figure 6-11 SNPs association results with PP for the different adjustment models.
SNPs that exceeds the GS:SFHS statistic threshold ($p<0.0012$), represented as red dotted line in the figure, for PP are in colour and labelled with the locus name, the remaining non significant SNPs are in greyscale with dotted lines. Abbreviations, PP^U: Unadjusted (i.e. measured) PP, PP^P: BP-adjustment was based on EPRs only, PP^Q: BP-adjustment was based on SRMs only, PP^R: BP-adjustment was based on both sources, PP^O: BP-adjustment was based on both sources, and using the total sample size in GS:SFHS.

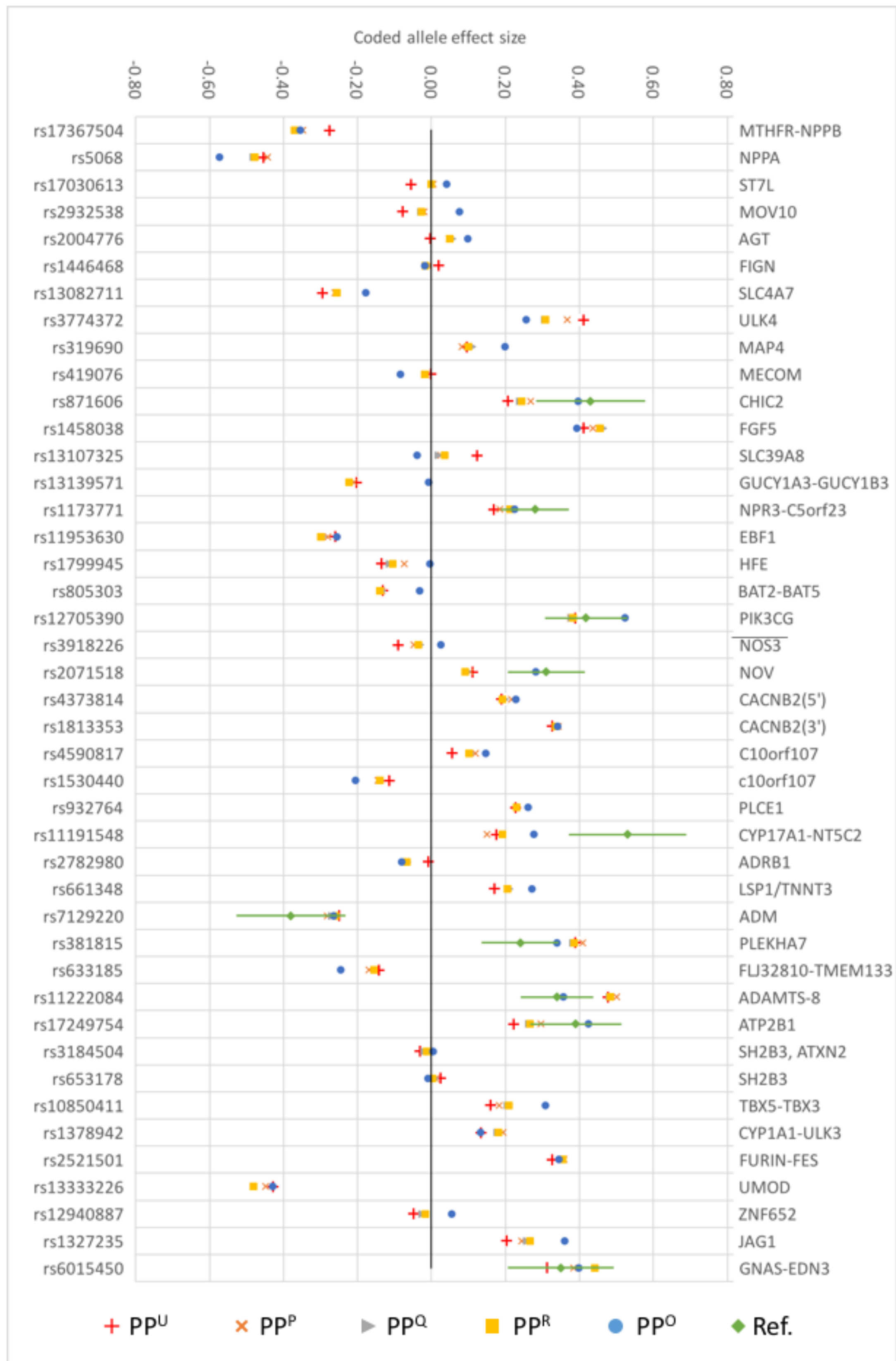


Figure 6-12 Coded allele effect size for PP in the different adjustment models.
SNP names are shown to the left, and loci to the right, with statistically significant loci for PP^O underlined.
Abbreviations, PP^U: Unadjusted (i.e. measured) PP, PP^P: BP-adjustment was based on EPRs only, PP^Q: BP-adjustment was based on SRMs only, PP^R: BP-adjustment was based on both sources, PP^O: BP-adjustment was based on both sources, and using the total sample size in GS:SFHS.

6.3.5 Association of genetic risk scores with BP traits

All the four GRSs were significantly associated with the four BP traits, except for the GRS of PP that was significantly associated with SBP and PP only (Table 6-5). Effect size of the different GRS ranged from 0.29-1.33 mmHg, with DBP GRS having the highest effect size on SBP level, as each one unit increases in GRS of DBP increases SBP level by 1.33 mmHg (95% CI: 1.13–1.53 mmHg). Classifying individuals based on their GRS (i.e. ±1&2 SD) and plotting the deviation from the mean of BP showed that difference between the group with highest GRS and the group with the lowest GRS can reach up to 7.2 mmHg in SBP. Subjects in the highest SBP GRS group had a mean of SBP that is above the population's mean by 3 mmHg, and subjects in the lowest SBP GRS group had a mean of SBP that the below the population's mean by 4.5 mmHg. Similar findings were observed for the other BP traits (Figure 6-13 & Figure 6-14). All the four GRSs showed a significant linear trend for higher mean BP levels with higher GRS groups ($p<7.5x10^{-10}$).

Hypertension GRS was also associated with higher prevalence of hypertension, as represented by a prevalence of 44% in the fourth quartile compared to 34% in the first quartile (Figure 6-15). This was also observed with the prvelance of treatment and awareness across the GRS quartiles were the fourth quartile had higher prevalence of trerated hypertension and awareness (Figure 6-16 & Figure 6-17). For controlled hypertension, the prevalence of controlled hypertension were lower in the third and fourth quartiles of GRS (Figure 6-18), but a linear trend of assocaition was not statically significant.

Table 6-5 Association of the GRS with BP traits

GRS	SBP level		DBP level		MAP level		PP level	
	β (95% CI)	*P*-value	β (95% CI)	*P* -value	β (95% CI)	*P*-value	β (95% CI)	*P*-value
SBP	0.85 (0.72-0.97)	8.67E-39	0.56 (0.47-0.64)	3.07E-40	0.65 (0.56-0.74)	2.09E-45	0.29 (0.20-0.38)	7.62E-11
DBP	1.33 (1.13-1.53)	2.21E-38	0.91 (0.78-1.04)	3.40E-43	1.05 (0.91-1.19)	4.43E-47	0.42 (0.28-0.56)	2.86E-09
MAP	1.18 (0.97-1.40)	8.51E-27	0.80 (0.66-0.94)	1.04E-29	0.93 (0.48-1.08)	1.64E-32	0.38 (0.23-0.53)	6.11E-07
PP	1.00 (0.60-1.39)	7.86E-07	0.21 (-0.05-0.46)	0.11	0.47 (0.19-0.75)	0.01	0.79 (0.52-1.06)	1.28E-08

Columns represent the BP traits, and rows represent the GRS

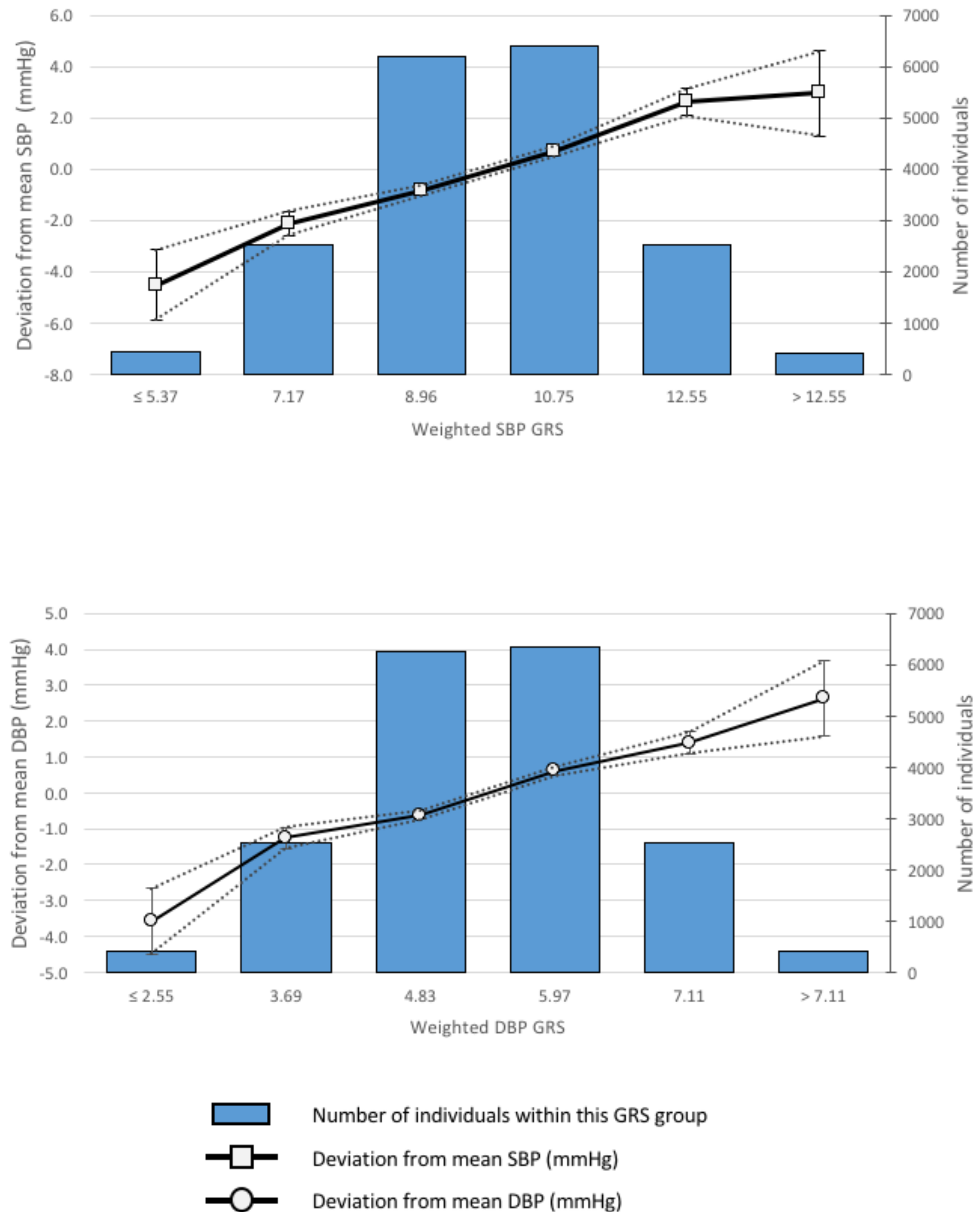


Figure 6-13 GRS for SBP and DBP.
The top figure shows the deviation in SBP (mmHg) from the mean SBP, and the bottom figure shows the deviation in DBP (mmHg) from the mean DBP, represented by the solid line and symbols. The dotted lines and whiskers above and below the solid line represent the upper and lower 95% CI, respectively. The shaded bars show sample size for each GRS category. The p values for slope across GRS were highly significant for linear trends: 3.99×10^{-28}, and 3.84×10^{-33} for SBP versus SBP GRS, and DBP versus DBP GRS, respectively.

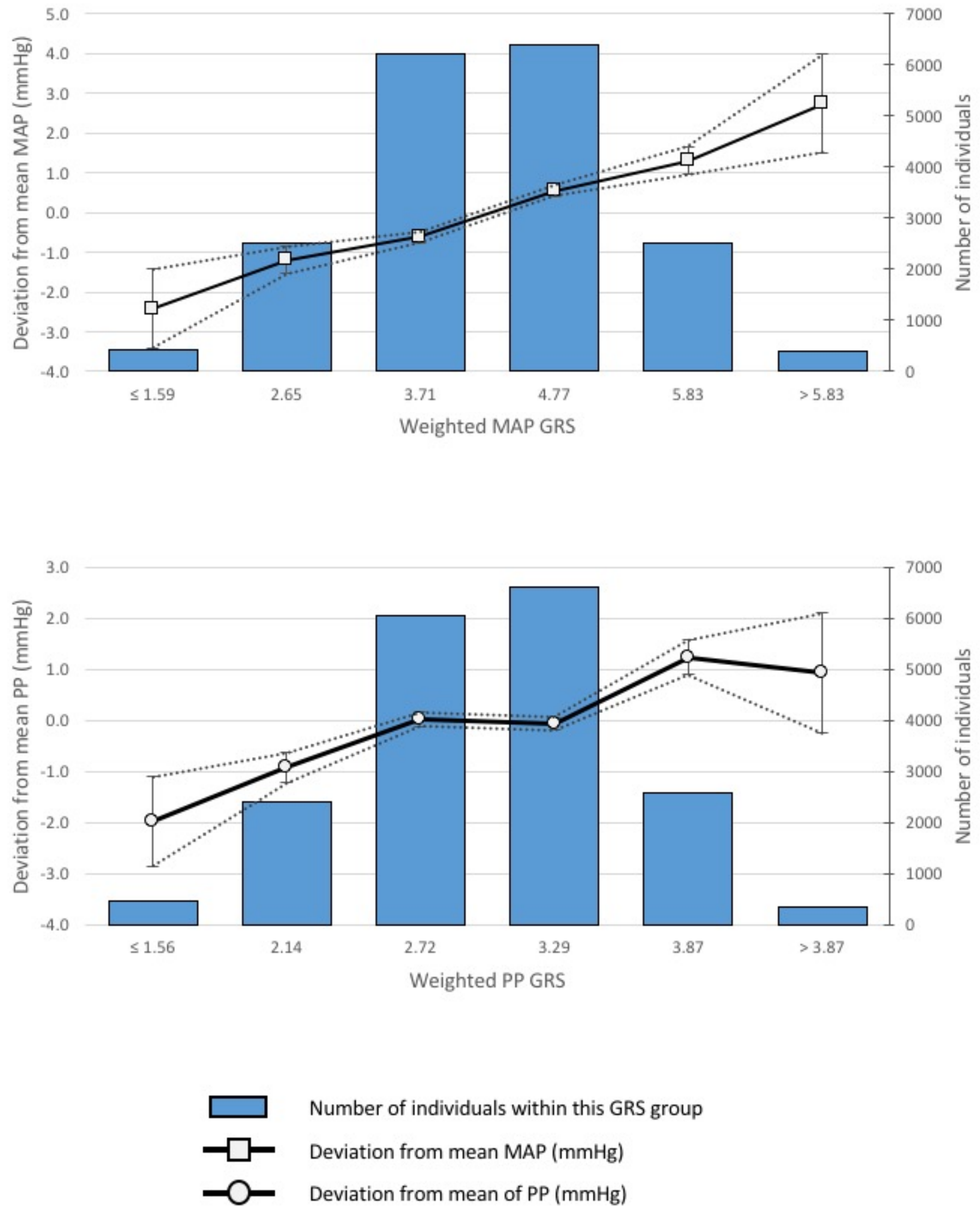


Figure 6-14 GRS for MAP and PP
The top figure shows the deviation in MAP (mmHg) from the mean MAP, and the bottom figure shows the deviation in PP (mmHg) from the mean PP, represented by the solid line and symbols. The dotted lines and whiskers above and below the solid line represent the upper and lower 95% confidence interval, respectively. The shaded bars show sample size for each GRS category. The p values for slope across GRS were all highly significant for linear trends: 1.80×10^{-21}, and 7.56×10^{-10} for MAP versus MAP GRS, and PP versus PP GRS, respectively.

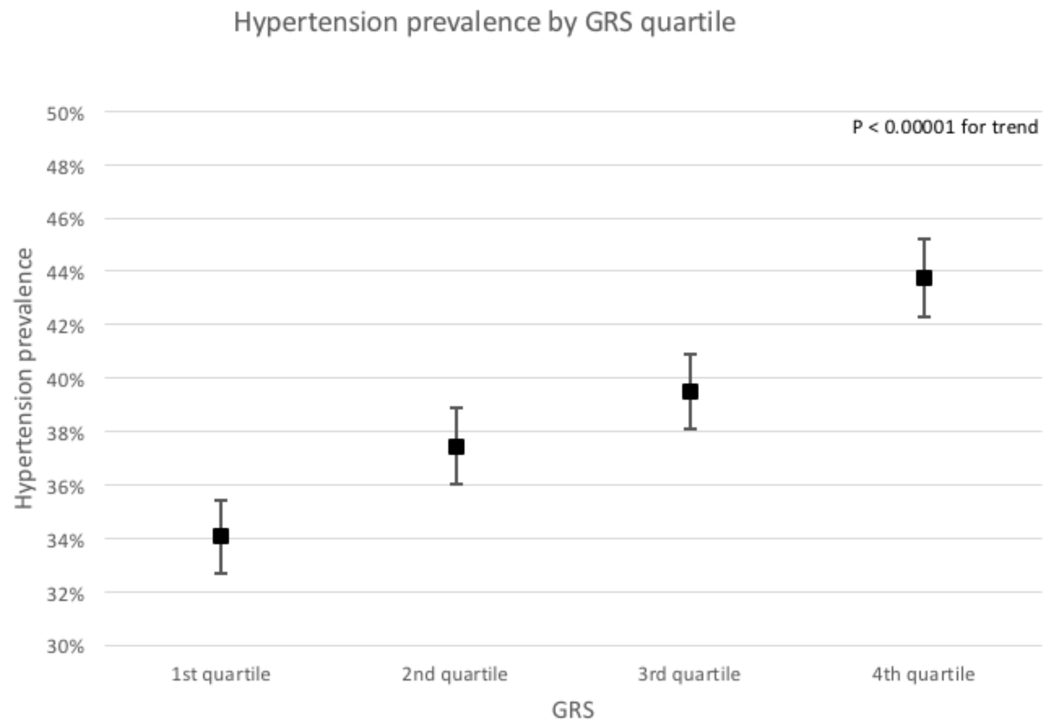


Figure 6-15 Prevalence of hypertension by GRS Quartiles.
Markers shows the prevalence of hypertension in each GRS quartiles with 95% CI; horizontal line shows the p value for linear trend.

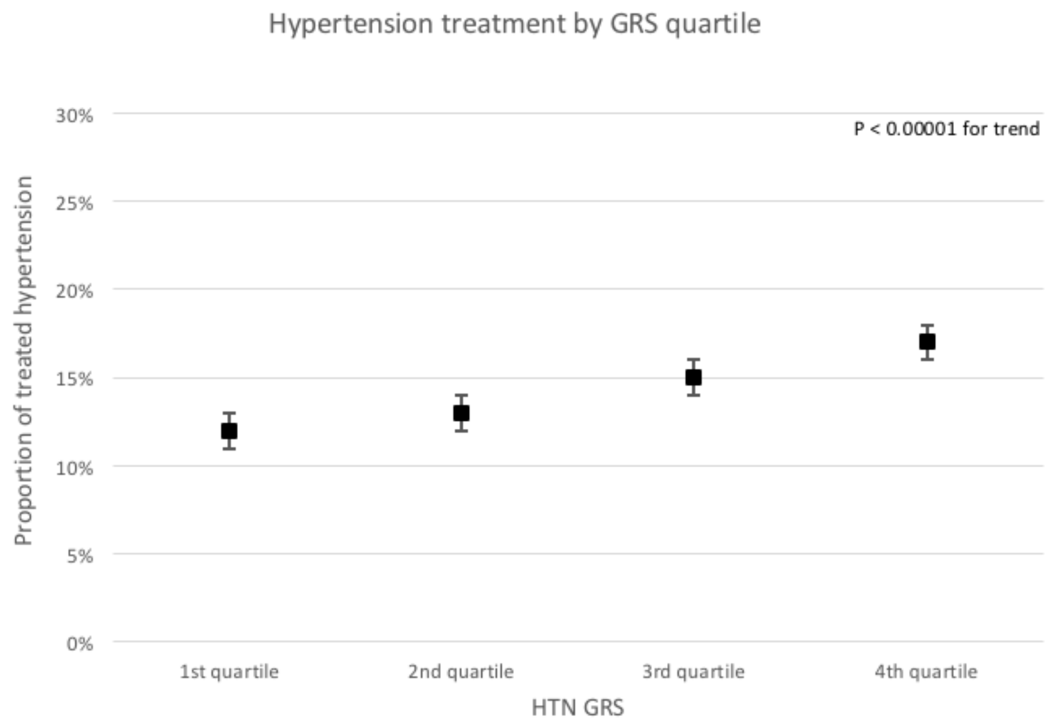


Figure 6-16 Hypertension treatment by GRS Quartiles.
Markers shows the prevalence of hypertension in each GRS quartiles with 95% CI; horizontal line shows the p value for linear trend.

