| | |
|---|---|
| Title | An investigation of the genetic contribution to cognitive resilience in healthy ageing |
| Author(s) | Fitzgerald, Joan |
| Publication Date | 2021-07 |
| Publisher | NUI Galway |
| Item record | http://hdl.handle.net/10379/17021 |

# An investigation of the genetic contribution to cognitive resilience in healthy ageing

By

Joan Fitzgerald, M.Sc.

A thesis submitted for the Degree of Doctor of Philosophy to the Discipline of Biochemistry, School of Natural Sciences, National University of Ireland, Galway.

Supervisor: Dr. Derek Morris

Discipline of Biochemistry, School of Natural Sciences, National University of Ireland, Galway.

Co-Supervisor: Prof. Gary Donohoe

Discipline of Psychology, School of Psychology, National University of Ireland, Galway.

*Submitted July 2021*

**TABLE OF CONTENTS:**

Table of content…… II
Declaration ..... VI
Statement of Contribution ..... VII
Acknowledgements ..... VIII
List of Tables ..... X
List of Figures ..... XII
List of abbreviations ..... XVI
Abstract ..... XXI

**1 INTRODUCTION ..... 1**

1.1 Age-related cognitive decline ..... 2
1.1.1 Genetic and cellular processes involved in cognitive decline ..... 3
1.1.2 Epigenetic factors influencing cognitive decline ..... 6
1.1.3 Psychological factors affecting cognitive decline ..... 7
1.1.4 Environmental factors associated with cognitive decline ..... 8
1.2 Theories of cognitive ageing ..... 9
1.2.1 Processing speed and cognitive decline ..... 9
1.2.2 Other theories of cognitive decline ..... 10
1.3 Cognitive resilience/reserve ..... 10
1.3.1 Theories of cognitive reserve ..... 11
1.3.1.1 Noradrenaline theory of cognitive resilience ..... 11
1.3.1.2 Theory of brain and cognitive reserve ..... 13
1.3.1.3 Theory of cognitive reserve, brain reserve and maintenance ..... 15
1.3.1.4 Theory of reserve, maintenance, and compensation ..... 15
1.3.2 Intelligence and resilience ..... 16
1.3.3 Measuring cognitive resilience ..... 17
1.3.3.1 Measuring cognition ..... 17
1.3.3.2 Measuring cognitive resilience ..... 18
1.3.4 Available datasets ..... 22
1.3.4.1 Datasets with longitudinal cognitive data ..... 22
1.4 Genetic component of cognitive resilience ..... 23
1.4.1 Genetics of cognition ..... 23

1.4.1.1 Genome-wide association studies (GWAS) and cognition ..... 25
1.4.1.2 Polygenic scores and cognition ..... 26
1.4.1.3 Rare variants ..... 27
1.4.1.4 Gene by environment interactions ..... 29
1.4.1.5 Current developments in cognitive research ..... 29
1.4.2 Genetics of cognitive resilience ..... 30
1.4.2.1 Overview of the genetics of cognitive resilience ..... 30
1.4.2.2 Genetic studies to date ..... 32
1.5 Limitation of research on genetics of cognitive resilience to date ..... 34
1.6 Project aims ..... 35
**2 MATERIALS AND METHODS ..... 37**
2.1 UK Biobank ..... 41
2.1.1 Participants ..... 41
2.1.2 Genetic data ..... 41
2.1.3 Cognitive data ..... 42
2.1.3.1 Verbal-numerical reasoning test (fluid intelligence) ..... 43
2.1.3.2 Reaction time (RT) ..... 43
2.1.3.3 Other baseline tests ..... 43
2.1.3.4 Web based tests: ..... 44
2.1.4 Creation of cognitive phenotypes ..... 44
2.1.4.1 Current cognitive performance ..... 44
2.1.4.2 Past cognitive performance ..... 46
2.1.5 Performing GWAS with UKB ..... 47
2.1.6 GWAS-by subtraction (GBS) ..... 47
2.1.7 Calculation of sample size after GBS ..... 48
2.1.8 Identification of genomic loci associated with resilience ..... 51
2.1.9 Function analysis of GWAS output ..... 51
2.1.10 Mapping SNPs to genes ..... 52
2.1.10.1 MAGMA gene-based analysis ..... 52
2.1.11 Comparison with published traits ..... 53
2.1.12 Mendelian randomisation ..... 56
2.2 Health and Retirement Study (HRS) ..... 56
2.2.1 Genetic data ..... 56