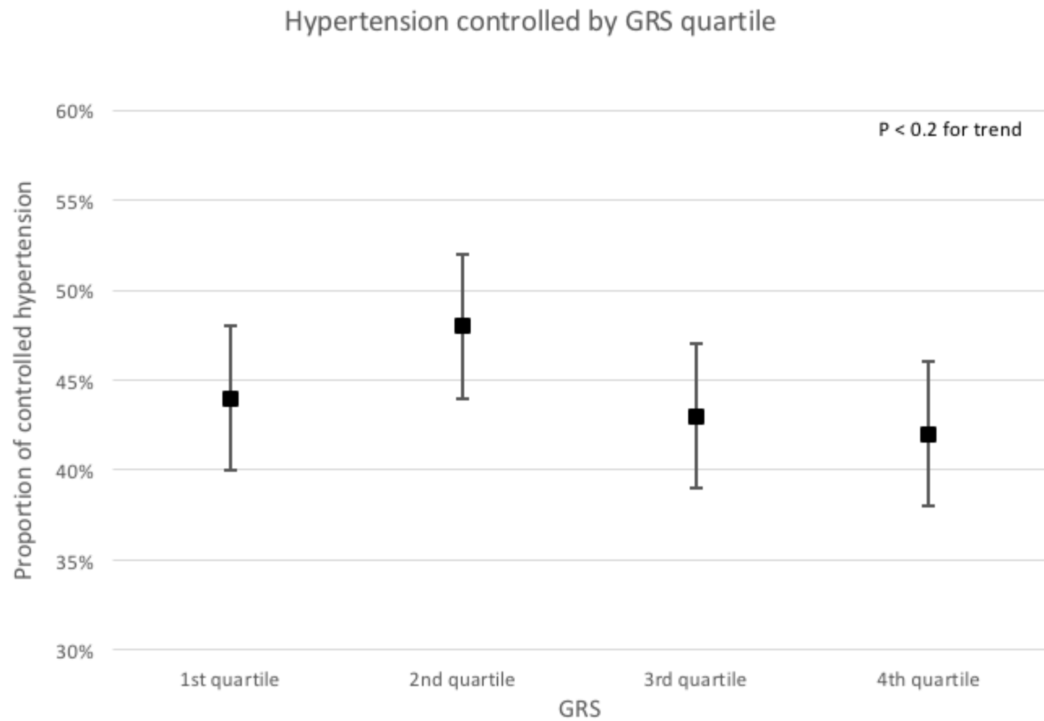


Figure 6-17 Hypertension controlled by GRS Quartiles
Markers shows the prevalence of hypertension in each GRS quartiles with 95% CI; horizontal line shows the p value for linear trend.

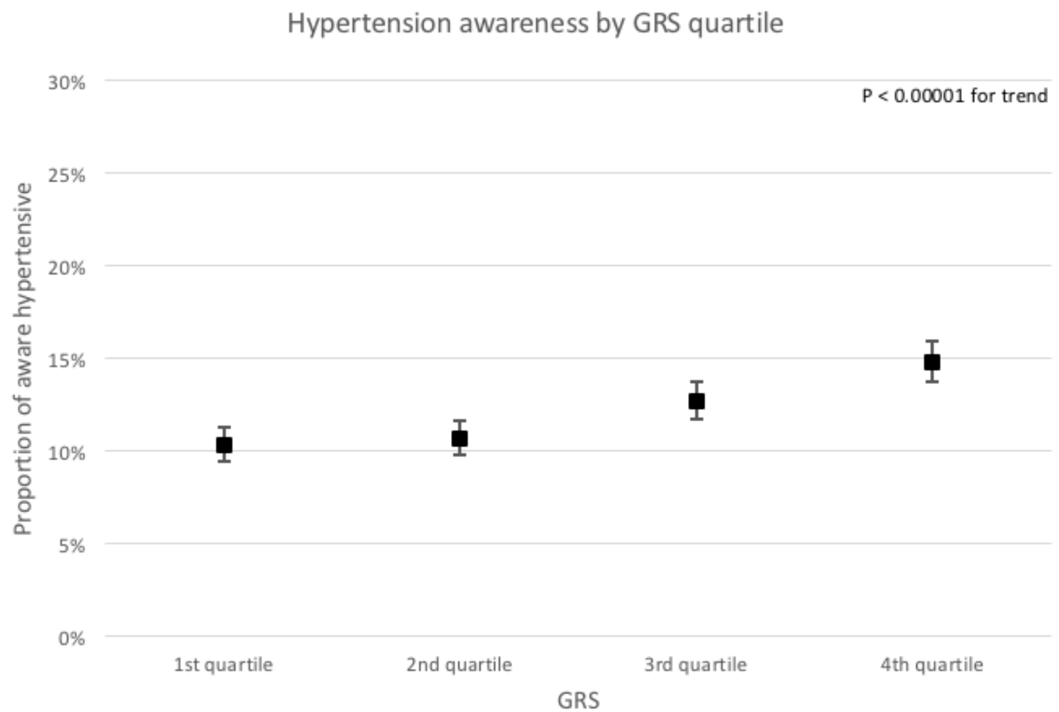


Figure 6-18 Hypertension awareness by GRS Quartiles
Markers shows the prevalence of hypertension in each GRS quartiles with 95% CI; horizontal line shows the p value for linear trend.

6.4 Discussion

The association of the previously identified SNPs in the large GWAS of BP was tested in one of the largest national family-based cohorts. We replicated the association for 11 SNPs at a stringent Bonferroni-corrected significance level, with directional consistency for almost all SNPs with the reported allele effect size and direction. The replicated SNPs at the Bonferroni-corrected significance level were nine SNPs for SBP (*MTHFR-NPPB*, *FGF5, NPR3-C5orf23, CACNB2(3')*, *CYP1A1-ULK3*, *FURIN-FES*, *JAG1*, *GNAS-EDN3*, and *UMOD)*, six SNPs for DBP (*FGF5*, *NPR3-C5orf23*, *CYP1A1-ULK3*, *FURIN-FES*, *GNAS-EDN3,* and *c10orf107*), nine SNPs for MAP (*MTHFR-NPPB*, *FGF5*, *NPR3-C5orf23*, *CYP1A1-ULK3*, *FURIN-FES*, *GNAS-EDN3*, *CACNB2(3'), c10orf107,* and *UMOD),* and one SNP with PP (*PIK3GG)*. For some of these loci, the associations identified in this study were not found with the primary BP traits reported in the reference study. For example, a significant association for the *UMOD* locus with SBP and MAP was found, which has not been reported previously. Assessing the effect of BP-treatment adjustment source on the association analyses has showed only a minimal impact of treatment exposure source compared to the effect of increasing sample size. The association of GRSs with BP traits were all statistically significant, with a linear trend of having higher BP levels in people of the highest extreme GRS.

6.4.1 SNP association with BP traits

The association of rs17367504 (*MTHFR-NPPB*) with SBP was first reported by the Global BPgen consortium for association with SBP, and then validated by the ICBP study which found a further significant association with DBP and hypertension.[107,133] This SNP is located in an intron of the *MTHFR* (methylnetetrahydrofolate reductase) gene in a region with many plausible candidate genes including *MTHFR*, *CLCN6*, *NPPA*, *NPPB* and *AGTRAP*. The precursors for the hormone atrial- and B-type natriuretic peptides (ANP, BNP) encoded by *NPPA* and *NPPB* is located near to this region, which contains rs5068 (modestly correlated [r =0.26] with rs17367504) that was also found to be associated with plasma ANP, BNP, and BP.[134] In this study, rs5068 was associated with SBP, MAP, and PP at the nominal p value only. Population-based studies have also showed that genetically determined increased circulating

concentration of ANP and BNP are associated with lower BP, and reduced risk of hypertension than genetic variants exhibiting lower peptide concentrations.[271] Interestingly, the findings of these GWAS were also corroborated by revealing the underlying molecular mechanisms by which a single base pair change in the *NPPA* gene (rs5068) prevents the binding of MicroRna-425 in subjects carrying the G allele and results in higher ANP levels.[272] This is consistent with the hypothesis that the risk allele reduces ANP levels, and hence increases BP level. Furthermore, abnormality in the natriuretic peptide system has been associated with obesity, type 2 diabetes mellitus, glucose intolerance and essential hypertension.[271]

SNP rs1458038 lies 3.4 kb upstream of *FGF5* (fibroblast growth factor 5), which is a member of the fibroblast growth factor (FGF) family that stimulates cell growth and proliferation in multiple cell types, including cardiomyocytes and has been associated with angiogenesis in the heart.[133] The locus was first reported in Global BPgen for association with DBP and rs16998073, which is in complete LD with the genotyped SNP in this study. The two ICBP studies have also confirmed the association with SBP, DBP, and MAP.[107,135] Studies in the East Asian populations have replicated the association of this variant with BP traits.[137,147] In this thesis, this variant was associated with three BP traits: SBP, DBP, and MAP, with the strongest association signal for SBP. It was also the variant with the highest variance explained individually (h^2 =0.1%, for SBP). In the Pharmacogenomic Evaluation of Antihypertensive Responses Study, the association of this SNP has reached a nominal significance level ($p < 0.05$) for association with better response to atenolol and hydrochlorothiazide therapies, with genotype effects in opposite directions as individuals with the risk allele for hypertension (T) responded better to atenolol than hydrochlorothiazide.[273] This represents one of the very few GWAS SNPs that has showed some pharmacogenetic effects.

The second locus found was to be associated with three BP traits (SBP, DBP, and MAP) in this study is located near *NPR3* (rs1173771). This SNP was first reported by the ICBP study for association with SBP, DBP and HTN in the European populations, and then replicated in the second ICBP study for association with SBP, MAP and PP.[107,135] This association were also reported in other populations,

such as East Asian and African American populations.[137,274] This locus encodes the natriuretic peptide clearance receptor (NPR-C), in which knockout mice exhibit reduced clearance of circulating natriuretic peptides and lower BP.[107] The family of natriuretic peptides elicit a number of vascular, renal, and endocrine effects that are important in the maintenance of BP and extracellular fluid volume. These effects are mediated by specific binding of the peptides to cell surface receptors in the vasculature, kidney, adrenal, and brain.

SNP rs12705390 was the only significant SNP associated with PP in our study. This SNP is in complete LD with SNP rs17477177 reported by the ICBP study for association with PP, and the same SNP genotyped in this study was recently reported for association with LTA-SBP and LTA-PP by Ganesh et al.[135,140] This variant is located 94 kb upstream of *PIK3CG* that encodes the phosphoinositide-3-kinase, catalytic, γ polypeptide protein (PI3Kγ). This protein phosphorylates phosphoinositides and modulates extracellular signals, including those elicited by E-cadherin–mediated cell–cell adhesion, which plays an important role in maintenance of the structural and functional integrity of epithelia.[38] Another association for PP was found for rs11222084 near to *ADAMTS8*, but this association was only significant for PP^P and only at nominal p-values for the full model (PP^O) (p-value =0.0044). This locus was also reported in the ICBP study for association with PP, and is located 1.6 kb downstream of *ADAMTS8*, which is highly expressed in macrophage-rich areas of human atherosclerotic plaques and may affect extracellular matrix remodeling.[135]

The locus *CACNB2* that contains SNP rs1813353 was first identified by the CHARGE study for association with DBP. The CHARGE study reported the association with rs11014166, which is in complete LD with rs1813353 (r =1).[133] The ICBP then validated the association and reported an association with SBP and hypertension.[107] Furthermore, the ICBP identified another independent SNP at the same locus, rs4373814, which was also genotyped in GS:SFHS but showed no significant association with any BP traits. The LTA study has lately reported the association of rs12258967 (r =0.69, with rs1813353) with SBP, DBP, and MAP.[140] This locus was also replicated in the Chinese population.[156] *CACNB2* encodes the β_2-subunit of a voltage-gated calcium channel, and is expressed in the cardiovascular tissue and a loss-of-function missense mutation in *CACNB2*

was identified in affected individuals with Brugada syndrome.[275] Variants in the promoter of *CACNB2* have shown evidence of pharmacogenetic effects in the INVEST-GENES cohort, as Caucasian patients taking CCBs (verapamil) who were homozygotes for the minor allele had increased risk of adverse cardiovascular outcomes relative to patients with AA, or AG genotypes.[276] This finding was also validated in Hispanic patients in the same study, and supported by an in vitro functional study that showed a significant increase in luciferase activity associated with the G allele, suggesting an increase in transcriptional activity compared to the A allele. However, the study did not replicate their findings in an independent sample.

SNP rs1530440 near to C*10orf107* was also replicated in GS for association with DBP and MAP. This SNP was first identified in the Global BPgen for association with DBP only, and then replicated in the ICBP SNPs along with another SNP that is in moderate LD with the Global BPgen (rs4590817, r =0.59).[107,133] Two other SNPs that are in moderate LD were also reported in the LTA study for association with LTA-DBP, LTA-MAP, and LTA-SBP.[140] These SNPs are located in 10q21, which is intronic and belongs to cluster of SNPs in *C10orf107*, an open reading frame of unknown function. This region has no clear neighbouring candidate genes of functional implication in BP.

The second SNP in this study that was found to be associated with three BP traits was rs1378942, in 15q24.1. This SNP was even associated with DBP and MAP in all the three differently adjusted models for BP-treatment exposure source models, using a smaller sample size (n =12,347). This locus was replicated in the two ICBP studies for association with SBP, DBP, and MAP, after it was first identified by the CHARGE study for association with DBP only. CHARGE reported the association with rs6495122 (*r =0.70,* versus rs1378942).[107,115,135] These SNPs cluster are located in the intron region of *CSK* (c-sre tyrosine kinase), and nearby genes include *CYP1A2* (cytochrome P450 enzyme), *LMAN1L* (lectin mannose- binding1 like) and *ARID3B* (encoding AT-rich interacting domain protein). *CYP1A2* is one of the major CYPs in human liver that are responsible for drug and xenobiotic chemical metabolisms, such as caffeine, theophylline, propranolol, and verapamil. A study in 553 young White individuals showed that the risk of hypertension associated with coffee intake

varied according to *CYP1A2* genotypes, with individuals carrying the allele with lower enzyme activity having higher risk of hypertension.[277] The association of *CYP1A2* variants with hypertension was also found to be modified by smoking status, as smoking is a well-known *CYP1A2* inducer. The three *CYP1A2* variants, including rs1378942, were associated with hypertension in non-smokers only, and higher *CYP1A2* activity was linearly associated with lower BP after quitting smoking.[278] In non-smokers, these variants were associated with higher reported caffeine intake, lower odds of hypertension and lower BP. This study has applied a Mendelian randomization approach to provide some evidence that caffeine intake is a modifiable factor for the observed association between *CYP1A2* variants and BP.[278] It must be noted that care is required when interpreting the results from Mendelian randomization studies when the direct product of the gene is not explored, as several conditions are needed to infer causality in observational epidemiology.[279,280]

The third SNP that has showed an association with three BP traits was rs2521501 near to *FURIN*, which was firstly identified by the ICBP study for association with SBP and DBP, and then replicated in the second ICBP study with additional association with MAP.[107,135] In this study, rs2521501 is the only SNP that was statistically significant with all the different treatment adjustment models for SBP, and with rs1378942 for MAP. Variants in *FURIN* were also found to be significantly associated with metabolic syndrome in Japanese individuals.[281] A very recent genome-wide expression quantitative trait loci study has also reported an independent variant in the *FURIN* locus associating a lower expression of *FURIN* with increased BP, possibly by increasing systemic vascular resistance.[282] Rs2521501is 11kb downstream of *FURIN*, an enzyme that belong to the proprotein convertase subtilisin/kexin (PCSK) family, a type 1 membrane-bound protease that processes latent precursor proteins into their biologically active protein and peptides, and encodes a type-1 membrane bound protease that is expressed in many tissues including neuroendocrine, liver, gut, and brain. FURIN was also found to be able to convert the (pro)renin receptor (PRR) into its active and soluble forms in the plasma.[283] Although, the importance of PRR in hypertension has been documented in different studies, the physiological roles of the PRR remain undetermined because the construction of *PRR* knockout mice has not been successful.[284]

UMOD is one locus that was identified from a GWAS using a dichotomous hypertension trait; this study used the BP-extreme approach to allow the maximum phenotypic separation between cases and control, and reported rs13333226, located in the promoter region of *UMOD*.[82] The BP-extreme study reported a suggestive association with SBP and DBP, with a direction of effect that is consistent with the odds of hypertension. A statistically significant association for this SNP with any BP quantitative trait has not been replicated by other studies. However, in this thesis a statistically significant association for rs13333226 is observed with quantitative BP traits (i.e. SBP and MAP). Another variant that is in complete LD with this SNP (rs4293393) was also found to be associated with serum creatinine concentration, uromodulin concentrations, and chronic kidney disease, with a stronger association at older age.[285,286] Yet, the association with hypertension was shown to be independent of renal function, suggesting a possible pleiotropic effect.[82] Functional studies in *UMOD* knockout mice showed that UMOD regulates sodium uptake in the thick ascending limb of the loop of Henle by modulating the effect of tumour necrosis factor-α on NKCC2A expression, making UMOD an important determinant of BP.[287] Significantly, a clear differences was observed in the BP level between UMOD knockout mice and wildtype, in that the knockout mice had significantly lower BP and were insensitive to salt-induced changes in BP. On the other hand, UMOD overexpression in transgenic mice let to salt-sensitive hypertension.[288] These findings highlight the importance of UMOD as a therapeutic target for lowering BP and preserving renal function.

The SNP rs1327235, was only associated with SBP in our study. This SNP was first identified by the ICBP study for association with SBP and DBP, but then conversely, in the second ICBP study an association was found only for MAP and DBP, but not with SBP.[135] An association was also identified in an Asian population with both SBP and DBP.[139] The SNP is located near to *JAG1* (Jagged 1) that encodes one of five cell surface ligands in the highly conserved Notch signalling pathway, which play a critical role in cellular fate determination and is active throughout development and across many organ systems.[289] Mutations in *JAG1* have been associated with several disorders such as the multi-system dominant disorder Alagille syndrome, which has cholestatic, skeletal, cardiac,

ocular, and facial characteristics and includes renal involvement with hypertension.[289]

The last SNP that was found to be associated with three BP traits (SBP, DBP, and MAP) was rs6015450, which is located in an intergenic region near to endothelin-3 (*EDN3)*. This SNP was firstly identified by the ICBP study for association with SBP, DBP, and hypertension, and was also associated with MAP in the second ICBP study.[107,135] The LTA study has also reported the association of another variant rs6092743 that is in moderate LD (r =0.59, versus rs6015450) with SBP, DBP, and MAP.[140] *EDN3* is a strong candidate for BP regulation, as the endothelins are widely expressed vasoactive peptides that exert proliferative, inflammatory, and fibrotic changes in blood vessels and other organs involved in the regulation of vascular tone and BP.[107] *GNAS* encodes the α-subunit of the heterotrimeric G-protein that mediates signal transduction at the β_1 and β_2 adrenergic receptors, influencing heart rate and smooth muscle tone. This genomic region also harbours a variant with pharmacogenetic effect, as SNP rs2273359 was found to influence BP response to hydrochlorothiazide, in which carriers of the minor allele G demonstrated consistently greater BP response to hydrochlorothiazide than the CC homozygotes (lower BP by 7/5 mm Hg).[290] However, the variant with the pharmacogenetic effect is not in LD with the two BP-associated variants (r2 = 0.003, ~189 kb apart). All these variants are intergenic and are unlikely to be the functional variant, suggesting that they may be in LD with undiscovered functional variants that influence gene expression or protein structure.

All the significant variants for SBP and DBP in our study were also significant in the Finnish study, except for the association of rs13333226 in *UMOD* with SBP and DBP, and for the association of rs2521501 in *FURIN-FES* with DBP.[263] Several other loci were significant only in the Finnish Study, which has replicated 22 loci out of the 32 tested variants. This discrepancy maybe due the difference in sample size between the two studies, as the Finnish Study was larger by more than 14,000 subjects. Also, population-specific genetic factors may play roles as demonstrated by finding significant association with *UMOD in* GS:SFHS only. It is impoartant to note that SNPs effect sizes in our study were compared aginst the effect sizes of the combined phase (i.e. joint phase) in the references study, as

the estimated effect size of the discovery phase would be inflated due to winner's curse.

6.4.2 Influence of phenotype measurement errors on association signal

In the second aim of this study, the influence of BP-treatment exposure source on the SNP association signals was assessed by constructing three models; the same sample was used in the three models to eliminate any effect of sample size, while keeping the random model to mimic the phenotypic noise in the overall model. To the best of our knowledge, there is no study that assessed the effect of BP-treatment exposure source in the genetic association signals using a real dataset. Several studies have assessed the concordance between SRMs history and clinical data, and showed a good concordance for antihypertensive medication particularly, and for medication of chronic use generally (discussed in Section 4.5.1 -p160). However, assessing if the association signals can be distorted by incorrect treatment adjustment in some individuals as a result of incorrect classification (i.e. treated\not treated) has never been done previously. Generally, this analysis demonstrated that SBP and DBP were more sensitive to treatment adjustment errors than MAP, which was the least affected with the same replicated SNPs in the two models.

A summary of the concordance results is revisited here to enable a better understanding of the results (discussed in Section 4.5.1 -p160). The concordance between the two treatment-exposure sources at the full sample level was found to be 94% (kappa statistics =78%, Appendix 2 -p251). For 13.1% of individuals, the two adjustments were similar and there was no difference in their classification (i.e. treated or not) between the two sources. Hence, adding a fixed value of 15/10 mmHg to their observed BP levels was justified by the two treatment-exposure sources. For participants with discordant treatment status, however, care is needed in interpreting the results as it depends on whether one source can be considered as superior or not. Generally, EPRs are considered a superior source to measure medication exposures, especially if participants have no access to BP-lowering medication without obtaining a prescription. Thus, the self-reported method had incorrectly classified 4.4% of the sample as treated, leading to unnecessarily adjustment in these subjects, and missed 1.5% of the

sample in which no adjustment was made. In total, the misclassification or phenotypic errors introduced to the BP phenotypes by adjusting the treatment based on self-reported methods was about 6%.

This study reinforces the findings from other studies of the importance of having accurate phenotypic characterisation, particularly for complex traits such as BP that are sensitive to several factors. BP phenotypes derived from treatment adjustment on the basis of EPRs can be considered more accurate than those adjusted on the basis of SRMs. Hence, finding an association signal with SBP^P that is not replicated in SBP^Q would imply that phenotypic noise introduced by BP adjustment in SBP^Q has altered the statistical power, and the similar concept can be applied to the remaining BP traits. For SBP, *FGF5* and *FURIN-FES* were significant with SBP^Q, but only *FURIN-FES* remained significant when using SBP^P. Similarly, *CYP1A1-ULK3* and *FURIN-FES* were significant with DBP^Q, but only *CYP1A1-ULK3* remained significant with DBP^P. For MAP, however, the same two loci (*CYP1A1-ULK3* and *FURIN-FES)* remained significant in both MAP^P and MAP^Q. For PP, *ADAMTS8* was found significant only with PP^P. These observation implies that MAP was more robust to the effect of phenotypic noise as introduced by the different adjustment sources. Moreover, *FURIN-FES* was found significant even with unadjusted MAP that was calculated from the observed BP values, supporting the finding that MAP is less impacted by the treatment status.

This analysis has shown that MAP may provides a better phenotype for genetic association analysis than SBP and DBP, as it was least affected by the phenotypic errors introduced by treatment adjustment. In addition, using MAP as a phenotype has captured all the significant SNPs that were also found to be associated with SBP, DBP, or both of them; with the exception of one SNP that was associated with SBP only (rs1327235, near to *JAG1*). These findings were consistent with previous results in the two ICBP studies, in which the first ICBP study reported the association of 29 SNPs with SBP, DBP or both; and the second ICBP study reported a further 8 new loci that were associated with PP and MAP.[107,135] Importantly, of the 22 SNPs that were associated with both of SBP and DBP in the first ICBP, the second ICBP has replicated the association with MAP for 18 SNPs. For the remaining four SNPs, three of them (*MOV10, BAT2-BAT5, CACNB2(5'))* were not even significant with any of SBP or DBP, and the

fourth SNP near to ZNF652 was found significant with DBP only in the second ICBP (Appendix 6 - p255). Furthermore, a significant association with MAP was also reported with SLC4A7 that was associated with DBP only in the first ICBP. These findings emphasize that SNPs reported for association with MAP are more likely to reflect physiological pathways related to both SBP and DBP [and are less likely to be influenced by slight phenotypic errors].

The four BP traits (SBP, DBP, MAP, and PP) that were examined in this analysis represent different components of BP that specifically reflect distinctive hemodynamic factors (discussed in more details in Section1.2.2 -p24). From a clinical point of view, it was determined that using a combination of two components (SBP and DBP, or MAP and PP) provided a better prediction values for CVD than using a single component.[20,291] In addition, the combination of SBP and DBP has a superior prediction for CVD risk, but the combination of MAP and PP provide greater insight into hemodynamics of altered arterial stiffness versus impaired peripheral resistance. Unlike the traditional components of BP (SBP and DBP), MAP represents a physiological component of BP that reflect the product of CO and peripheral resistance minus central venous pressure.[20,291] Furthermore, MAP is highly correlated with both SBP and DBP, with a phenotypic correlation of 91% and 94% in this study, similar to the reported correlation in other studies (Table 5-3 -p182).[233]

6.4.3 SNPs with potential pharmacogenetic effects

A simulation study performed by Masca et al has showed that methods that adjust for BP treatment by adding a fixed value could be biased for genetic variants that affect the level of response, leading to false positives or false negatives depending on the direction of the pharmacogenetic effect.[292] For instance, a genetic variant that influences treatment efficacy could yield spurious association with BP, or, conversely, a genetic variant that truly influences BP could be masked if it is also involved in a pharmacogenetic interaction. In the other hand, Mascal et al has also suggested that inference about the presence of such interaction could be made by comparing the effect size estimate from the measured BP (i.e. the unadjusted model) and the model adjusted by adding a fixed value. [292] Hence, the presence of the such interaction can be assessed in this thesis by comparing the unadjusted model to the

adjusted model (for instance SBP^U to SBP^R). The random model is chosen here as it represents a model with the average effect size of the treatment adjustment. Based on Mosac et al simulation study, SNPs with large different in effect size between the two models (i.e. adjusted and unadjusted) have potential pharmacogenetic effects.

The top five SNPs with the largest discripincy in the effect size between the two models are in loci near to *SLC39A8*, *GNAS-EDN3*, *MTHFR-NPPB*, *AGT*, and *C10orf107*. The *SLC39A8* was reported to be associated with SBP, DBP, HDL, and BMI, which shows that this variant may has a pleiotropic effects or influence a common pathway that impact these traits.[107,293-294] The *GNAS-EDN3* locus has been previously reported to have a pharmacogenetic effects, as it was found to be associated with BP response to hydrochlorothiazide in black hypertensives.[290] The SNP near to *MTHFR-NPPB* is pharmocogenetically interesting SNP, considering its adjacency to the NPPA/NPPB genes and the concurrent possible association with diuretics treatment in particular. The SNP near to *AGT* is also interested for possible pharmacogenetic interaction as the RAAS is a drug target for multiple class of antihypertensive drug classes. In the other extrem, the bottom five loci with the lowest discripincy between the two models were in loci near to *SH2B3 (*2 SNPs*)*, *BAT2-BAT5, CYP17A1-NT5C2,* and *NOV*. These variants can be assumed to have a minimial possible pharmacogenetic interaction. This exploration of the potential role of pharmacogenetic interaction should be further confirmed by biological and clinical evidences. For instance, by performeing a clinical trail to assess the presence of such pharmacogenetic interaction.

6.4.4 Impacts of genetic risk scores on BP traits

Testing the aggregate association of combined SNPs with the phenotype may provide a better prediction for the risk to the disease than a single SNP, as each SNP individually is likely to confer a low effect size. The present analysis examined the association of four GRSs with the four quantitative BP traits and hypertension prevalence, and showed that GRS is positively associated with BP level and hypertension prevalence, with an average effect of 0.8 mmHg increase per one standard deviation in the significant GRS across the different BP traits. The four GRSs were weighted according to previously reported effect estimates,

and based on SNPs that showed significant association with each trait only. The DBP GRS (i.e. GRS that included significant SNPs for DBP and their weighted effect size for DBP) showed the highest effect size for the four BP traits, even higher effect size than the match GRS for each trait. For instance, SBP level increased only 0.85 mmHg (95%CI: 0.72–0.97) with each one unit of SBP GRS, but increased 1.38 mmHg (95%CI: 1.13–1.53) with each one unit of DBP GRS.

Several studies have assessed the association of the GRS with BP traits, either within the same GWAS such as the CHARGE and ICBP studies, or as a separate study that aim to replicate the findings of GWAS. The CHARGE Consortium assessed the conjoint effect of the top ten SNPs in their study, and showed that the variation explained by them was 1%, and that higher GRS was associated with higher BP level, which can reach up to 5 mmHg and 3 mmHg for SBP and DBP, respectively.[115] The ICBP study constructed a GRS based on a higher number of SNP (29 SNPs), and observed a strong association with both of SBP and DBP in European and non-European ancestry groups (i.e. East-Asian, South Asian, and African).[107] Furthermore, individuals in the top decile of the GRS had a higher prevalence of hypertension (29%) compared with those in the bottom decile (16%), and those in the top quintiles had higher SBP and DBP levels by 4.6 mmHg and 3.0 mmHg, respectively. These GRS were positively associated with left ventricular wall thickness, stroke, and coronary artery diseases, but no kidney disease or kidney function. In this thesis, higher prevalence of treated hypertension and awareness were even observed in those with GRS of the 4 quartiles. The awareness and treated hypertension may represent a clinically diagnosed hypertension that is more strict difention than the study definition of hypertension. In the other hand, controlled hypertension has not shown a trend in the similar direction of the other indices. This study is the first to assess the impact of the GRS on the prevalence of awareness, treatment, and controlled hypertension.

Other studies with smaller sample sizes have also examined the impact of GRS in specific-population samples. A longitudinal study of Swedish individuals demonstrated a positive association of GRS (based on 29 SNPs) with higher BP and incidence of hypertension independently from traditional risk factors; however, the result of discrimination analysis does not show any improvement in

the prediction of incident hypertension on top of traditional risk factors.[295] Another study in a Finnish population conducted by Havulinna et al. concluded that the two GRSs based on the 29 SNPs were strongly associated with risk of incident coronary artery disease, stroke and composite CVD.[263] These findings were also confirmed recently in the Swedish population by findings that GRS was independently associated with ischemic stroke in three Swedish case-control studies.[296] However, both studies reported that GRS did not improve CVD risk discrimination over and above the traditional risk factors. A very recent study in the Korean population has also reached to the same conclusion, although they only used four SNPs in their GRS.[297]

While these studies have demonstrated a modest improvement in risk discrimination for GRS over the traditional risk factors, the importance of such factor should not be disregarded as it has been shown that modest increments in BP level, even if based on a single BP measurement, are associated with substantial increases in CVD risk.[3,5,6] Moreover, individuals with higher GRS are consistently exposed to the impact of these genetic variants over their life time, meaning that even a small increase in BP may translate into comparatively large effect when compounded over a life time. An important factor that needs to be considered is the fact that a risk scoring system such as the Framingham Risk Score is limited to a specific time-period (between 10 to 30 years), as risk factors included in the scores change over time. However, a risk score based on GRS is a lifetime score that is invariant over time.

An important limitation in the application of GRS in this study, and most similar studies so far, is not considering any possible interaction between the included genetic variants themselves, and/or other factors such as demographic or environmental factors. Furthermore, the included variants are from GWAS, which often discovers SNPs in genomic regions that have no clear physiological impact on BP. Hence, including genetic variants that are identified from newer GWAS or pathway-based analysis could improve the prediction of GRS. In addition, further studies are required to clarify whether different scores are need in different populations, or the need to account for any confounders before applying the GRS

7 General discussion and conclusion

The studies comprising this PhD thesis provide a unique insight into the genetic and epidemiological features of hypertension, through a large family-based cohort study (GS:SFHS). High BP is the leading risk factor for mortality worldwide, responsible for 13% of deaths globally. While more and more people are being treated for hypertension, the numbers of people with uncontrolled BP remain largely unchanged, and many people still remain undiagnosed. The diagnosis, treatment and follow-up of hypertensive patients is a core part of primary care workload and successful management of hypertension requires an engaged partnership between clinicians and their patients. Hypertension epidemiology in terms of prevalence, awareness, treatment and control is conducted through health surveys or cohort studies of unrelated individuals, and interventions are delivered on an individual basis. The family-design and population-based sampling approach taken in GS:SFHS allows investigation of genes, environment, and hypertension epidemiology with a valid inference to the population.

The strengths of GS:SFHS include participants that were recruited from across Scotland, which has a high prevalence of common diseases such as coronary heart disease, stroke, cancer, and diabetes. Even compared to other UK countries, Scotland had a higher rate of mortality that is partially explained by baseline risk factors. [169] Furthermore, the population in Scotland is relatively static and stable, providing an ideal population to measure heritable and lifestyle factors for complex traits. Although GS:SFHS cannot be considered truly representative of the Scottish population, it includes a large sample size with a wide range of socio-economic and clinical features. Furthermore, the availability of clinical, self-reported, and prescription data has allowed an assessment of management trends of hypertension in this cohort, which can be generalized to the Scottish population. While the family-based nature of the cohort represents one of the important strengths of GS:SFHS, the availability of high-quality DNA samples make it a good candidate for future exome and whole-genome sequencing analysis. The current study can serve as a first step to evaluate the genotypic and phenotypic features of GS:SFHS in relation to BP, and inform selection of individuals or families to undergo further detailed genomic analysis, such as whole-genome sequencing.

A significant variation in hypertension management trends were observed between countries and populations, which may reflect different population-specific factors that may contribute to such variation including genetic, environmental, public-health policy, and socio-economic factors. Knowing the presence of such variation between populations and countries necessitates a cautious interpretation of studies from other populations, and also that each population should be independently examined to identify factors playing a role in this population. Nevertheless, the importance of lowering BP by either pharmacological or non-pharmacological approaches have consistently been demonstrated in all populations, such that a 10 mmHg reduction in SBP is associated with a 22% reduction in CHD and a 41% reduction in stroke.[70] On the other hand, even a small incremental increase of 2 mmHg can increase the risk of stroke by 10% and CHD by 7%.[5] It is therefore important to assess even the factors that impose a small effect size, such as common genetic variants.

A family-cohort design of related individuals with adjustment for family-clustering would likely account for important residual or unmeasured confounding factors, which could bias both the main effect and the interaction estimate.[167] However, this cannot be ruled out in observational studies as a result of differential recall or ascertainment bias. Furthermore, the statistical power for gene discovery is likely to be higher in studies containing extended-pedigree, as they represent a more homogenous sample. Member of families are more likely to share both genetic background and environmental factors exposure, and hence analysis of phenotypes can be modelled by genetic and environmental factors. For genetic background, a polygenic component can be modelled based on the genetic background that can be derived as a function of the degree of relationship. For instance, monozygotic twins share the extreme of genetic background, suggesting that phenotypic variation is to due to epigenetic, environmental, or interaction factors. The extent of shared background decreases as kinship coefficient decreases, leaving the remaining variability to environmental factors. Similarly, close-relatives tend to have more homogeneous environmental factors, as they are more likely to be living in the same house with similar socioeconomic status, and may share similar lifestyle habits such as diet, physical activity, smoking, and alcohol consumption. Whilst strictly controlling for these factors as in animal studies is not possible, the

residual noise variance is lower in family studies comparing to population studies with unrelated individuals. This degree of nature control for both genetic background and environmental factors enhances the power to detect novel associations as a result of reducing the residual noise variance.[167]

The family-based design is typically assumed to be less powerful for genetic association studies than a population-based design, because the addition of a non-independent subject to the sample is considered not equal to one whole extra unrelated person, but only adds a fraction of information depending on the degree of familial correlation.[167] However, other studies have showed that the opposite might actually be true, as the power to detect an effect can be increased by including relatives and conducting a multi-level analysis to accommodate familial correlation.[168] Even the addition of an identical twin that has perfectly correlated genotype, phenotype, and residuals cannot be considered a redundant duplication, as neither complex phenotypes nor residuals are identical in any pair of twins. A critical determinant of study power is error variance, where the power to estimate all model parameters is increased as the unexplained error variance decreases. For family data, a higher proportion of the error variance can be explained by the extra information that is not available in unrelated individuals.[168]

Family data can offer a deeper level of genotype QC, especially with respect to the detection of Mendelian errors. This can also be an important step to detect and possibly resolve any sample mix-ups. Moreover, a particular advantage that was also applied in this study is to impute the missing genotypes, by inferring the most likely genotypes based on the observation in other genotyped relatives. This is performed based on information at other markers and family relationship. In this study, the imputation was limited to a subset of SNPs that were not recalled for technical reasons. However, this approach can be extended to impute the genotype of relatives that were not actually genotyped in the study with available phenotypes, leading to more efficient and powerful statistical test.[298]

This study was able to determine the familial risk of hypertension in Scotland, showing that subjects with one or two hypertensive parents have 42% or 95% higher risk of having hypertension, respectively, compared to participants

without any hypertensive parents. Whilst the rates of hypertension awareness and treatment in GS:SFHS were, respectively, 25.3% and 31.2%, the rate of treated hypertension in GS:SFHS was much lower than the rate of treated hypertension reported in the SHeS of 2008/2009 and 2010/2011 (49% and 48%, respectively). The rate of controlled hypertension in GS:SFHS was 54.3%, with higher rates in women (60%) compared to men (47.8%). Furthermore, parental history was inversely associated with control rate, such that participants with two hypertensive parents had a lower rate of controlled hypertension. These findings open up avenues for further research into implementing hypertension care in the community, using family as a unit of intervention, and assessing if this can improve awareness and ultimately improve the level of controlled BP across the country.

The availability of self-reported medication and pharmacy refill records allowed an assessment of phenotypic accuracy when the hypertension phenotype is based on a history of taking antihypertensive therapy. Interestingly, around 10% of the hypertensive population reported taking antihypertensive therapy without any corroborative evidence from pharmacy data. As all hypertensive drugs in Scotland are obtained through pharmacies from primary care prescriptions, it is unlikely that these subjects were sourcing their antihypertensive drugs from elsewhere. Thus, this indicates the misclassification rate, if SRM was used to classify subjects as hypertensive, can be high and indicates the importance of corroborating SRM history with other records. Assessment of the SRMs reliability before performing epidemiological and genetic analysis allowed an approximate estimation of the phenotypic errors. For instance, 4.7% of hypertensive subjects were classified as hypertensive only because they have reported taking BP-lowering medication in SRMs, with no confirmation in EPRs. In quantitative analyses, the proportion of participants who were incorrectly exposed/not exposed to treatment adjustment was 6%. To the best of our knowledge, these findings have not been previously reported for BP-lowering medications, as most BP genetic studies had only one source of medication history, and do not assess the reliability of treatment status.

An important step in this study concerned the quality of data available in the GS:SFHS, with a range of procedures undertaken to ensure high quality data was

used in the analysis. Amending the PCQ structure with the introduction of some questions in the middle of the study recruitment period raised several challenges to optimally merge all the information. This led to losing some information, due to the inability to combine participants' answers between the two phases. The other challenge was in transforming the inconsistent medication names that were hand-written by participants into standard medication names to retrieve the participant SRMs in PCQ-1. The availability of EPRs and SRMs has allowed a critical step to assess the reliability of reporting BP-lowering medications, and the chance to increase sample size to include those participants without EPRs in the familial and genetic analyses. Assessing the adherence to BP-lowering medication would be possible given the presence of these two sources of information; however, the amount of prescribed medication (i.e. prescribed pill count) was not available in all the prescriptions, and thus it was not possible to calculate the frequency of prescriptions.

The clinical and health implications of the epidemiological part of this thesis can be summarized in as follows: First, although the SHeS has shown some improvement in the management trends of hypertension, this study indicates significantly more improvement is still needed in the detection and treatment of hypertension. Second, the rate of awareness, treatment, and controlled hypertension were significantly lower in the least deprived area of Scotland. This was in accordance with report from other studies that participants from more deprived areas received at least equivalent and sometimes higher quality of care for hypertension than those from less deprived areas.[227] Hence, further work is needed to raise awareness in the least deprived areas about the risk of undiagnosed hypertension, as these individuals may make fewer visits for doctors and hence have a lower chance of getting their BP reviewed. Third, participants with parental history of hypertension were significantly more likely to be treated and aware of their hypertension and less likely to be controlled compared to those without any parental history of hypertension. Though, for controlled hypertension the association was not statistically signifincant. Therefore, a history of hypertension should instigate a review of all eligible family members for hypertension status and treatment, if required.

Several statistical methods were used in this study to assess the familial aggregation of hypertension and BP traits. Firstly, the familial influences on the hypertension prevalence and treatment were assessed by reporting λ_S of 1.6, and of 2.04, respectively. These values were calculated using a counting method that is assumed to be unbiased and consistent when applied to samples that are not ascertained via any particular phenotype.[243] Secondly, the familial correlation of BP traits was much higher among first-degree relatives than other types of relatives, with higher correlation between siblings than parent-offspring, reflecting the level of shared genetics. Thirdly, the heritability of BP traits ranged from 24% to 32%, which is within the range reported in other populations. The familial aggregation of hypertension and BP can be attributed to shared environmental and genetic factors. These two factors can be intimately connected and it is difficult to partition their independent influences, as some of the environmental factors such as obesity and sodium intake have a heritable component themselves. For instance, the bivariate analysis between BP traits and BMI showed that shared genetic factors explain 54% of the phenotypic correlation. Lastly, the genetic correlation between the different BP traits suggests that SBP, DBP, and MAP share common genetic factors, but the observed incomplete pleiotropy indicates the presence of an independent set of gene(s) controlling each BP components. Interestingly, PP showed the lowest correlation with the three remaining BP traits, suggesting that PP may be influenced by a different set of genes.

SNPs association tests were performed using an emerging method that is based on LMM approach, to test for association in the presence of sample structure. LMM methods perform a total association test that includes both within-family and between-family effects, and hence they offer a comprehensive approach that has been found to substantially outperform other typical family-based statistical methods.[265] However, as this study only genotyped a small number of markers, pedigree data were used in constructing the GRM rather than genetic data. This was done by calculating theoretical pair-wise kinship coefficients between each two individuals. Hence, the accuracy of the constructed GRM relies on the quality of pedigree records, and is therefore sensitive to errors. An alternative approach that is usually applied in large GWAS of unrelated individuals with a high density of genotyped markers is to construct GRM on the

basis of genotyped markers, rather than being fixed at their known theoretical values, as implemented in software such as GCTA.[299] The marker-based GRM can overcome issues of incomplete pedigree, inbreeding, selection, and drift. However, comparing a model fitted with marker-based GRM to a model fitted with pedigree-based GRM has showed comparable results.[125] Estimating the impact of the pedigree-based GRM in our study is not possible without having a large number of markers to obtain accurate estimates of both population structure and relative kinship within the sample.

The association of 43 SNPs in 38 loci that were reported in large GWAS meta-analysis of BP was assessed in GS:FHS, showing 11 SNPs were significantly associated with at least one BP trait at a stringent Bonferroni corrected significance level ($p<0.0012$), and 36 SNPs at a nominal p-value ($p<0.05$). Almost all the SNPs showed directional consistency with the reported allele effect size and direction in the reference study. SNPs that were statistically significant were in loci near the following genes: *MTHFR-NPPB, FGF5, NPR3-C5orf23, PIK3CG, CACNB2(3'), C10orf107, CYP1A1-ULK3, FURIN-FES, UMOD, JAG1,* and *GNAS-EDN3*. All the significant SNPs were associated with the primary traits as reported in the reference study, with the exception of the *UMOD* locus (rs13333226) which was found significantly associated with SBP and MAP. The *UMOD* locus (rs13333226) was reported for association with hypertension with only a suggestive association with SBP and DBP.[82] The total variance explained by all the 43 genotyped SNPs was 1.4%, 1.5%, 1.6%, and 0.8% for SBP, DBP, MAP, and PP, respectively. Considering the heritability estimates from the family-based analysis, these common variants on average explained less than 0.05 of BP heritability [e.g. 0.014/0.30 (h^2)]. Therefore, this study has showed that the proportion of h^2 that is explained by BP SNPs identified by large GWAS was approximately 5%. This is much lower than the proportion of 80% for SBP reported by Vattikuti et al. based on more than 400,000 markers.[160] Consequently, this discrepancy between the two numbers is a result of the difference in number of markers that were used to estimate the h^2_{SNP}, supporting the theory that several markers are yet to be discovered for BP traits.[107]

The problem of missing heritability has attained much focus in the field of complex traits genetics, for which proposed reasons include, rare variants in novel pathways that are undetectable by the traditional GWAS approach; many more variants with smaller effect size that need even larger sample sizes to capture the variants; poor modelling of the genetic effects by ignoring the epistasis factors; structural variants such as CNVs that are poorly captured by existing genotyping platforms; over-estimated heritability in the family-based studies; and inadequate accounting for shared environment among family members.[159] GWAS with larger sample sizes and a wider range of allele frequency continue to detect additional variants, and this implies that variants detected so far represent only a small fraction of those that influence a given trait. The ICBP study predicted that there are potentially 116 variants yet to be discovered with similar effect size to those identified so far.[107] The phenotypic variance explained by these variants is only 2.2%. Hence, dissecting the genetic architecture of complex traits is a challenging process that requires larger sample size, better phenotyping, consideration of non-genetic risk factors, focused study designs, and an integration of multiple sources of phenotypic and genetic information.[300]

Given the complexity of traits, such as BP, and the shortcomings of GWAS to explain much of the phenotypic variance, this has resulted in great interest in exploring the role of rare variants. This is also made possible by the increased affordability of exome- and whole-genome sequencing. However, this is more challenging as association studies that focus on rare variants (MAF <1%) require even larger sample sizes than that currently used in the largest GWASs. Essentially, finding rare variants with large effect size do not equate to explaining a large proportion of phenotypic variance at the population level. It is important to reiterate that complex diseases are more likely to have highly polygenic architecture that is consistent with the following features: (a) rare cases clustered within families with likely Mendelian forms of the disease, (b) a majority of cases at population level with little or no family history, (c) two different cases may carry different sets of risk alleles (genetic heterogeneity). Hence, dissecting the genetic architecture of such complex traits requires a combination of large-scale GWAS and more targeted approaches at both the population- and family-level to identify the remaining genetic variants.

Phenotypic accuracy represents a major challenge for most genetic studies, and can impact the effect size of genetic association or even the proportion of phenotypic variance explained by the findings of GWAS.[81,85] Heritability estimates can also be influenced by measurements errors, as it was found that heritability of more accurate BP phenotypes, using approaches such as LTA or ABPM, resulted in higher heritability estimates.[239,301,301] Combining individuals based on a broad clinical definition such as hypertension can average the effect size across them, while in fact they may have distinct underlying casual variants. Preferably, individuals should be classified into smaller groups on the basis of related phenotypes to reduce the chances of genetic heterogeneity. Advancement in high throughput "omics" technology can provide phenotypes that are closer to the level of gene action "endophenotypes" such as gene transcription, microRNA levels, proteomics, and metabolomics. For instance, the effects of a mutation that alters transcription factor binding can be clearly observed at the level of gene expression but may not be detected at the level of disease risk.[81] Another way to improve the phenotype is to combine multiple phenotypic information for the same subjects and adopt a multivariate approach by jointly modelling multiple traits, or by stratifying the phenotype by the values of another phenotype, similar to what has been done in stratifying type 2 diabetes cases based on their BMI values.[303,304]

The impact of BP-lowering medications on quantitative genetic studies has been demonstrated in several studies.[87] However, no previous study has examined the impact of using SRMs or EPRs as a basis for adjustment of treatment on subjects taking BP-lowering medications. Although the pharmacoepidemiological analyses have showed good concordance between the SRMs and EPRs, the effect of treatment-exposure source was observed in the SNPs association signals, especially for SBP and DBP. Interestingly, MAP was the least impacted BP trait with overlapping findings in the phenotypes generated by both SRMs and EPRs. This may imply that MAP is a more robust BP phenotype that is less impacted by the BP-lowering treatment. Further studies are required to explore if MAP represent a better phenotype for genetic association studies than both of SBP and DBP. The question of whether MAP provides a better phenotype than both of SBP and DBP, or it simply reflects different physiological pathways require further work. Using MAP only for dissecting the genetic architecture of BP can:

(1) capture genetic variants that show strong association with both SBP and DBP, (2) provide better phenotype quality that are less impacted by BP-lowering medications, and (3) reduce the multiple testing penalty (i.e. testing the association of SNP with one phenotype instead of two phenotypes). Importantly, MAP is one component of BP that represents a physiological (rather than traditional) component of BP, and corresponds to the product of CO and PVR minus central venous pressure. In this study, MAP was highly correlated with both SBP and DBP (ρ_p of 0.91 and 0.94, respectively), and the genetic correlation was even higher (ρ_G of 0.94 and 0.97, respectively).

This thesis was limited by some factors due to the available resources; first, the BP values used are based on two measurements at a single time-point. This is clearly not the optimal BP phenotype, as discussed in this thesis. Second, changing the PCQ structure in the middle of the study has limited the use of some information. Third, the findings of this study may not be applicable to other populations, particularly of non-European ancestry, as this study was restricted to the Scottish population. Fourth, although the cohort recruited in the study can be considered the largest family-based available, the sample size is still not adequate to detect low frequency variants of modest effect reported in previous GWAS, even at nominal $p < 0.05$.

Appendix

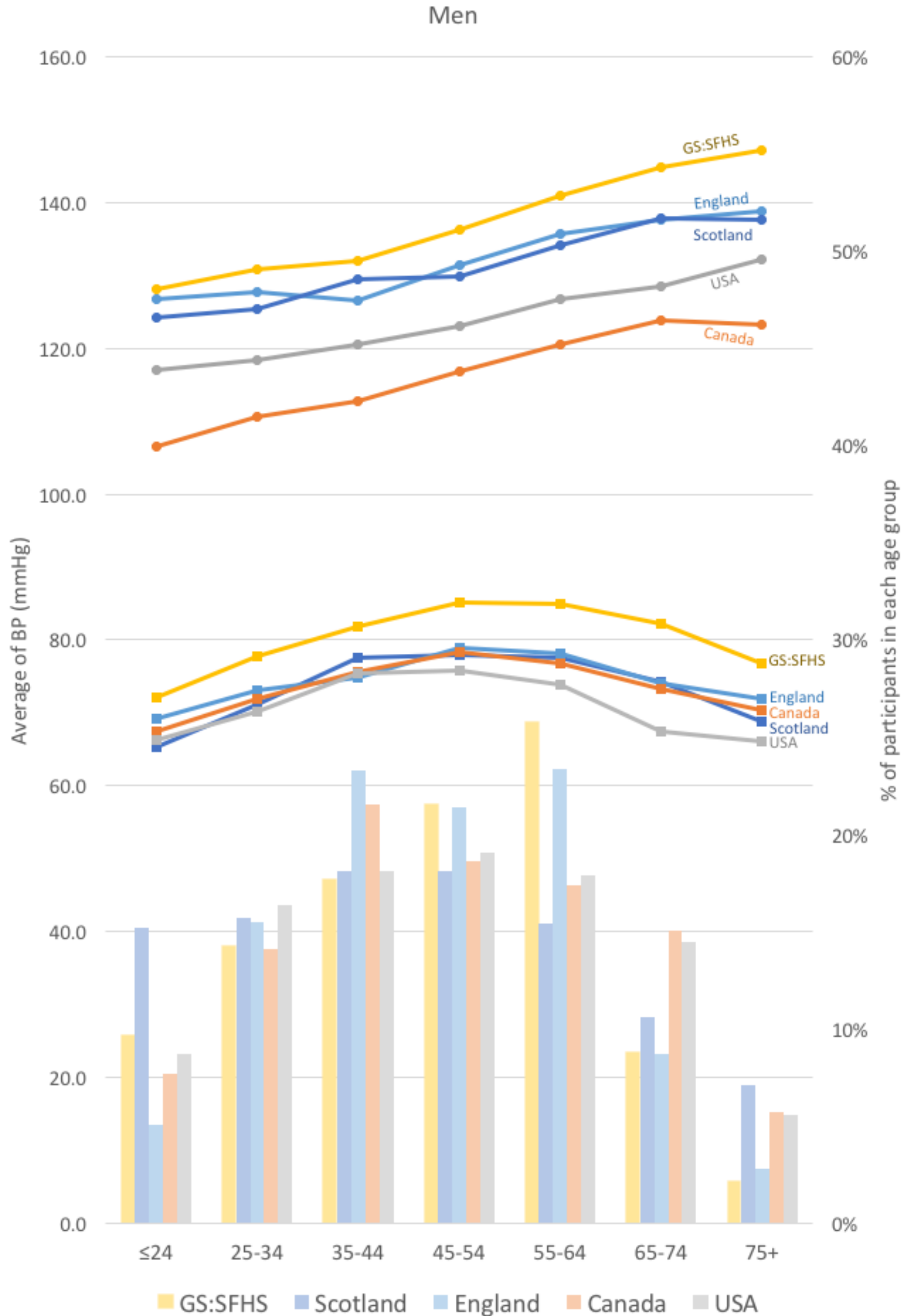
Appendix 1: 2011 Mid-year household population estimates for Scotland by age and sex.

Age group (years)	*Men*	*Women*
18 – 24	252,000	253,100
25 – 34	337,900	332,300
35 – 44	335,500	365,380
45 – 54	370,270	405,150
55 – 64	319,520	337,250
65 – 74	221,450	253,430
75+	152,370	231,530
Total	1,989,010	2,178,140
Data are reproduced from http://www.gov.scot/Publications/2012/09/8038/1		

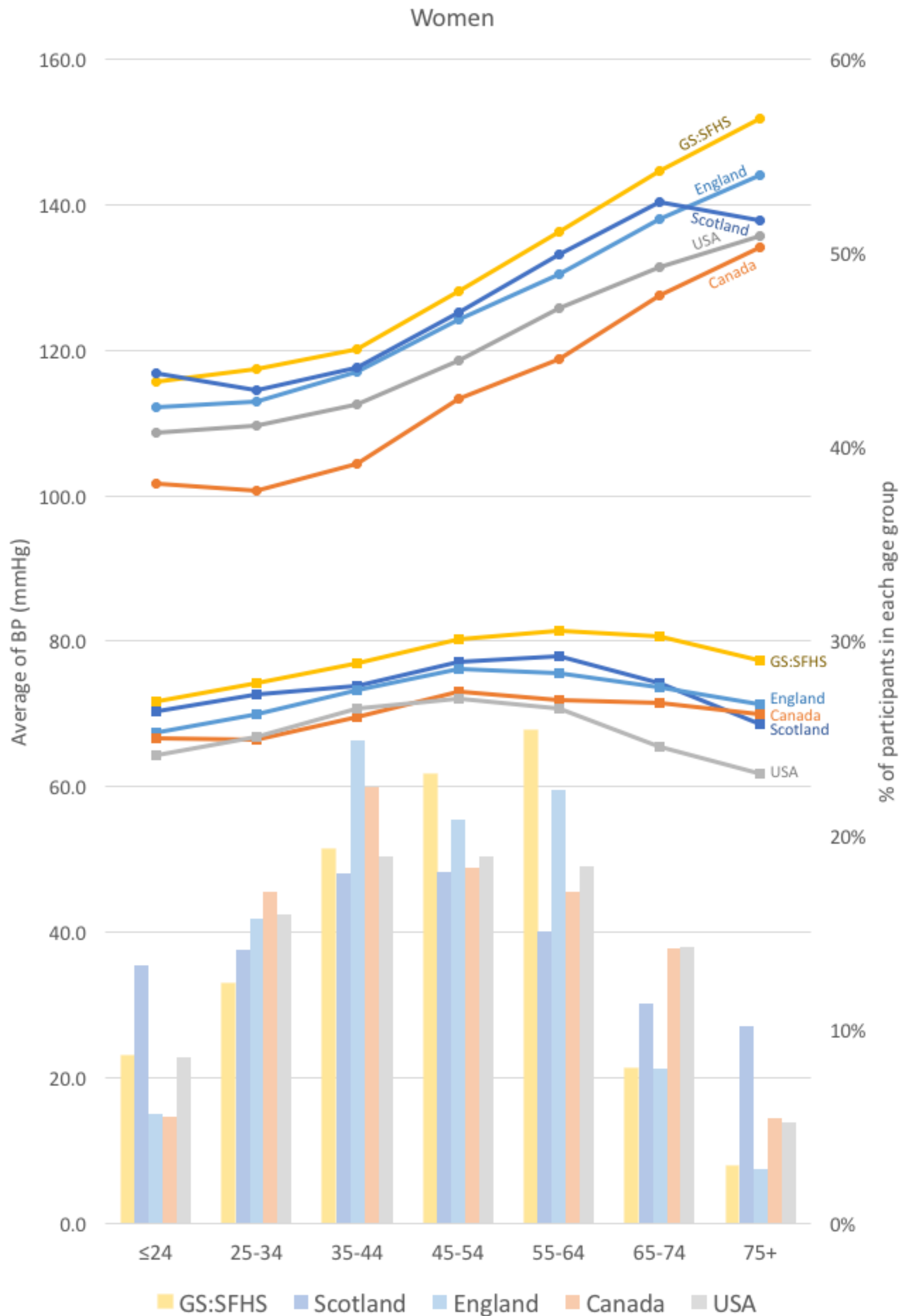
Appendix 2 Concordance of BP-lowering medication in the total population

		$Treatment^{Prescr}$		
		Yes	No	Total
$Treatment^{Ques}$	Yes	1624 13.1% [12.5%–13.7%]	542 4.4% [4.5%–4.7%]	2166 17.5% [16.9%–18.2%]
	No	184 1.5% [1.3%–1.7%]	9997 81% [80.3%–81.6%]	10181 82.5% [81.8%–83.1%]
Total		1808 14.6% [14.0%–15.3%]	10540 85.4% [84.7%–86.0%]	12,347
Measurements of validity	Sensitivity	90%	Specificity	95%
	PPV	75%	NPV	85%
Measurements of reliability	p_o	94%	p_e	73%
	K	78%	PI	68%
	BI	3%	PABAK	88%
	P_{pos}	82%	P_{neg}	96%
$Treatment^{PCQ1}$	Yes	673 11.7% [10.9%–12.5%]	316 5.54% [4.9%–6.1%]	989 17.3% [16.3%–18.3%]
	No	60 1.1% [0.8%–1.3%]	4678 81.6% [80.6%–82.6%]	4738 82.7% [81.7%–83.7%]
Total		733 12.8% [11.7%–13.7%]	4994 87.2% [86.3%–88.0%]	5727
Measurements of validity	Sensitivity	92%	Specificity	94%
	PPV	75%	NPV	85%
Measurements of reliability	p_o	93%	p_e	74%
	K	74%	PI	70%
	BI	4%	PABAK	87%
	P_{pos}	78%	P_{neg}	96%
$Treatment^{PCQ2}$	Yes	951 14.3% [13.5%–15.2%]	226 3.4% [3%–3.8%]	1177 17.8 [16.7%–18.7%]
	No	124 1.9% [1.6%–2.2%]	5319 80.4% [79.4%–81.3%]	5443 82.2% [81.3%–83.1%]
Total		1075 16.2% [15.4%–17.2%]	5545 83.8% [82.9%–84.6%]	6,620
Measurements of validity	Sensitivity	88%	Specificity	96%
	PPV	81%	NPV	84%
Measurements of reliability	p_o	95%	p_e	72%
	K	81%	PI	66%
	BI	2%	PABAK	89%
	P_{pos}	84%	P_{neg}	97%

Appendix 3 Average of BP in men by age groups in GS:SFHS, Scotland, England, USA, and Canada.
Left vertical axis shows the average of SBP and DBP by each age groups, the right vertical axis shows the percentage of participants fall in each age groups as represented by the bar chart.



Appendix 4 Average of BP in women by age groups in GS:SFHS, Scotland, England, USA, and Canada.
Left vertical axis shows the average of SBP and DBP by each age groups, the right vertical axis shows the percentage of participants fall in each age groups as represented by the bar chart.



Appendix 5 SNPs included in each GRS, and effect size as reported by the reference study

SN	SNP	CA	Chr.	SBP[1]	DBP[1]	MAP[2]	PP[2]	HTN
1	rs17367504	G	1	-0.90	-0.55	-0.53	-	-0.10
2	rs5068	G	1	-	-	-	-	-
3	rs17030613	C	1	-	-	-	-	-
4	rs2932538	G	1	0.39	0.24	0.25	-	0.05
5	rs2004776	T	1	-	-	-	-	-
6	rs1446468	A	2	-	-	-0.34	-	-
7	rs13082711	T	3	-0.32	-0.24	-0.34	-	-0.03
8	rs3774372	T	3	-	-0.37	-	-	-
9	rs319690	A	3	-	-	0.30	-	-
10	rs419076	T	3	0.41	0.24	0.34	-	0.03
11	rs871606	T	4	-	-	-	0.43	-
12	rs1458038	T	4	0.71	0.46	0.40	-	0.07
13	rs13107325	T	4	-0.98	-0.68	-0.63	-	-0.10
14	rs13139571	C	4	0.32	0.26	0.29	-	0.04
15	rs1173771	G	5	0.50	0.26	0.28	0.28	0.06
16	rs11953630	T	5	-0.41	-0.28	-	-	-0.05
17	rs1799945	G	6	0.63	0.46	-	-	0.09
18	rs805303	G	6	0.38	0.23	-	-	0.05
19	rs12705390	A	7	-	-	-	-	-
20	rs3918226	T	7	-	-	-	-	-
21	rs2071518	T	8	-	-	-	0.31	-
22	rs4373814	G	10	-0.37	-0.22	-	-	-0.05
23	rs1813353	A	10	0.57	0.41	-	-	0.08
24	rs4590817	G	10	0.65	0.42	0.58	-	0.10
25	rs1530440	T	10	-	-	-	-	-
26	rs932764	G	10	0.48	0.18	-	-	0.06
27	rs11191548	T	10	1.10	0.46	-	0.53	0.10
28	rs2782980	T	10	-	-	0.34	-	-
29	rs661348	C	11	-	-	-	-	-
30	rs7129220	G	11	-0.62	-0.30	-	-0.38	-0.04
31	rs381815	T	11	0.57	0.35	0.30	0.24	0.06
32	rs633185	G	11	-0.56	-0.33	-0.33	-	-0.07
33	rs11222084	T	11	-	-	-	0.34	-
34	rs17249754	G	12	0.93	0.52	0.56	0.39	0.13
35	rs3184504	T	12	0.60	0.45	-	-	0.06
36	rs653178	T	12	-	-	-0.43	-	-
37	rs10850411	T	12	0.35	0.25	-	-	0.05
38	rs1378942	C	15	0.61	0.42	0.39	-	0.07
39	rs2521501	T	15	0.65	0.36	0.34	-	0.06
40	rs13333226	G	16	-	-	-	-	-0.14
41	rs12940887	T	17	0.36	0.27	0.25	-	0.05
42	rs1327235	G	20	0.34	0.30	0.26	-	0.03
43	rs6015450	G	20	0.90	0.56	0.52	0.35	0.11
Total number of SNPs included in the GRS				27	28	21	9	28

Abbreviation, Ch; chromosome, CA; coded allele. SNP effect size is reported in the reference study, for SBP and DBP from **ICBP1 (ref [107]), and for MAP and PP from ICBP2 (ref. [135]**

Appendix 6 Reported association results for SBP, DBP and MAP from the two ICBP studies.

Nearby genes	SNP	Ch	CA	ICBP1					ICBP 2						
				CAF	SBP		DBP		CAF	SBP		DBP		MAP	
					β (se)	p-value	β (se)	p-value		β (se)	p-value	β (se)	p-value	β (se)	p-value
MTHFR- NPPB	rs17367504	1	G	0.15	-0.9(0.09)	**8.72E-22**	-0.55(0.06)	**3.55E-19**	0.17	-0.83(0.09)	**1.38E-18**	-0.52(0.06)	**5.81E-18**	-0.53(0.07)	**2.18E-16**
MOV10	rs2932538	1	G	0.75	0.39(0.06)	**1.17E-09**	0.24(0.04)	**9.88E-10**	0.83	0.42(0.08)	2.72E-07	0.24(0.05)	5.24E-07	0.25(0.06)	8.25E-06
SLC4A7	rs13082711	3	T	0.78	-0.32(0.07)	1.51E-06	-0.24(0.04)	**3.77E-09**	0.80	-0.38(0.08)	5.28E-06	-0.32(0.05)	**4.37E-09**	-0.34(0.06)	**4.62E-09**
ULK4	rs3774372	3	T	0.83	-0.07(0.08)	0.39	-0.37(0.05)	**9.02E-14**	0.72	0.07(0.09)	4.26E-01	-0.26(0.05)	1.17E-06	-0.14(0.06)	2.77E-02
MECOM	rs419076	3	T	0.47	0.41(0.06)	**1.78E-13**	0.24(0.03)	**2.12E-12**	0.44	0.5(0.07)	**4.09E-13**	0.3(0.04)	**1.11E-11**	0.34(0.05)	**8.11E-13**
FGF5	rs1458038	4	T	0.29	0.71(0.07)	**1.47E-23**	0.46(0.04)	**8.46E-25**	0.3	0.58(0.08)	**9.42E-14**	0.4(0.05)	**1.71E-15**	0.4(0.05)	**2.88E-14**
SLC39A8	rs13107325	4	T	0.05	-0.98(0.13)	**3.27E-14**	-0.68(0.08)	**2.28E-17**	0.12	-0.9(0.14)	**2.16E-10**	-0.6(0.09)	**2.11E-11**	-0.63(0.1)	**1.30E-10**
GUCY1A3- GUCY1B3	rs13139571	4	C	0.76	0.32(0.07)	1.16E-06	0.26(0.04)	**2.17E-10**	0.77	0.44(0.08)	5.60E-08	0.3(0.05)	**2.97E-10**	0.29(0.06)	2.69E-07
NPR3-C5orf23	rs1173771	5	G	0.60	0.5(0.06)	**1.79E-16**	0.26(0.04)	**9.11E-12**	0.53	0.52(0.07)	**1.39E-13**	0.23(0.05)	3.57E-07	0.28(0.05)	**3.51E-09**
EBF1	rs11953630	5	T	0.37	-0.41(0.06)	**3.02E-11**	-0.28(0.04)	**3.81E-13**	0.34	-0.48(0.07)	**1.86E-11**	-0.31(0.05)	**3.44E-11**	-0.33(0.05)	**1.51E-11**
HFE	rs1799945	6	G	0.14	0.63(0.09)	**7.69E-12**	0.46(0.06)	**1.45E-15**	0.18	0.59(0.1)	**2.04E-09**	0.43(0.06)	**8.05E-12**	0.47(0.07)	**6.55E-12**
BAT2-BAT5	rs805303	6	G	0.61	0.38(0.06)	**1.49E-11**	0.23(0.03)	**2.98E-11**	0.58	0.31(0.07)	8.52E-06	0.12(0.04)	2.66E-03	0.15(0.05)	1.45E-03
CACNB2(5')	rs4373814	10	G	0.55	-0.37(0.06)	**4.81E-11**	-0.22(0.03)	**4.36E-10**	0.41	-0.28(0.07)	8.60E-05	-0.16(0.04)	1.83E-04	-0.18(0.05)	3.01E-04
CACNB2(3')	rs1813353	10	T	0.68	0.57(0.08)	**2.56E-12**	0.41(0.05)	**2.30E-15**	0.65	0.5(0.08)	**2.77E-11**	0.35(0.05)	**5.38E-13**	0.4(0.05)	**7.01E-15**
C10orf107	rs4590817	10	G	0.84	0.65(0.09)	**3.97E-12**	0.42(0.06)	**1.29E-12**	0.83	0.79(0.1)	**9.24E-17**	0.51(0.06)	**5.01E-17**	0.58(0.07)	**2.14E-18**
PLCE-1	rs932764	10	G	0.44	0.48(0.06)	**7.10E-16**	0.18(0.04)	8.06E-07	0.43	0.42(0.07)	**1.28E-09**	0.15(0.05)	1.00E-03	0.2(0.05)	2.22E-05
CYP17A1- NT5C3	rs11191548	10	T	0.91	1.10(0.1)	**6.90E-26**	0.46(0.06)	**9.44E-13**	0.94	1.06(0.12)	**1.44E-18**	0.56(0.08)	**6.35E-13**	0.65(0.08)	**2.19E-15**
ADM	rs7129220	11	G	0.89	-0.62(0.09)	**2.97E-12**	-0.3(0.06)	6.44E-08	0.87	-0.68(0.11)	**1.21E-09**	-0.26(0.07)	3.13E-04	-0.33(0.08)	2.07E-05
PLEKHA7	rs381815	11	T	0.26	0.57(0.09)	**5.27E-11**	0.35(0.06)	**5.34E-10**	0.30	0.57(0.08)	**2.16E-13**	0.34(0.05)	**9.98E-12**	0.3(0.05)	**2.75E-08**
FLJ32810- TMEM34	rs633185	11	G	0.28	-0.56(0.07)	**1.21E-17**	-0.33(0.04)	**1.95E-15**	0.32	-0.5(0.08)	**6.89E-11**	-0.29(0.05)	**4.09E-09**	-0.33(0.05)	**6.58E-10**
ATP2B1	rs17249754	12	G	0.84	0.93(0.11)	**1.82E-18**	0.52(0.07)	**1.16E-14**	0.89	0.91(0.09)	**6.02E-22**	0.46(0.06)	**3.08E-14**	0.56(0.07)	**1.21E-17**
ATXN2	rs3184504	12	T	0.47	0.6(0.07)	**3.83E-18**	0.45(0.04)	**3.59E-25**	0.41	0.56(0.07)	**9.92E-16**	0.44(0.04)	**7.22E-23**	0.41(0.05)	**3.64E-18**
TBX5-TBX3	rs10850411	12	T	0.70	0.35(0.07)	5.38E-08	0.25(0.04)	**5.43E-10**	0.72	0.38(0.08)	8.47E-07	0.25(0.05)	**2.56E-08**	0.25(0.05)	4.44E-06
CSK	rs1378942	15	C	0.35	0.61(0.06)	**5.69E-23**	0.42(0.04)	**2.69E-26**	0.33	0.59(0.07)	**7.96E-17**	0.42(0.05)	**4.38E-20**	0.39(0.05)	**1.63E-15**
FES	rs2521501	15	G	0.31	0.65(0.07)	**5.20E-19**	0.36(0.05)	**1.89E-15**	0.63	0.58(0.09)	**1.16E-10**	0.37(0.05)	**1.01E-12**	0.34(0.06)	**2.88E-08**
GOSR2	rs17608766	17	T	0.86	-0.56(0.09)	**1.13E-10**	-0.13(0.05)	0.02	0.91	-0.73(0.1)	**4.68E-13**	-0.2(0.07)	2.28E-03	-0.36(0.07)	1.94E-07
ZNF652	rs12940887	17	T	0.38	0.36(0.06)	**1.79E-10**	0.27(0.04)	**2.29E-14**	0.42	0.28(0.07)	9.03E-05	0.26(0.05)	**1.36E-08**	0.25(0.05)	2.47E-07
JAG1	rs1327235	20	G	0.46	0.34(0.06)	**1.87E-08**	0.3(0.04)	**1.41E-15**	0.12	0.35(0.07)	5.11E-07	0.26(0.04)	**1.44E-10**	0.26(0.05)	**4.35E-08**
GNAS- EDN3	rs6015450	20	G	0.12	0.9(0.09)	**3.87E-23**	0.56(0.06)	**5.63E-23**	0.07	0.85(0.11)	**5.94E-15**	0.51(0.07)	**4.39E-13**	0.52(0.07)	**1.58E-12**

Statisically significant SNPs are in bold. Abbreviation, Ch; chromosome, CA; coded allele, CAF: coded allele frequency. ICBP1 (ref [107]), ICBP2 (ref [135]).
Cells are highlighted in colour as following: green for SNPs associated with both SBP and DBP, blue for SNP associated with SBP only, orange for SNP associated with DBP only, red for association that was not replicated in the second ICBP.

References

1. Lim, S.S., Vos, T., Flaxman, A.D., Danaei, G., Shibuya, K., Adair-Rohani, H., AlMazroa, M.A., Amann, M., Anderson, H.R., Andrews, K.G., et al. (2012). A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. The Lancet *380*, 2224-2260.

2. Lawes, C.M.M., Vander Hoorn, S., Rodgers, A., International Society of Hypertension (2008). Global burden of blood-pressure-related disease, 2001. Lancet *371*, 1513–1518.

3. Kearney, P.M., Whelton, M., Reynolds, K., Muntner, P., Whelton, P.K., and He, J. (2005). Global burden of hypertension: analysis of worldwide data. Lancet *365*, 217–223.

4. World Health Organization (WHO) (2013). A Global Brief on Hypertension: Silent Killer, Global Public Health Crisis. Available from: http://www.thehealthwell.info/node/466541 [Accessed: February 2016].

5. Lewington, S., Clarke, R., Qizilbash, N., Peto, R., Collins, R., Prospective Studies Collaboration (2002). Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. The Lancet *360*, 1903–1913.

6. Vasan, R.S., Larson, M.G., Leip, E.P., Evans, J.C., O’Donnell, C.J., Kannel, W.B., and Levy, D. (2001). Impact of high-normal blood pressure on the risk of cardiovascular disease. N. Engl. J. Med. *345*, 1291–1297.

7. Lifton, R.P., Gharavi, A.G., and Geller, D.S. (2001). Molecular Mechanisms of Human Hypertension. Cell *104*, 545–556.

8. Oparil, S., Zaman, M.A., and Calhoun, D.A. (2003). Pathogenesis of hypertension. Ann. Intern. Med. *139*, 761–776.

9. Beevers, G., Lip, G.Y., and O'Brien, E. (2001). ABC of hypertension: The pathophysiology of hypertension. BMJ *322*, 912–916.

10. Atlas, S.A. (2007). The renin-angiotensin aldosterone system: pathophysiological role and pharmacologic inhibition. J Manag Care Pharm *13*, 9-20.

11. Riet, Te, L., van Esch, J.H.M., Roks, A.J.M., van den Meiracker, A.H., and Danser, A.H.J. (2015). Hypertension: renin-angiotensin-aldosterone system alterations. Circ. Res. *116*, 960–975.

12. Mu, S., Shimosawa, T., Ogura, S., Wang, H., Uetake, Y., Kawakami-Mori, F., Marumo, T., Yatomi, Y., Geller, D.S., Tanaka, H., et al. (2011). Epigenetic modulation of the renal [beta]-adrenergic-WNK4 pathway in salt-sensitive hypertension. Nat Med *17*, 573–580.

13. Coffman, T.M. (2011). Under pressure: the search for the essential

mechanisms of hypertension. Nat Med *17*, 1402–1409.

14. Dharmashankar, K., and Widlansky, M.E. (2010). Vascular Endothelial Function and Hypertension: Insights and Directions. Curr. Hypertens. Rep. *12*, 448–455.

15. Sesso, H.D., Stampfer, M.J., Rosner, B., Hennekens, C.H., Gaziano, J.M., Manson, J.E., and Glynn, R.J. (2000). Systolic and diastolic blood pressure, pulse pressure, and mean arterial pressure as predictors of cardiovascular disease risk in Men. Hypertension *36*, 801–807.

16. Strandberg, T.E., and Pitkala, K. (2003). What is the most important component of blood pressure: systolic, diastolic or pulse pressure? Curr. Opin. Nephrol. Hypertens. *12*, 293–297.

17. García-Donaire, J.A., and Ruilope, L.M. (2010). Systolic Pressure, Diastolic Pressure, or Pulse Pressure as a Cardiovascular Risk Factor in Renal Disease. Curr. Hypertens. Rep. *12*, 307–312.

18. MACMAHON, S. (1990). Blood pressure, stroke, and coronary heart disease *1Part 1, prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. The Lancet *335*, 765–774.

19. Kannel, W.B., Gordon, T., and Schwartz, M.J. (1971). Systolic versus diastolic blood pressure and risk of coronary heart disease. Am. J. Cardio. *27*, 335–346.

20. Franklin, S.S., Larson, M.G., Khan, S.A., Wong, N.D., Leip, E.P., Kannel, W.B., and Levy, D. (2001). Does the relation of blood pressure to coronary heart disease risk change with aging? The Framingham Heart Study. Circulation *103*, 1245–1249.

21. Domanski, M., Mitchell, G., Pfeffer, M., Neaton, J.D., Norman, J., Svendsen, K., Grimm, R., Cohen, J., Stamler, J., MRFIT Research Group (2002). Pulse pressure and cardiovascular disease-related mortality: follow-up study of the Multiple Risk Factor Intervention Trial (MRFIT). JAMA *287*, 2677–2683.

22. Franklin, S.S., Lopez, V.A., Wong, N.D., Mitchell, G.F., Larson, M.G., Vasan, R.S., and Levy, D. (2009). Single versus combined blood pressure components and risk for cardiovascular disease: the Framingham Heart Study. Circulation *119*, 243–250.

23. Mancia, G., Fagard, R., Zanchetti, A., Böhm, M., Christiaens, T., Cifkova, R., De Backer, G., Dominiczak, A., Galderisi, M., Grobbee, D.E., et al. (2013). 2013 ESH/ESC guidelines for the management of arterial hypertension: the Task Force for the Management of Arterial Hypertension of the European Society of Hypertension (ESH) and of the European Society of Cardiology (ESC). Eur. Heart J. *34*, 2159–2219.

24. Bliziotis, I.A., Destounis, A., and Stergiou, G.S. (2012). Home versus ambulatory and office blood pressure in predicting target organ damage in hypertension. J Hypertens *30*, 1289–1299.

25. Stergiou, G.S., and Bliziotis, I.A. (2011). Home blood pressure monitoring in

the diagnosis and treatment of hypertension: a systematic review. Am. J. Hypertens. *24*, 123–134.

26. Beevers, G., Lip, G.Y.H., and O'Brien, E. (2001). Blood pressure measurement. BMJ *322*, 981–985.

27. Sebo, P., Pechère-Bertschi, A., Herrmann, F.R., Haller, D.M., and Bovier, P. (2014). Blood pressure measurements are unreliable to diagnose hypertension in primary care. J Hypertens *32*, 509–517.

28. Jones, D.W., Appel, L.J., Sheps, S.G., Roccella, E.J., and Lenfant, C. (2003). Measuring blood pressure accurately: new and persistent challenges. JAMA *289*, 1027–1030.

29. Handler, J. (2009). The importance of accurate blood pressure measurement. Perm J *13*, 51–54.

30. PLATT, R. (1959). The nature of essential hypertension. The Lancet 2, 55–57.

31. Zanchetti A. (1986). Platt versus Pickering: an episode in recent medical history. By J. D. Swales, editor. An essay review. Medical History 30, 94-96.

32. Oldham, P.D., Pickering, G., Fraser Roberts, J.A., and Sowry, G.S.C. (1960). THE NATURE OF ESSENTIAL HYPERTENSION. The Lancet *275*, 1085–1093.

33. National Clinical Guideline Centre The clinical management of primary hypertension in adults. Avilable online: https://www.nice.org.uk/guidance/cg127

34. Krause, T., Lovibond, K., Caulfield, M., McCormack, T., and Williams, B. (2011). Management of hypertension: summary of NICE guidance. BMJ *343*, d4891–d4891.

35. National Institute of Health (2003). JNC 7 express: The seventh report of the Joint National Committee on prevention, detection, evaluation, and treatment of high blood pressure (Maryland: National Institutes of Health).

36. James, P.A., Oparil, S., Carter, B.L., Cushman, W.C., Dennison-Himmelfarb, C., Handler, J., Lackland, D.T., LeFevre, M.L., MacKenzie, T.D., Ogedegbe, O., et al. (2014). 2014 evidence-based guideline for the management of high blood pressure in adults: report from the panel members appointed to the Eighth Joint National Committee (JNC 8). JAMA *311*, 507–520.

37. Padmanabhan, S., Newton-Cheh, C., and Dominiczak, A.F. (2012). Genetic basis of blood pressure and hypertension. Trends in Genetics *28*, 397–408.

38. Padmanabhan, S., Caulfield, M., and Dominiczak, A.F. (2015). Genetic and Molecular Aspects of Hypertension. Circ. Res. *116*, 937–959.

39. Coffman, T.M., and Crowley, S.D. (2014). The inextricable role of the kidney in hypertension. J. Clin. Invest. *124*, 2341–2347.

40. Munroe, P.B., Barnes, M.R., and Caulfield, M.J. (2013). Advances in blood pressure genomics. Circ. Res. *112*, 1365–1379.

41. Wilson, P.W.F., D'Agostino, R.B., Sullivan, L., Parise, H., and Kannel, W.B. (2002). Overweight and obesity as determinants of cardiovascular risk: the Framingham experience. Arch. Intern. Med. *162*, 1867–1872.

42. Wang, N.-Y., Young, J.H., Meoni, L.A., Ford, D.E., Erlinger, T.P., and Klag, M.J. (2008). Blood Pressure Change and Risk of Hypertension Associated With Parental Hypertension: The Johns Hopkins Precursors Study. Arch. Intern. Med. *168*, 643–648.

43. Carson, A.P., Howard, G., Burke, G.L., Shea, S., Levitan, E.B., and Muntner, P. (2011). Ethnic Differences in Hypertension Incidence Among Middle-Aged and Older Adults The Multi-Ethnic Study of Atherosclerosis. Hypertension *57*, 1101–1107.

44. Adrogué, H.J., and Madias, N.E. (2007). Sodium and Potassium in the Pathogenesis of Hypertension. N. Engl. J. Med. *356*, 1966–1978.

45. Carretero, O.A., and Oparil, S. (2000). Essential Hypertension : Part I: Definition and Etiology. Circulation *101*, 329–335.

46. Cheng, S., Xanthakis, V., Sullivan, L.M., and Vasan, R.S. (2012). Blood pressure tracking over the adult life course: patterns and correlates in the Framingham heart study. Hypertension *60*, 1393–1399.

47. Falaschetti, E., Gillespie, C., Robitaille, C., McAlister, F.A., and Johansen, H. (2013). Hypertension prevalence, awareness, treatment and control in national surveys from England, the USA and Canada, and correlation with stroke and ischaemic heart disease mortality: a cross-sectional study. BMJ Open *3*, e003423–e003423.

48. Berenson, G.S., Chen, W., DasMahapatra, P., Fernandez, C., Giles, T., Xu, J., and Srinivasan, S.R. (2011). Stimulus response of blood pressure in black and white young individuals helps explain racial divergence in adult cardiovascular disease: The Bogalusa Heart Study. J Am Soc Hypertens *5*, 230–238.

49. Wang, X., Poole, J.C., Treiber, F.A., Harshfield, G.A., Hanevold, C.D., and Snieder, H. (2006). Ethnic and gender differences in ambulatory blood pressure trajectories: results from a 15-year longitudinal study in youth and young adults. Circulation *114*, 2780–2787.

50. Jones, D.W., and Hall, J.E. (2006). Racial and ethnic differences in blood pressure: biology and sociology. Circulation *114*, 2757–2759.

51. Lane, D., Beevers, D.G., and Lip, G.Y.H. (2002). Ethnic differences in blood pressure and the prevalence of hypertension in England. J Hum Hypertens *16*, 267–273.

52. DeMarco, V.G., Aroor, A.R., and Sowers, J.R. (2014). The pathophysiology of hypertension in patients with obesity. Nat Rev Endocrinol *10*, 364–376.

53. Schmieder, R.E., and Messerli, F.H. (1993). Does obesity influence early target organ damage in hypertensive patients? Circulation *87*, 1482–1488.

54. Jordan, J., Yumuk, V., Schlaich, M., Nilsson, P.M., Zahorska-Markiewicz, B.,

Grassi, G., Schmieder, R.E., Engeli, S., and Finer, N. (2012). Joint statement of the European Association for the Study of Obesity and the European Society of Hypertension: obesity and difficult to treat arterial hypertension. J Hypertens *30*, 1047–1055.

55. Sacks, F.M., Svetkey, L.P., Vollmer, W.M., Appel, L.J., Bray, G.A., Harsha, D., Obarzanek, E., Conlin, P.R., Miller, E.R., Simons-Morton, D.G., et al. (2001). Effects on Blood Pressure of Reduced Dietary Sodium and the Dietary Approaches to Stop Hypertension (DASH) Diet. N. Engl. J. Med. *344*, 3–10.

56. He, F.J., and MacGregor, G.A. (2003). How Far Should Salt Intake Be Reduced? Hypertension *42*, 1093–1099.

57. Xin, X., He, J., Frontini, M.G., Ogden, L.G., Motsamai, O.I., and Whelton, P.K. (2001). Effects of Alcohol Reduction on Blood Pressure A Meta-Analysis of Randomized Controlled Trials. Hypertension *38*, 1112–1117.

58. Lawlor, D.A., Nordestgaard, B.G., Benn, M., Zuccolo, L., Tybjaerg-Hansen, A., and Smith, G.D. (2013). Exploring causal associations between alcohol and coronary heart disease risk factors: findings from a Mendelian randomization study in the Copenhagen General Population Study. Eur. Heart J. *34*, 2519–2528.

59. Chen, L., Davey Smith, G., Harbord, R.M., and Lewis, S.J. (2008). Alcohol Intake and Blood Pressure: A Systematic Review Implementing a Mendelian Randomization Approach. PloS Med. *5*, e52.

60. Balijepalli, C., sch, C.L.O., Bramlage, P., Erbel, R., Humphries, K.H., ckel, K.-H.J.O., and Moebus, S. (2013). Percentile distribution of blood pressure readings in 35683 men and women aged 18 to 99 years. J Hum Hypertens *28*, 193–200.

61. Danaei, G., Finucane, M.M., Lin, J.K., Singh, G.M., Paciorek, C.J., Cowan, M.J., Farzadfar, F., Stevens, G.A., Lim, S.S., Riley, L.M., et al. (2011). National, regional, and global trends in systolic blood pressure since 1980: systematic analysis of health examination surveys and epidemiological studies with 786 country-years and 5·4 million participants. Lancet *377*, 568–577.

62. Pereira, M., Lunet, N., Azevedo, A., and Barros, H. (2009). Differences in prevalence, awareness, treatment and control of hypertension between developing and developed countries. J Hypertens *27*, 963–975.

63. Wolf-Maier, K., Cooper, R.S., Kramer, H., Banegas, J.R., Giampaoli, S., Joffres, M.R., Poulter, N., Primatesta, P., Stegmayr, B., and Thamm, M. (2004). Hypertension Treatment and Control in Five European Countries, Canada, and the United States. Hypertension *43*, 10–17.

64. Hertz, R.P., Unger, A.N., Cornell, J.A., and Saunders, E. (2005). Racial Disparities in Hypertension Prevalence, Awareness, and Management. Arch. Intern. Med. *165*, 2098–2104.

65. Chow, C.K., Teo, K.K., Rangarajan, S., Islam, S., Gupta, R., Avezum, A., Bahonar, A., Chifamba, J., Dagenais, G., Diaz, R., et al. (2013). Prevalence, Awareness, Treatment, and Control of Hypertension in Rural and Urban Communities in High-, Middle-, and Low-Income Countries. JAMA *310*, 959–968.

66. Rutherford, L., Hinchliffe, S., Sharp, C., Bromley, C., Dowling, S., Gray, L., Hughes, T., Leyland, A.H., McNeill, G., and Marcinkiewicz, A. The Scottish Health Survey 2013: Volume 1: Main Report. Avilable online: http://www.gov.scot/Publications/2014/12/9982

67. Lloyd-Sherlock, P., Beard, J., Minicuci, N., Ebrahim, S., and Chatterji, S. (2014). Hypertension among older adults in low- and middle-income countries: prevalence, awareness and control. Int J Epidemiol *43*, 116–128.

68. Falaschetti, E., Mindell, J., Knott, C., and Poulter, N. (2014). Hypertension management in England: a serial cross-sectional study from 1994 to 2011. The Lancet *383*, 1912–1919.

69. Dickinson, H.O., Mason, J.M., Nicolson, D.J., Campbell, F., Beyer, F.R., Cook, J.V., Williams, B., and Ford, G.A. (2006). Lifestyle interventions to reduce raised blood pressure: a systematic review of randomized controlled trials. J Hypertens *24*, 215–233.

70. Law, M.R., Morris, J.K., and Wald, N.J. (2009). Use of blood pressure lowering drugs in the prevention of cardiovascular disease: meta-analysis of 147 randomised trials in the context of expectations from prospective epidemiological studies. BMJ *338*, b1665–b1665.

71. Gaziano, T.A., Bitton, A., Anand, S., Weinstein, M.C., and Hypertension, F.T.I.S.O. (2009). The global cost of nonoptimal blood pressure. J Hypertens *27*, 1472–1477.

72. Di Cesare, M., Bennett, J.E., Best, N., Stevens, G.A., Danaei, G., and Ezzati, M. (2013). The contributions of risk factor trends to cardiometabolic mortality decline in 26 industrialized countries. Int J Epidemiol *42*, 838–848.

73. Whelton, P.K., He, J., Appel, L.J., Cutler, J.A., Havas, S., Kotchen, T.A., Roccella, E.J., Stout, R., Vallbona, C., Winston, M.C., et al. (2002). Primary Prevention of Hypertension: Clinical and Public Health Advisory From the National High Blood Pressure Education Program. JAMA *288*, 1882–1888.

74. Burton, P.R., Tobin, M.D., and Hopper, J.L. (2005). Key concepts in genetic epidemiology. Lancet *366*, 941–951.

75. Alghamdi, J., and Padmanabhan, S. (2014). Fundamentals of Complex Trait Genetics and Association Studies. In Handbook of Pharmacogenomics and Stratified Medicine, (Elsevier), pp. 235–257.

76. Lander, E.S. (2011). Initial impact of the sequencing of the human genome. Nature *470*, 187–197.

77. Karlsson, E.K., Kwiatkowski, D.P., and Sabeti, P.C. (2014). Natural selection and infectious disease in human populations. Nat. Rev. Genet. *15*, 379–393.

78. Frazer, K.A., Murray, S.S., Schork, N.J., and Topol, E.J. (2009). Human genetic variation and its contribution to complex traits. Nat. Rev. Genet. *10*, 241–251.

79. Burton, P.R., Bowden, J.M., and Tobin, M.D. (2008). Epidemiology and

Genetic Epidemiology. In Handbook of Statistical Genetics, (Chichester: John Wiley & Sons, Ltd), pp. 1109–1140.

80. Plomin, R., Haworth, C.M.A., and Davis, O.S.P. (2011). Common disorders are quantitative traits. Nat. Rev. Genet. *10*, 872–878.

81. Almasy, L. (2012). The role of phenotype in gene discovery in the whole genome sequencing era. Hum. Genet. *131*, 1533–1540.

82. Padmanabhan, S., Melander, O., Johnson, T., Di Blasio, A.M., Lee, W.K., Gentilini, D., Hastie, C.E., Menni, C., Monti, M.C., Delles, C., et al. (2010). Genome-wide association study of blood pressure extremes identifies variant near UMOD associated with hypertension. PLoS Genet. *6*, e1001177.

83. Green, A.E., Munafò, M.R., DeYoung, C.G., Fossella, J.A., Fan, J., and Gray, J.R. (2008). Using genetic data in cognitive neuroscience: from growing pains to genuine insights. Nat. Rev. Neurosci. *9*, 710–720.

84. Zheng, G., and Tian, X. (2005). The impact of diagnostic error on testing genetic association in case-control studies. Statist. Med. *24*, 869–882.

85. Liao, J., Li, X., Wong, T.Y., Wang, J.J., Khor, C.-C., Tai, E.S., Aung, T., Teo, Y.Y., and Cheng, C.-Y. (2014). Impact of measurement error on testing genetic association with quantitative traits. *9*, e87044.

86. Wojczynski, M.K., and Tiwari, H.K. (2008). Definition of Phenotype. In Genetic Dissection of Complex Traits, (Academic Press), pp. 75–105.

87. Tobin, M.D., Sheehan, N.A., Scurrah, K.J., and Burton, P.R. (2005). Adjusting for treatment effects in studies of quantitative traits: antihypertensive therapy and systolic blood pressure. Statist. Med. *24*, 2911–2935.

88. Rana, B.K., Dhamija, A., Panizzon, M.S., Spoon, K.M., Vasilopoulos, T., Franz, C.E., Grant, M.D., Jacobson, K.C., Kim, K., Lyons, M.J., et al. (2014). Imputing Observed Blood Pressure for Antihypertensive Treatment: Impact on Population and Genetic Analyses. Am. J. Hypertens. *27*, hpt271–hpt837.

89. Ott, J., Kamatani, Y., and Lathrop, M. (2011). Family-based designs for genome-wide association studies. Nat. Rev. Genet. *12*, 465–474.

90. Laird, N.M., and Lange, C. (2006). Family-based designs in the age of large-scale gene-association studies. Nat. Rev. Genet. *7*, 385–394.

91. Clerget-Darpoux, F., and Elston, R.C. (2007). Are linkage analysis and the collection of family data dead? Prospects for family studies in the age of genome-wide association. Hum Hered *64*, 91–96.

92. Lange, E.M., Sun, J., Lange, L.A., Zheng, S.L., Duggan, D., Carpten, J.D., Grönberg, H., Isaacs, W.B., Xu, J., and Chang, B.-L. (2008). Family-based samples can play an important role in genetic association studies. Cancer Epidemiol. Biomarkers Prev. *17*, 2208–2214.

93. Price, A.L., Zaitlen, N.A., Reich, D., and Patterson, N. (2010). New

approaches to population stratification in genome-wide association studies. Nat. Rev. Genet. *11*, 459–463.

94. Kong, A., Steinthorsdottir, V., Masson, G., Thorleifsson, G., Sulem, P., Besenbacher, S., Jonasdottir, A., Sigurdsson, A., Kristinsson, K.T., Jonasdottir, A., et al. (2009). Parental origin of sequence variants associated with complex diseases. Nature *462*, 868–874.

95. Mott, R., Yuan, W., Kaisaki, P., Gan, X., Cleak, J., and Edwards, A. (2014). The architecture of parent-of-origin effects in mice. Cell *156*, 332–342.

96. Veltman, J.A., and Brunner, H.G. (2012). De novo mutations in human genetic disease. Nat. Rev. Genet. *13*, 565–575.

97. Benyamin, B., Visscher, P.M., and McRae, A.F. (2009). Family-based genome-wide association studies. Pharmacogenomics *10*, 181–190.

98. Gordon, D., Heath, S.C., and Ott, J. (1999). True pedigree errors more frequent than apparent errors for single nucleotide polymorphisms. Hum Hered *49*, 65–70.

99. Chen, W.-M., and Abecasis, G.R. (2007). Family-based association tests for genomewide association scans. Am. J. Hum. Genet. *81*, 913–926.

100. Visscher, P.M., Andrew, T., and Nyholt, D.R. (2008). Genome-wide association studies of quantitative traits with related individuals: little (power) lost but much to be gained. Eur. J. Hum. Genet. *16*, 387–390.

101. Sahana, G., Guldbrandtsen, B., Janss, L., and Lund, M.S. (2010). Comparison of association mapping methods in a complex pedigreed population. Genet. Epidemiol. *34*, 455–462.

102. Khoury, M.J., Beaty, T.H., and KUNG-YEE, L. (1988). CAN FAMILIAL AGGREGATION OF DISEASE BE EXPLAINED BY FAMILIAL AGGREGATION OF ENVIRONMENTAL RISK FACTORS? Am. J. Epidemiol. *127*, 674–683.

103. Risch, N. (1990). Linkage strategies for genetically complex traits. II. The power of affected relative pairs. Am. J. Hum. Genet. *46*, 229–241.

104. Matthews, A.G., Finkelstein, D.M., and Betensky, R.A. (2008). Analysis of familial aggregation studies with complex ascertainment schemes. Statist. Med. *27*, 5076–5092.

105. Tenesa, A., and Haley, C.S. (2013). The heritability of human disease: estimation, uses and abuses. Nat. Rev. Genet. *14*, 139–149.

106. Sham, P.C., and Cherny, S.S. (2011). Genetic Architecture of Complex Diseases. In Analysis of Complex Disease Association Studies, (Elsevier), pp. 1–13.

107. Ehret, G.B., Munroe, P.B., Rice, K.M., Bochud, M., Johnson, A.D., Chasman, D.I., Smith, A.V., Tobin, M.D., Verwoert, G.C., Hwang, S.-J., et al. (2011). Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. Nature *478*, 103–109.

108. Ehret, G.B., and Caulfield, M.J. (2013). Genes for blood pressure: an opportunity to understand hypertension. Eur. Heart J. *0*, 4551–455.

109. Kotchen, T.A., Kotchen, J.M., Grim, C.E., George, V., Kaldunski, M.L., Cowley, A.W., Hamet, P., and Chelius, T.H. (2000). Genetic Determinants of Hypertension Identification of Candidate Phenotypes. Hypertension *36*, 7–13.

110. HAVLIK, R.J., GARRISON, R.J., FEINLEIB, M., Kannel, W.B., CASTELLI, W.P., and MCNAMARA, P.M. (1979). BLOOD PRESSURE AGGREGATION IN FAMILIES. Am. J. Epidemiol. *110*, 304–312.

111. Kupper, N., Willemsen, G., RIESE, H., Posthuma, D., Boomsma, D.I., and De Geus, E.J.C. (2005). Heritability of daytime ambulatory blood pressure in an extended twin design. Hypertension *45*, 80–85.

112. Usher, C.L., and McCarroll, S.A. (2015). Complex and multi-allelic copy number variation in human disease. Briefings in Functional Genomics *14*, elv028–elv338.

113. Zhang, D., Qian, Y., Akula, N., Alliey-Rodriguez, N., Tang, J., Bipolar Genome Study, Gershon, E.S., and Liu, C. (2011). Accuracy of CNV Detection from GWAS Data. PLoS ONE 6, e14511.

114. Solovieff, N., Cotsapas, C., Lee, P.H., Purcell, S.M., and Smoller, J.W. (2013). Pleiotropy in complex traits: challenges and strategies. Nat. Rev. Genet. *14*, 483–495.

115. Levy, D., Ehret, G.B., Rice, K., Verwoert, G.C., Launer, L.J., Dehghan, A., Glazer, N.L., Morrison, A.C., Johnson, A.D., Aspelund, T., et al. (2009). Genome-wide association study of blood pressure and hypertension. Nat. Genet. *41*, 677–687.

116. Hindorff, L.A., Sethupathy, P., Junkins, H.A., Ramos, E.M., Mehta, J.P., Collins, F.S., and Manolio, T.A. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc. Natl. Acad. Sci. U.S.A. *106*, 9362–9367.

117. ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.

118. Schork, A.J., Thompson, W.K., Pham, P., Torkamani, A., Roddey, J.C., Sullivan, P.F., Kelsoe, J.R., O'donovan, M.C., Furberg, H., Tobacco, T., et al. (2013). All SNPs Are Not Created Equal: Genome-Wide Association Studies Reveal a Consistent Pattern of Enrichment among Functionally Annotated SNPs. PLoS Genet. *9*, e1003449.

119. Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., et al. (2012). Systematic localization of common disease-associated variation in regulatory DNA. Science *337*, 1190–1195.

120. Anderson, C.A., Pettersson, F.H., Clarke, G.M., Cardon, L.R., Morris, A.P., and Zondervan, K.T. (2010). Data quality control in genetic case-control association studies. Nat Protoc *5*, 1564–1573.

121. Weale, M.E. (2010). Quality Control for Genome-Wide Association Studies. In Genetic Variation: Methods and Protocols, R.M. Barnes, and G. Breen, eds. (Totowa, NJ: Humana Press), pp. 341–372.

122. Turner, S., Armstrong, L.L., Bradford, Y., Carlson, C.S., Crawford, D.C., Crenshaw, A.T., de Andrade, M., Doheny, K.F., Haines, J.L., Hayes, G., et al. (2001). Quality Control Procedures for Genome-Wide Association Studies. In Current Protocols in Human Genetics, (Hoboken, NJ, USA: John Wiley & Sons, Inc.), pp. 1.19.1–1.19.18.

123. Laurie, C.C., Doheny, K.F., Mirel, D.B., Pugh, E.W., Bierut, L.J., Bhangale, T., Boehm, F., Caporaso, N.E., Cornelis, M.C., Edenberg, H.J., et al. (2010). Quality control and quality assurance in genotypic data for genome-wide association studies. Genet. Epidemiol. *34*, 591–602.

124. Sham, P.C., and Purcell, S.M. (2014). Statistical power and significance testing in large-scale genetic studies. Nat. Rev. Genet. *15*, 335–346.

125. Yu, J., Pressoir, G., Briggs, W.H., Vroh Bi, I., Yamasaki, M., Doebley, J.F., McMullen, M.D., Gaut, B.S., Nielsen, D.M., Holland, J.B., et al. (2005). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. *38*, 203–208.

126. Aulchenko, Y.S., de Koning, D.J., and Haley, C. (2007). Genomewide Rapid Association Using Mixed Model and Regression: A Fast and Simple Method For Genomewide Pedigree-Based Quantitative Trait Loci Association Analysis. Genetics *177*, 577–585.

127. Yang, J., Zaitlen, N.A., Goddard, M.E., Visscher, P.M., and Price, A.L. (2014). Advantages and pitfalls in the application of mixed-model association methods. Nat. Genet. *46*, 100–106.

128. Luft, F.C. (2001). Twins in Cardiovascular Genetic Research. Hypertension *37*, 350–356.

129. Tobin, M.D., Tomaszewski, M., Braund, P.S., Hajat, C., Raleigh, S.M., Palmer, T.M., Caulfield, M., Burton, P.R., and Samani, N.J. (2008). Common variants in genes underlying monogenic hypertension and hypotension and blood pressure in the general population. Hypertension *51*, 1658–1664.

130. Baysal, B.E. (2008). Clinical and molecular progress in hereditary paraganglioma. J. Med. Genet. *45*, 689–694.

131. Burton, P.R., Clayton, D.G., Cardon, L.R., Duncanson, A., Kwiatkowski, D.P., Mccarthy, M.I., Barrett, J.C., Donnelly, P., Marchini, J.L., Morris, A.P., et al. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature *447*, 661–678.

132. Levy, D., Larson, M.G., Benjamin, E.J., Newton-Cheh, C., Wang, T.J., Hwang, S.-J., Vasan, R.S., and Mitchell, G.F. (2007). Framingham Heart Study 100K Project: genome-wide associations for blood pressure and arterial stiffness. BMC Med. Genet. *8*, 1–11.

133. Newton-Cheh, C., Johnson, T., Gateva, V., Tobin, M.D., Bochud, M., Coin,

L., Najjar, S.S., Zhao, J.H., Heath, S.C., Eyheramendy, S., et al. (2009). Genome-wide association study identifies eight loci associated with blood pressure. Nat. Genet. *41*, 666–676.

134. Newton-Cheh, C., Larson, M.G., Vasan, R.S., Levy, D., Bloch, K.D., Surti, A., Guiducci, C., Kathiresan, S., Benjamin, E.J., Struck, J., et al. (2009). Association of common variants in NPPA and NPPB with circulating natriuretic peptides and blood pressure. Nat. Genet. *41*, 348–353.

135. Wain, L.V., Verwoert, G.C., O'reilly, P.F., Shi, G., Johnson, T., Johnson, A.D., Bochud, M., Rice, K.M., Henneman, P., Smith, A.V., et al. (2011). Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. Nat. Genet. *43*, 1005–1011.

136. Salvi, E., Kutalik, Z., Glorioso, N., Benaglio, P., Frau, F., Kuznetsova, T., Arima, H., Hoggart, C., Tichet, J., Nikitin, Y.P., et al. (2012). Genomewide association study using a high-density single nucleotide polymorphism array and case-control design identifies a novel essential hypertension susceptibility locus in the promoter region of endothelial NO synthase. Hypertension *59*, 248–255.

137. Kato, N., Takeuchi, F., Tabara, Y., Kelly, T.N., Go, M.J., Sim, X., Tay, W.T., Chen, C.-H., Zhang, Y., Yamamoto, K., et al. (2011). Meta-analysis of genome-wide association studies identifies common variants associated with blood pressure variation in east Asians. Nat. Genet. *43*, 531–538.

138. Franceschini, N., Fox, E., Zhang, Z., Edwards, T.L., Nalls, M.A., Sung, Y.J., Tayo, B.O., Sun, Y.V., Gottesman, O., Adeyemo, A., et al. (2013). Genome-wide association analysis of blood-pressure traits in African-ancestry individuals reveals common associated genes in African and non-African populations. Am. J. Hum. Genet. *93*, 545–554.

139. Lu, X., Wang, L., Lin, X., Huang, J., Charles Gu, C., He, M., Shen, H., He, J., Zhu, J., Li, H., et al. (2015). Genome-wide association study in Chinese identifies novel loci for blood pressure and hypertension. Hum. Mol. Genet. *24*, 865–874.

140. Ganesh, S.K., Chasman, D.I., Larson, M.G., Guo, X., Verwoert, G., Bis, J.C., Gu, X., Smith, A.V., Yang, M.-L., Zhang, Y., et al. (2014). Effects of Long-Term Averaging of Quantitative Blood Pressure Traits on the Detection of Genetic Associations. Am. J. Hum. Genet. *95*, 49–65.

141. Yadav, S., Cotlarciuc, I., Munroe, P.B., Khan, M.S., Nalls, M.A., Bevan, S., Cheng, Y.-C., Chen, W.-M., Malik, R., McCarthy, N.S., et al. (2013). Genome-wide analysis of blood pressure variability and ischemic stroke. Stroke *44*, 2703–2709.

142. Tomaszewski, M., Debiec, R., Braund, P.S., Nelson, C.P., Hardwick, R., Christofidou, P., Denniff, M., Codd, V., Rafelt, S., Van Der Harst, P., et al. (2010). Genetic Architecture of Ambulatory Blood Pressure in the General Population: Insights From Cardiovascular Gene-Centric Array. Hypertension *56*, 1069–1076.

143. Simino, J., Shi, G., Bis, J.C., Chasman, D.I., Ehret, G.B., Gu, X., Guo, X., Hwang, S.-J., Sijbrands, E., Smith, A.V., et al. (2014). Gene-age interactions in

blood pressure regulation: a large-scale investigation with the CHARGE, Global BPgen, and ICBP Consortia. Am. J. Hum. Genet. *95*, 24–38.

144. Simino, J., Sung, Y.J., Kume, R., Schwander, K., and Rao, D.C. (2013). Gene-alcohol interactions identify several novel blood pressure loci including a promising locus near SLC16A9. Front Genet *4*, 277.

145. Sung, Y.J., las Fuentes, de, L., Schwander, K.L., Simino, J., and Rao, D.C. (2015). Gene-smoking interactions identify several novel blood pressure loci in the Framingham Heart Study. Am. J. Hypertens. *28*, 343–354.

146. Basson, J., Sung, Y.J., Schwander, K., Kume, R., Simino, J., las Fuentes, de, L., and Rao, D. (2014). Gene-education interactions identify novel blood pressure Loci in the Framingham Heart Study. Am. J. Hypertens. *27*, 431–444.

147. Takeuchi, F., Isono, M., Katsuya, T., Yamamoto, K., Yokota, M., Sugiyama, T., Nabika, T., Fujioka, A., Ohnaka, K., Asano, H., et al. (2010). Blood Pressure and Hypertension Are Associated With 7 Loci in the Japanese Population. Circulation *121*, 2302–2309.

148. Ganesh, S.K., Tragante, V., Guo, W., Guo, Y., Lanktree, M.B., Smith, E.N., Johnson, T., Castillo, B.A., Barnard, J., Baumert, J., et al. (2013). Loci influencing blood pressure identified using a cardiovascular gene-centric array. Hum. Mol. Genet. *22*, 1663–1678.

149. Johnson, A.D., Newton-Cheh, C., Chasman, D.I., Ehret, G.B., Johnson, T., Rose, L., Rice, K., Verwoert, G.C., Launer, L.J., Gudnason, V., et al. (2011). Association of hypertension drug target genes with blood pressure and hypertension in 86,588 individuals. Hypertension *57*, 903–910.

150. Johnson, T., Gaunt, T.R., Newhouse, S.J., Padmanabhan, S., Tomaszewski, M., Kumari, M., Morris, R.W., Tzoulaki, I., O'Brien, E.T., Poulter, N.R., et al. (2011). Blood Pressure Loci Identified with a Gene-Centric Array. Am. J. Hum. Genet. *89*, 688–700.

151. Wang, Y., O’connell, J.R., McArdle, P.F., Wade, J.B., Dorff, S.E., Shah, S.J., Shi, X., Pan, L., Rampersaud, E., Shen, H., et al. (2009). Whole-genome association study identifies STK39 as a hypertension susceptibility gene. Proc. Natl. Acad. Sci. U.S.A. *106*, 226–231.

152. Tragante, V., Barnes, M.R., Ganesh, S.K., Lanktree, M.B., Guo, W., Franceschini, N., Smith, E.N., Johnson, T., Holmes, M.V., Padmanabhan, S., et al. (2014). Gene-centric meta-analysis in 87,736 individuals of European ancestry identifies multiple blood-pressure-related loci. Am. J. Hum. Genet. *94*, 349–360.

153. Hong, K.-W., Min, H., Heo, B.-M., Joo, S.E., Kim, S.S., and Kim, Y. (2012). Recapitulation of genome-wide association studies on pulse pressure and mean arterial pressure in the Korean population. J. Hum. Genet. *57*, 391–393.

154. Kelly, T.N., Takeuchi, F., Tabara, Y., Edwards, T.L., Kim, Y.J., Chen, P., Li, H., Wu, Y., Yang, C.-F., Zhang, Y., et al. (2013). Genome-wide association study meta-analysis reveals transethnic replication of mean arterial and pulse pressure loci. Hypertension *62*, 853–859.

155. Ho, J.E., Levy, D., Rose, L., Johnson, A.D., Ridker, P.M., and Chasman, D.I. (2011). Discovery and replication of novel blood pressure genetic loci in the Women's Genome Health Study. J Hypertens *29*, 62–69.

156. Lin, Y., Lai, X., Chen, B., Xu, Y., Huang, B., Chen, Z., Zhu, S., Yao, J., Jiang, Q., Huang, H., et al. (2011). Genetic variations in CYP17A1, CACNB2 and PLEKHA7 are associated with blood pressure and/or hypertension in she ethnic minority of China. Atherosclerosis *219*, 709–714.

157. Lango Allen, H., Estrada, K., Lettre, G., Berndt, S.I., Weedon, M.N., Rivadeneira, F., Willer, C.J., Jackson, A.U., Vedantam, S., Raychaudhuri, S., et al. (2010). Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature *467*, 832–838.

158. Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., Mccarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., et al. (2009). Finding the missing heritability of complex diseases. Nature *461*, 747–753.

159. Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., et al. (2010). Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. *42*, 565–569.

160. Vattikuti, S., Guo, J., and Chow, C.C. (2012). Heritability and genetic correlations explained by common SNPs for metabolic syndrome traits. PLoS Genet. *8*, e1002637.

161. MacRae, C.A., and Vasan, R.S. (2011). Next-generation genome-wide association studies: time to focus on phenotype? Circ Cardiovasc Genet *4*, 334–336.

162. Kulminski, A.M., Culminskaya, I., Arbeev, K.G., Arbeeva, L., Ukraintseva, S.V., Stallard, E., Wu, D., and Yashin, A.I. (2015). Birth Cohort, Age, and Sex Strongly Modulate Effects of Lipid Risk Alleles Identified in Genome-Wide Association Studies. PLoS ONE 10, e0136319.

163. Wray, N.R., Lee, S.H., and Kendler, K.S. (2012). Impact of diagnostic misclassification on estimation of genetic correlations using genome-wide genotypes. Eur. J. Hum. Genet. *20*, 668–674.

164. van der Sluis, S., Verhage, M., Posthuma, D., and Dolan, C.V. (2010). Phenotypic complexity, measurement bias, and poor phenotypic resolution contribute to the missing heritability problem in genetic association studies. PLoS ONE *5*, e13929.

165. Chasman, D.I., Paré, G., and Ridker, P.M. (2009). Population-based genomewide genetic analysis of common clinical chemistry analytes. Clin. Chem. *55*, 39–51.

166. Bovet, P., Gervasoni, J.-P., Ross, A.G., Mkamba, M., Mtasiwa, D.M., Lengeler, C., Burnier, M., and Paccaud, F. (2003). Assessing the prevalence of hypertension in populations: are we doing it right? J Hypertens *21*, 509–517.

167. Borecki, I.B., and Province, M.A. (2008). Genetic and genomic discovery using family studies. Circulation *118*, 1057–1063.

168. Krull, J.L. (2007). Using multilevel analyses with sibling data to increase analytic power: an illustration and simulation study. Dev Psychol *43*, 602–619.

169. McCartney, G., Russ, T.C., Walsh, D., Lewsey, J., Smith, M., Davey Smith, G., Stamatakis, E., and Batty, G.D. (2014). Explaining the excess mortality in Scotland compared with England: pooling of 18 cohort studies. J Epidemiol Community Health *69*, jech–2014–204185–27.

170. Generation Scotland. Available online: http://www.generationscotland.org

171. Smith, B.H., Campbell, A., Linksted, P., Fitzpatrick, B., Jackson, C., Kerr, S.M., Deary, I.J., MacIntyre, D.J., Campbell, H., McGilchrist, M., et al. (2013). Cohort Profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. Int J Epidemiol *42*, 689–700.

172. Smith, B.H., Campbell, H., Blackwood, D., Connell, J., Connor, M., Deary, I.J., Dominiczak, A.F., Fitzpatrick, B., Ford, I., Jackson, C., et al. (2006). Generation Scotland: the Scottish Family Health Study; a new resource for researching genes and heritability. BMC Med. Genet. *7*, 1–9.

173. Haddow, G., Cunningham-Burley, S., Bruce, A., and Parry, S. (2008). Generation Scotland: consulting publics and specialists at an early stage in a genetic database's development. Critical Public Health *18*, 139–149.

174. Haddow, G., Cunningham-Burley, S., Bruce, A., and Parry, S. (2004). Generation Scotland Preliminary Consultation Exercise 2003–04: Public and Stakeholder Views from Focus Groups and Interviews. Avilable online: http://www.innogen.ac.uk/working-papers/537

175. Cui, J.S., Hopper, J.L., and Harrap, S.B. (2003). Antihypertensive Treatments Obscure Familial Contributions to Blood Pressure Variation. Hypertension *41*, 207–210.

176. Hunt, S.C., Ellison, R.C., Atwood, L.D., Pankow, J.S., Province, M.A., and Leppert, M.F. (2002). Genome scans for blood pressure and hypertension: the National Heart, Lung, and Blood Institute Family Heart Study. Hypertension *40*, 1–6.

177. Joint Formulary Committee British National Formulary. Avilable online: https://www.medicinescomplete.com/mc/bnf/current/

178. Paganini-Hill, A., Paganini-Hill, A., ROSS, R.K., and Ross, R.K. (1982). Reliability of recall of drug usage and other health-related information. Am. J. Epidemiol. *116*, 114–122.

179. Monster, T.B.M., Janssen, W.M.T., De Jong, P.E., de Jong-van den Berg, L.T.W., PREVEND Study Group Prevention of REnal and Vascular ENT Stage Disease (2002). Pharmacy data in epidemiological studies: an easy to obtain and reliable tool. Pharmacoepidemiol Drug Saf *11*, 379–384.

180. Haapea, M., Miettunen, J., Lindeman, S., Joukamaa, M., and Koponen, H. (2010). Agreement between self-reported and pharmacy data on medication use in the Northern Finland 1966 Birth Cohort. Int J Methods Psychiatr Res *19*, 88–96.

181. Lau, H.S., Lau, H.S., de Boer, A., de Boer, A., Beuning, K.S., Beuning, K.S., Porsius, A., and Porsius, A. (1997). Validation of pharmacy records in drug exposure assessment. J Clin Epidemiol *50*, 619–625.

182. Haukka, J., Haukka, J., Suvisaari, J., Suvisaari, J., Tuulio-Henriksson, A., Tuulio-Henriksson, A., Lönnqvist, J., and Lönnqvist, J. (2007). High concordance between self-reported medication and official prescription database information. Eur. J. Clin. Pharmacol. *63*, 1069–1074.

183. Nielsen, M.W., Nielsen, M.W., Søndergaard, B., Søndergaard, B., Kjøller, M., Kjøller, M., Hansen, E.H., and Hansen, E.H. (2008). Agreement between self-reported data on medicine use and prescription records vary according to method of analysis and therapeutic group. J Clin Epidemiol *61*, 919–924.

184. Scottish Executive Scottish Index of Multiple Deprivation: 2009 General Report. Avilable online:
http://www.gov.scot/resource/doc/289599/0088642.pdf

185. Kutyavin, I.V., Kutyavin, I.V., Afonina, I.A., Afonina, I.A., Mills, A., Mills, A., Gorn, V.V., Gorn, V.V., Lukhtanov, E.A., Lukhtanov, E.A., et al. (2000). 3'-minor groove binder-DNA probes increase sequence specificity at PCR extension temperatures. Nucleic Acids Res. *28*, 655–661.

186. Livak, K.J., and Livak, K.J. (1999). Allelic discrimination using fluorogenic probes and the 5' nuclease assay. Genet. Anal. *14*, 143–149.

187. Livak, K.J., Flood, S.J., Marmaro, J., Giusti, W., and Deetz, K. (1995). Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. PCR Methods Appl. *4*, 357–362.

188. Kim, S., and Misra, A. (2007). SNP Genotyping: Technologies and Biomedical Applications. Annu. Rev. Biomed. Eng. *9*, 289–320.

189. Ragoussis, J. (2009). Genotyping technologies for genetic research. Annu Rev Genomics Hum Genet *10*, 117–133.

190. Teo, Y.Y. (2008). Common statistical issues in genome-wide association studies: a review on power, data quality control, genotype calling and population structure. Curr. Opin. Lipidol. *19*, 133–143.

191. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., et al. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. Am. J. Hum. Genet. *81*, 559–575.

192. Abecasis, G.R., Cherny, S.S., Cookson, W.O., and Cardon, L.R. (2001). Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. Nat. Genet. *30*, 97–101.

193. Smith, J.G., and Newton-Cheh, C. (2009). Genome-Wide Association Study in Humans. In Ischemic Stroke in Mice and Rats, (Totowa, NJ: Humana Press), pp. 231–258.

194. Wittke-Thompson, J.K., Pluzhnikov, A., and Cox, N.J. (2005). Rational inferences about departures from Hardy-Weinberg equilibrium. Am. J. Hum. Genet. *76*, 967–986.

195. Wigginton, J.E., Wigginton, J.E., and Abecasis, G.R. (2005). PEDSTATS: descriptive statistics, graphics and quality assessment for gene mapping data. Bioinformatics *21*, 3445–3447.

196. Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T., Hindorff, L., et al. (2014). The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res. *42*, D1001–D1006.

197. Hindorff, L.A., H, J., Absher, D., Hall, P., Klemm, A., and Manolio, T.A. A Catalog of Published Genome-Wide Association Studies. Avilable online: http://www.ebi.ac.uk/gwas/

198. Johnson, A.D., Handsaker, R.E., Pulit, S.L., Nizzari, M.M., O'donnell, C.J., and de Bakker, P.I.W. (2008). SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics *24*, 2938–2939.

199. Adeyemo, A., Gerry, N., Chen, G., Herbert, A., Doumatey, A., Huang, H., Zhou, J., Lashley, K., Chen, Y., Christman, M., et al. (2009). A Genome-Wide Association Study of Hypertension and Blood Pressure in African Americans. PLoS Genet. *5*, e1000564.

200. Welch, C., Petersen, I., Walters, K., Morris, R.W., Nazareth, I., Kalaitzaki, E., White, I.R., Marston, L., and Carpenter, J. (2012). Two-stage method to remove population- and individual-level outliers from longitudinal data in a primary care database. Pharmacoepidemiol Drug Saf *21*, 725–732.

201. Wickham, H. (2009). ggplot2: Elegant Graphics for Data Analysis (New York: Springer).

202. Goldberg, S.I., Niemierko, A., and Turchin, A. (2008). Analysis of data errors in clinical research databases. AMIA Annu Symp Proc *2008*, 242–246.

203. Kerr, S.M., Campbell, A., Murphy, L., Hayward, C., Jackson, C., Wain, L.V., Tobin, M.D., Dominiczak, A., Morris, A., Smith, B.H., et al. (2013). Pedigree and genotyping quality analyses of over 10,000 DNA samples from the Generation Scotland: Scottish Family Health Study. BMC Med. Genet. *14*, 1–7.

204. Klungel, O.H., de Boer, A., Paes, A.H., Herings, R.M., Seidell, J.C., and Bakker, A. (2000). Influence of question structure on the recall of self-reported drug use. J Clin Epidemiol *53*, 273–277.

205. West, S.L., Ritchey, M.E., and Poole, C. (2013). Validity of Pharmacoepidemiologic Drug and Diagnosis Data. In Textbook of Pharmacoepidemiology, (Chichester, UK: John Wiley & Sons Ltd), pp. 203–227.

206. Schneeweiss, S., and Avorn, J. (2005). A review of uses of health care utilization databases for epidemiologic research on therapeutics. J Clin Epidemiol *58*, 323–337.

207. McCartney, G., Walsh, D., Whyte, B., and Collins, C. (2012). Has Scotland always been the “sick man” of Europe? An observational study from 1855 to 2006. Eur J Public Health *22*, 756–760.

208. Hanlon, P., Lawder, R.S., Buchanan, D., Redpath, A., Walsh, D., Wood, R., Bain, M., Brewster, D.H., and Chalmers, J. (2005). Why is mortality higher in Scotland than in England and Wales? Decreasing influence of socioeconomic deprivation between 1981 and 2001 supports the existence of a “Scottish Effect.” Journal of Public Health *27*, 199–204.

209. Landis, J.R., and Koch, G.G. (1977). The measurement of observer agreement for categorical data. Biometrics *33*, 159–174.

210. Byrt, T., Bishop, J., and Carlin, J.B. (1993). Bias, prevalence and kappa. J Clin Epidemiol *46*, 423–429.

211. Sim, J., and Wright, C.C. (2005). The kappa statistic in reliability studies: use, interpretation, and sample size requirements. Phys Ther *85*, 257–268.

212. Feinstein, A.R., and Cicchetti, D.V. (1990). High agreement but low kappa: I. The problems of two paradoxes. J Clin Epidemiol *43*, 543–549.

213. Hoehler, F.K. (2000). Bias and prevalence effects on kappa viewed in terms of sensitivity and specificity. J Clin Epidemiol *53*, 499–503.

214. Cicchetti, D.V., and Feinstein, A.R. (1990). High agreement but low kappa: II. Resolving the paradoxes. J Clin Epidemiol *43*, 551–558.

215. Chobanian, A.V., Bakris, G.L., Black, H.R., Cushman, W.C., Green, L.A., Izzo, J.L., Jones, D.W., Materson, B.J., Oparil, S., Wright, J.T., et al. (2003). Seventh report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. Hypertension *42*, 1206–1252.

216. Diez-Roux, A.V. (2000). Multilevel Analysis in Public Health Research. Annu. Rev. Public. Health. *21*, 171–192.

217. Bates, D., Maechler, M., Ben Bolker, and Walker, S. (2014). *lme4: Linear mixed-effects models using Eigen and S4.* R package version 1.1-7.

218. Richardson, K., Kenny, R.A., Peklar, J., and Bennett, K. (2013). Agreement between patient interview data on prescription medication use and pharmacy records in those aged older than 50 years varied by therapeutic group and reporting of indicated health conditions. J Clin Epidemiol *66*, 1308–1316.

219. Caskie, G.I.L., Willis, S.L., Warner Schaie, K., and Zanjani, F.A.K. (2005). Congruence of medication information from a brown bag data collection and pharmacy records: findings from the Seattle longitudinal study. Exp Aging Res *32*, 79–103.

220. Klungel, O.H., de Boer, A., Paes, A.H.P., Herings, R.M.C., Seidell, J.C., and

Bakker, A. (1999). Agreement between self-reported antihypertensive drug use and pharmacy records in a population-based study in The Netherlands. Pharm World Sci *21*, 217–220.

221. Boudreau, D.M., Daling, J.R., Malone, K.E., Gardner, J.S., Blough, D.K., and Heckbert, S.R. (2004). A validation study of patient interview data and pharmacy records for antihypertensive, statin, and antidepressant medication use among older women. Am. J. Epidemiol. *159*, 308–317.

222. Caskie, G.I.L., and Willis, S.L. (2004). Congruence of self-reported medications with pharmacy prescription records in low-income older adults. Gerontologist *44*, 176–185.

223. Gama, H., Correia, S., and Lunet, N. (2009). Questionnaire design and the recall of pharmacological treatments: a systematic review. Pharmacoepidemiol Drug Saf *18*, 175–187.

224. Wolf-Maier, K., Cooper, R.S., Banegas, J.R., Giampaoli, S., Hense, H.-W., Joffres, M., Kastarinen, M., Poulter, N., Primatesta, P., Rodríguez-Artalejo, F., et al. (2003). Hypertension Prevalence and Blood Pressure Levels in 6 European Countries, Canada, and the United States. JAMA *289*, 2363–2369.

225. Joffres, M.R., Campbell, N.R.C., Manns, B., and Tu, K. (2007). Estimate of the benefits of a population-based reduction in dietary sodium additives on hypertension and its related health care costs in Canada. Can J Cardiol *23*, 437–443.

226. Polonia, J., Martins, L., Pinto, F., and Nazare, J. (2014). Prevalence, awareness, treatment and control of hypertension and salt intake in Portugal: changes over a decade. The PHYSA study. J Hypertens *32*, 1211–1221.

227. Hammouche, S., and Holland, R. (2011). Does quality of care for hypertension in primary care vary with postcode area deprivation? An observational study. BMC Health Services Research *11*, 297.

228. Ashworth, M., Medina, J., and Morgan, M. (2008). Effect of social deprivation on blood pressure monitoring and control in England: a survey of data from the quality and outcomes framework. BMJ *337*, a2030.

229. MONGEAU, J.G., BIRON, P., and Sing, C.F. (2009). The Influence of Genetics and Household Environment upon the Variability of Normal Blood Pressure: The Montreal Adoption Survey. Clinical and Experimental Hypertension. Part a: Theory and Practice *8*, 653–660.

230. Harrap, S.B., Stebbing, M., Hopper, J.L., Hoang, H.N., and Giles, G.G. (2000). Familial patterns of covariation for cardiovascular risk factors in adults: The Victorian Family Heart Study. Am. J. Epidemiol. *152*, 704–715.

231. Hunt, S.C., Williams, R.R., and Barlow, G.K. (1986). A comparison of positive family history definitions for defining risk of future disease. J Chronic Dis *39*, 809–821.

232. Daniels, P.R., Kardia, S.L.R., Hanis, C.L., Brown, C.A., Hutchinson, R., Boerwinkle, E., Turner, S.T., Genetic Epidemiology Network of Arteriopathy

study (2004). Familial aggregation of hypertension treatment and control in the Genetic Epidemiology Network of Arteriopathy (GENOA) study. Am. J. Med. *116*, 676–681.

233. van Rijn, M.J.E., Schut, A.F.C., Aulchenko, Y.S., Deinum, J., Sayed-Tabatabaei, F.A., Yazdanpanah, M., Isaacs, A., Axenovich, T.I., Zorkoltseva, I.V., Zillikens, M.C., et al. (2007). Heritability of blood pressure traits and the genetic contribution to blood pressure variance explained by four blood-pressure-related genes. J Hypertens *25*, 565–570.

234. Alwan, H., Ehret, G., Ponte, B., Pruijm, M., Ackermann, D., Guessous, I., Staessen, J.A., Asayama, K., Kutalik, Z., Vuistiner, P., et al. (2015). Heritability of ambulatory and office blood pressure in the Swiss population. J Hypertens *33*, 2061–2067.

235. Mitchell, G.F., Destefano, A.L., Larson, M.G., Benjamin, E.J., Chen, M.-H., Vasan, R.S., Vita, J.A., and Levy, D. (2005). Heritability and a genome-wide linkage scan for arterial stiffness, wave reflection, and mean arterial pressure: the Framingham Heart Study. Circulation *112*, 194–199.

236. Fava, C., Burri, P., Almgren, P., Groop, L., Hulthén, U.L., and Melander, O. (2004). Heritability of ambulatory and office blood pressure phenotypes in Swedish families. J Hypertens *22*, 1717–1721.

237. Wang, X., and Snieder, H. (2013). Familial Aggregation of Blood Pressure. In Pediatric Hypertension, (Totowa, NJ: Humana Press), pp. 195–209.

238. Wang, X., Ding, X., Su, S., Yan, W., Harshfield, G., Treiber, F., and Snieder, H. (2009). Genetic influences on daytime and night-time blood pressure: similarities and differences. J Hypertens *27*, 2358–2364.

239. Kupper, N., Willemsen, G., Riese, H., Posthuma, D., Boomsma, D.I., and de Geus, E.J.C. (2004). Heritability of Daytime Ambulatory Blood Pressure in an Extended Twin Design. Hypertension *45*, 80–85.

240. Bochud, M. (2011). Estimating Heritability from Nuclear Family and Pedigree Data. In Methods in Molecular Biology, (Totowa, NJ: Humana Press), pp. 171–186.

241. Caulfield, M., Munroe, P., Pembroke, J., Samani, N., Dominiczak, A., Brown, M., Benjamin, N., Webster, J., Ratcliffe, P., O'Shea, S., et al. (2003). Genome-wide mapping of human loci for essential hypertension. Lancet *361*, 2118–2123.

242. Wang, X., and Zhang, H. (2001). Ascertainment in Genetics Studies. In Encyclopedia of Life Sciences. John Wiley & Sons, Ltd, Chichester. Available online: http://www.els.net/.

243. Olson, J.M., and Cordell, H.J. (2000). Ascertainment bias in the estimation of sibling genetic risk parameters. Genet. Epidemiol. *18*, 217–235.

244. S.A.G.E (2012). Statistical Analysis for Genetic Epidemiology.

245. Almasy, L., and Blangero, J. (1998). Multipoint quantitative-trait linkage

analysis in general pedigrees. Am. J. Hum. Genet. *62*, 1198–1211.

246. Fermino, R.C., Seabra, A., Garganta, R., and Maia, J.A.R. (2009). Genetic factors in familial aggregation of blood pressure of Portuguese nuclear families. Arq. Bras. Cardiol. *92*, 199–204.

247. Fuentes, R.M., Notkola, I.L., Shemeikka, S., Tuomilehto, J., and Nissinen, A. (2000). Familial aggregation of blood pressure: a population-based family study in eastern Finland. J Hum Hypertens *14*, 441–445.

248. Saunders, C.L., and Gulliford, M.C. (2006). Heritabilities and shared environmental effects were estimated from household clustering in national health survey data. J Clin Epidemiol *59*, 1191–1198.

249. Hu, Y., He, L., Wu, Y., Ma, G., Li, L., and Hu, Y. (2013). Familial correlation and aggregation of body mass index and blood pressure in Chinese Han population. BMC Public Health *13*, 686.

250. van Dongen, J., Willemsen, G., Chen, W.-M., De Geus, E.J.C., and Boomsma, D.I. (2013). Heritability of metabolic syndrome traits in a large population-based sample. J. Lipid Res. *54*, 2914–2923.

251. Pilia, G., Chen, W., Scuteri, A., Orru, M., Albai, G., Dei, M., Lai, S., Usala, G., Lai, M., Loi, P., et al. (2005). Heritability of Cardiovascular and Personality Traits in 6,148 Sardinians. PLoS Genet. *preprint*, e132.

252. Allison, D.B., Neale, M.C., Kezis, M.I., Alfonso, V.C., Heshka, S., and Heymsfield, S.B. (1996). Assortative mating for relative weight: Genetic implications. Behav Genet *26*, 103–111.

253. Kwagyan, J., Tabe, C.E., Xu, S., Maqbool, A.R., Gordeuk, V.R., and Randall, O.S. (2005). The impact of body mass index on pulse pressure in obesity. J Hypertens *23*, 619–624.

254. Jelenkovic, A., Poveda, A., and Rebato, E. (2010). A statistical investigation into the sharing of common genetic factors between blood pressure and obesity phenotypes in nuclear families from the Greater Bilbao (Spain). J Hypertens 28, 723–731.

255. Biino, G., Parati, G., Concas, M.P., Adamo, M., Angius, A., Vaccargiu, S., and Pirastu, M. (2013). Environmental and genetic contribution to hypertension prevalence: data from an epidemiological survey on Sardinian genetic isolates. PLoS ONE 8, e59612.

256. Wu, T., Snieder, H., Li, L., Cao, W., Zhan, S., Lv, J., Gao, W., Wang, X., Ding, X., and Hu, Y. (2011). Genetic and environmental influences on blood pressure and body mass index in Han Chinese: a twin study. Hypertension Research *34*, 173–179.

257. Marioni, R.E., Davies, G., Hayward, C., Liewald, D., Kerr, S.M., Campbell, A., Luciano, M., Smith, B.H., Padmanabhan, S., Hocking, L.J., et al. (2014). Molecular genetic contributions to socioeconomic status and intelligence. Intelligence *44*, 26–32.

258. Dubay, C., Vincent, M., Samani, N.J., Hilbert, P., Kaiser, M.A., Beressi, J.P., Kotelevtsev, Y., Beckmann, J.S., Soubrier, F., and Sassard, J. (1993). Genetic determinants of diastolic and pulse pressure map to different loci in Lyon hypertensive rats. Nat. Genet. *3*, 354–357.

259. Llamas, B., Lau, C., Cupples, W.A., Rainville, M.-L., Souzeau, E., and Deschepper, C.F. (2006). Genetic determinants of systolic and pulse pressure in an intercross between normotensive inbred rats. Hypertension *48*, 921–926.

260. Cui, J., Hopper, J.L., and Harrap, S.B. (2002). Genes and Family Environment Explain Correlations Between Blood Pressure and Body Mass Index. Hypertension *40*, 7–12.

261. Tarnoki, A.D., Tarnoki, D.L., Bogl, L.H., Medda, E., Fagnani, C., Nisticò, L., Stazi, M.A., Brescianini, S., Lucatelli, P., Boatta, E., et al. (2013). Association of body mass index with arterial stiffness and blood pressure components: a twin study. Atherosclerosis *229*, 388–395.

262. Guttmacher, A.E., Collins, F.S., and Carmona, R.H. (2004). The Family History — More Important Than Ever. N. Engl. J. Med. *351*, 2333–2336.

263. Havulinna, A.S., Kettunen, J., Ukkola, O., Osmond, C., Eriksson, J.G., Kesäniemi, Y.A., Jula, A., Peltonen, L., Kontula, K., Salomaa, V., et al. (2013). A blood pressure genetic risk score is a significant predictor of incident cardiovascular events in 32 669 individuals. Hypertension *61*, 987–994.

264. Lu, X., Huang, J., Wang, L., Chen, S., Yang, X., Li, J., Cao, J., Chen, J., Li, Y., Zhao, L., et al. (2015). Genetic Predisposition to Higher Blood Pressure Increases Risk of Incident Hypertension and Cardiovascular Diseases in Chinese. Hypertension *66*, 786–792.

265. Manichaikul, A., Chen, W.-M., Williams, K., Wong, Q., Sale, M.M., Pankow, J.S., Tsai, M.Y., Rotter, J.I., Rich, S.S., and Mychaleckyj, J.C. (2012). Analysis of family- and population-based samples in cohort genome-wide association studies. Hum. Genet. *131*, 275–287.

266. Zhang, Z., Ersoz, E., Lai, C.-Q., Todhunter, R.J., Tiwari, H.K., Gore, M.A., Bradbury, P.J., Yu, J., Arnett, D.K., Ordovas, J.M., et al. (2010). Mixed linear model approach adapted for genome-wide association studies. Nat. Genet. *42*, 355–360.

267. Kang, H.M., Sul, J.H., Service, S.K., Zaitlen, N.A., Kong, S.-Y., Freimer, N.B., Sabatti, C., and Eskin, E. (2010). Variance component model to account for sample structure in genome-wide association studies. Nat. Genet. *42*, 348–354.

268. Zhou, X., and Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. Nat. Genet. *44*, 821–824.

269. Sinnwell, J.P., Therneau, T.M., and Schaid, D.J. (2014). The kinship2 R package for pedigree data. Hum Hered *78*, 91–93.

270. Chen, W.-M., Manichaikul, A., and Rich, S.S. (2009). A Generalized Family-Based Association Test for Dichotomous Traits. Am. J. Hum. Genet. *85*, 364–376.

271. Zois, N.E., Bartels, E.D., Hunter, I., Kousholt, B.S., Olsen, L.H., and Goetze, J.P. (2014). Natriuretic peptides in cardiometabolic regulation and disease. Nat Rev Cardiol *11*, 403–412.

272. Arora, P., Wu, C., Khan, A.M., Bloch, D.B., Davis-Dusenbery, B.N., Ghorbani, A., Spagnolli, E., Martinez, A., Ryan, A., Tainsh, L.T., et al. (2013). Atrial natriuretic peptide is negatively regulated by microRNA-425. J. Clin. Invest. *123*, 3378–3382.

273. Gong, Y., McDonough, C.W., Wang, Z., Hou, W., Cooper-Dehoff, R.M., Langaee, T.Y., Beitelshees, A.L., Chapman, A.B., Gums, J.G., Bailey, K.R., et al. (2012). Hypertension Susceptibility Loci and Blood Pressure Response to Antihypertensives: Results From the Pharmacogenomic Evaluation of Antihypertensive Responses Study. Circ Cardiovasc Genet *5*, 686–691.

274. Zhu, X., Young, J.H., Fox, E., Keating, B.J., Franceschini, N., Kang, S., Tayo, B., Adeyemo, A., Sun, Y.V., Li, Y., et al. (2011). Combined admixture mapping and association analysis identifies a novel blood pressure genetic locus on 5p13: contributions from the CARe consortium. Hum. Mol. Genet. *20*, 2285–2295.

275. Antzelevitch, C., Pollevick, G.D., Cordeiro, J.M., Casis, O., Sanguinetti, M.C., Aizawa, Y., Guerchicoff, A., Pfeiffer, R., Oliva, A., Wollnik, B., et al. (2007). Loss-of-function mutations in the cardiac calcium channel underlie a new clinical entity characterized by ST-segment elevation, short QT intervals, and sudden cardiac death. Circulation *115*, 442–449.

276. Niu, Y., Gong, Y., Langaee, T.Y., Davis, H.M., Elewa, H., Beitelshees, A.L., Moss, J.I., Cooper-Dehoff, R.M., Pepine, C.J., and Johnson, J.A. (2010). Genetic variation in the beta2 subunit of the voltage-gated calcium channel and pharmacogenetic association with adverse cardiovascular outcomes in the INternational VErapamil SR-Trandolapril STudy GENEtic Substudy (INVEST-GENES). Circ Cardiovasc Genet *3*, 548–555.

277. Palatini, P., Ceolotto, G., Ragazzo, F., Dorigatti, F., Saladini, F., Papparella, I., Mos, L., Zanata, G., and Santonastaso, M. (2009). CYP1A2 genotype modifies the association between coffee intake and the risk of hypertension. J Hypertens *27*, 1594–1601.

278. Guessous, I., Dobrinas, M., Kutalik, Z., Pruijm, M., Ehret, G., Maillard, M., Bergmann, S., Beckmann, J.S., Cusi, D., Rizzi, F., et al. (2012). Caffeine intake and CYP1A2 variants associated with high caffeine intake protect non-smokers from hypertension. Hum. Mol. Genet. *21*, 3283–3292.

279. Bochud, M., Chiolero, A., Elston, R.C., and Paccaud, F. (2008). A cautionary note on the use of Mendelian randomization to infer causation in observational epidemiology. Int J Epidemiol *37*, 414–416.

280. Guessous, I., Eap, C.B., and Bochud, M. (2014). Blood Pressure in Relation to Coffee and Caffeine Consumption. Curr. Hypertens. Rep. *16*, 1–9.

281. Ueyama, C., Horibe, H., Yamase, Y., Fujimaki, T., Oguri, M., Kato, K., Arai, M., Watanabe, S., Murohara, T., and Yamada, Y. (2015). Association of FURIN and ZPR1 polymorphisms with metabolic syndrome. Biomed Rep *3*, 641–

647.

282. Turpeinen, H., Seppälä, I., Lyytikäinen, L.-P., Raitoharju, E., Hutri-Kähönen, N., Levula, M., Oksala, N., Waldenberger, M., Klopp, N., Illig, T., et al. (2015). A genome-wide expression quantitative trait loci analysis of proprotein convertase subtilisin/kexin enzymes identifies a novel regulatory gene variant for FURIN expression and blood pressure. Hum. Genet. *134*, 627–636.

283. Cousin, C., Bracquart, D., Contrepas, A., Corvol, P., Muller, L., and Nguyen, G. (2009). Soluble form of the (pro)renin receptor generated by intracellular cleavage by furin is secreted in plasma. Hypertension *53*, 1077–1082.

284. Ichihara, A., Sakoda, M., Kurauchi-Mito, A., Narita, T., Kinouchi, K., Murohashi-Bokuda, K., and Itoh, H. (2010). Possible roles of human (pro)renin receptor suggested by recent clinical and experimental findings. Hypertension Research *33*, 177–180.

285. Gudbjartsson, D.F., Holm, H., Indridason, O.S., Thorleifsson, G., Edvardsson, V., Sulem, P., de Vegt, F., d'Ancona, F.C.H., Heijer, den, M., Wetzels, J.F.M., et al. (2010). Association of variants at UMOD with chronic kidney disease and kidney stones-role of age and comorbid diseases. PLoS Genet. *6*, e1001039.

286. Köttgen, A., Hwang, S.-J., Larson, M.G., Van Eyk, J.E., Fu, Q., Benjamin, E.J., Dehghan, A., Glazer, N.L., Kao, W.H.L., Harris, T.B., et al. (2010). Uromodulin levels associate with a common UMOD variant and risk for incident CKD. J. Am. Soc. Nephrol. *21*, 337–344.

287. Graham, L.A., Padmanabhan, S., Fraser, N.J., Kumar, S., Bates, J.M., Raffi, H.S., Welsh, P., Beattie, W., Hao, S., Leh, S., et al. (2014). Validation of uromodulin as a candidate gene for human essential hypertension. Hypertension *63*, 551–558.

288. Trudu, M., Janas, S., Lanzani, C., Debaix, H., Schaeffer, C., Ikehata, M., Citterio, L., Demaretz, S., Trevisani, F., Ristagno, G., et al. (2013). Common noncoding UMOD gene variants induce salt-sensitive hypertension and kidney damage by increasing uromodulin expression. Nat Med *19*, 1655–1660.

289. Grochowski, C.M., Loomes, K.M., and Spinner, N.B. (2016). Jagged1 (JAG1): Structure, expression, and disease associations. Gene *576*, 381–384.

290. Turner, S.T., Boerwinkle, E., O'connell, J.R., Bailey, K.R., Gong, Y., Chapman, A.B., McDonough, C.W., Beitelshees, A.L., Schwartz, G.L., Gums, J.G., et al. (2013). Genomic association analysis of common variants influencing antihypertensive response to hydrochlorothiazide. Hypertension *62*, 391–397.

291. Franklin, S.S., Gustin, W., Wong, N.D., Larson, M.G., Weber, M.A., Kannel, W.B., and Levy, D. (1997). Hemodynamic patterns of age-related changes in blood pressure. The Framingham Heart Study. Circulation *96*, 308–315.

292. Masca, N., Sheehan, N.A., and Tobin, M.D. (2010). Pharmacogenetic interactions and their potential effects on genetic analyses of blood pressure.

Statist. Med. 30, 769–783.

293. Global Lipids Genetics Consortium, Willer, C.J., Schmidt, E.M., Sengupta, S., Peloso, G.M., Gustafsson, S., Kanoni, S., Ganna, A., Chen, J., Buchkovich, M.L., et al. (2013). Discovery and refinement of loci associated with lipid levels. Nat Genet 45, 1274–1283.

294. Speliotes, E.K., Willer, C.J., Berndt, S.I., Monda, K.L., Thorleifsson, G., Jackson, A.U., Lango Allen, H., Lindgren, C.M., Luan, J., Mägi, R., et al. (2010). Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nat Genet 42, 937–948.

295. Fava, C., Sjögren, M., Montagnana, M., Danese, E., Almgren, P., Engström, G., Nilsson, P., Hedblad, B., Guidi, G.C., Minuz, P., et al. (2013). Prediction of blood pressure changes over time and incidence of hypertension by a genetic risk score in Swedes. Hypertension *61*, 319–326.

296. Fava, C., Sjögren, M., Olsson, S., Lövkvist, H., Jood, K., Engström, G., Hedblad, B., Norrving, B., Jern, C., Lindgren, A., et al. (2015). A genetic risk score for hypertension associates with the risk of ischemic stroke in a Swedish case-control study. Eur. J. Hum. Genet. *23*, 969–974.

297. Lim, N.-K., Lee, J.-Y., Lee, J.-Y., Park, H.-Y., and Cho, M.-C. (2015). The Role of Genetic Risk Score in Predicting the Risk of Hypertension in the Korean population: Korean Genome and Epidemiology Study. PLoS ONE 10, e0131603.

298. Visscher, P.M., and Duffy, D.L. (2006). The value of relatives with phenotypes but missing genotypes in association studies for quantitative traits. Genet. Epidemiol. *30*, 30–36.

299. Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: A Tool for Genome-wide Complex Trait Analysis. Am. J. Hum. Genet. *88*, 76–82.

300. Robinson, M.R., Wray, N.R., and Visscher, P.M. (2014). Explaining additional genetic variation in complex traits. Trends in Genetics *30*, 124–132.

301. Bochud, M., Bovet, P., Elston, R.C., Paccaud, F., Falconnet, C., Maillard, M., Shamlaye, C., and Burnier, M. (2005). High heritability of ambulatory blood pressure in families of East African descent. Hypertension *45*, 445–450.

302. Hottenga, J.-J., Boomsma, D.I., Kupper, N., Posthuma, D., Snieder, H., Willemsen, G., and De Geus, E.J.C. (2005). Heritability and Stability of Resting Blood Pressure. Twin Res Hum Genet *8*, 499–508.

303. O'reilly, P.F., Hoggart, C.J., Pomyen, Y., Calboli, F.C.F., Elliott, P., Järvelin, M.-R., and Coin, L.J.M. (2012). MultiPhen: joint model of multiple phenotypes can increase discovery in GWAS. PLoS ONE 7, e34861.

304. Perry, J.R.B., Voight, B.F., Yengo, L., Amin, N., Dupuis, J., Ganser, M., Grallert, H., Navarro, P., Li, M., Qi, L., et al. (2012). Stratifying Type 2 Diabetes Cases by BMI Identifies Genetic Risk Variants in LAMA1 and Enrichment for Risk Variants in Lean Compared to Obese Cases. PLoS Genet. *8*, e1002741.