2.2.2 Cognitive phenotypes ..... 57
2.2.2.1 Total Cognitive performance ..... 58
2.2.2.2 Recall ..... 58
2.2.3 Creating longitudinal phenotypes with mixed effect modelling ..... 59
2.2.4 GWAS of cognitive change over time ..... 59
2.3 Other Bioinformatic tools used ..... 59
**3 INVESTIGATION OF RESILIENCE IN THE UK BIOBANK ..... 63**
3.1 Introduction ..... 63
3.2 Selection of Cognitive variables in the UKB ..... 63
3.2.1 Current cognitive performance ..... 63
3.2.2 Proxy measures ..... 65
3.2.3 Generation of the resilience phenotype ..... 66
3.2.4 GWAS of Resilience groups ..... 68
3.2.5 Extraction of Res from EY+Res ..... 70
3.2.5.1 Functional level ..... 70
3.2.5.2 SNP level ..... 71
3.2.6 Extraction of Resilience using SEM ..... 73
3.2.6.1 Initial phenotype development ..... 73
3.2.7 Discovery and replication analysis ..... 77
3.2.8 Analysis of the full sample ..... 80
3.2.9 Correlation analysis ..... 83
3.3 Conclusion ..... 84
**4 FUNCTIONAL ANALYSIS OF RESILIENCE ..... 85**
4.1 Introduction ..... 85
4.2 Description of genetic loci ..... 85
4.2.1 Conditional Analysis ..... 101
4.2.2 Fine Mapping ..... 101
4.2.3 Gene mapping ..... 105
4.2.4 Genome-wide gene-based association analysis (GWGAS) ..... 118
4.2.5 Tissue, cell type and pathway enrichment analysis ..... 123
4.2.5.1 Gene – tissue expression ..... 123
4.2.5.2 Gene-cell type expression ..... 127

4.2.5.3 Pathway enrichment analysis ..... 129
4.2.6 Genetic correlations with other traits ..... 130
4.2.7 Mendelian randomisation ..... 137
4.3 Conclusion ..... 139

**5 AD HOC TESTING TO EXPLORE GWAS OF *RESILIENCE*..... 140**

5.1 Introduction: ..... 140
5.2 Analysis ..... 140
5.2.1 Examination of the relationship of *Resilience* with RT ..... 140
5.2.1.1 Functional analysis ..... 142
5.2.1.2 Pathway enrichment analysis: ..... 145
5.2.1.3 Tissue and cell type enrichment analysis ..... 145
5.2.1.4 LD regression analysis ..... 149
5.2.1.5 Mendelian randomisation ..... 151
5.2.2 Effect of large locus on chromosome 3 ..... 153
5.2.3 Exploration of HRS to verify findings ..... 157
5.2.3.1 Generation of GWAS of cogntive change in HRS ..... 157
5.2.3.2 Examining the relationship of the HRS GWAS with *Resilience* ..... 159
5.3 Conclusion to Chapter 5 ..... 163

**6 DISCUSSION ..... 169**

6.1 Generating a resilience GWAS: ..... 169
6.2 Replication and GWAS output ..... 170
6.3 Replication of findings in an external dataset: ..... 171
6.4 Limitation of current analysis ..... 172
6.5 Understanding the genetic contribution to cognitive resilience ..... 173
6.6 Future studies ..... 174
6.7 Further use of tools used in my research ..... 178
6.8 Challenges ..... 178
6.9 Concluding remarks ..... 180
Oral presentations ..... 212
Poster presentations ..... 212
Awards ..... 212
Publications and Funding ..... 213

## Declaration

I declare that this thesis has not been submitted as an exercise at this or any other university.

I declare that this thesis is entirely my own work, except where otherwise stated.

Signed:

Joan Fitzgerald

## Statement of Contribution

The data used for my thesis was acquired from publicly available resources. The publications that arose from my research were co-authored by members of the CogGene group at NUIG, including Dr. Derek Morris, Prof. Gary Donohoe, Laura Fahey, Dr. Pilib Ó Broin, and Dr. Laurena Holleran.

## Acknowledgements

This thesis is the culmination of 7 years of post-graduate studies at NUIG. I am very grateful to NUIG for the support that is gives to mature students, and I felt included in university life along with my younger colleagues. After I retired from the pharmaceutical industry after 35 years, I wrote a bucket list of what I wanted to do for the next number of years and one of the challenges I gave myself was to study psychology and another was to get a PhD. I did study psychology, but I was attracted back toward the natural sciences where I followed up my higher diploma in psychology with an MSc in Clinical Neuroscience. During the MSc we did a module on genetics which was fascinating to me. When I initially studied genetics (BSc 1976), it was only a small subsection of my biochemistry course (given that the knowledge about the subject was in its infancy).

I was delighted when Derek Morris agreed to take me on for my MSc project. I was even more delighted when he allowed me to use the very limited skills, I had acquired in bioinformatics to examine another of my passions – which was healthy ageing. I had followed the work of The NEIL Memory Research Unit at Trinity College and had volunteered to be tested by them and had attended a number of public meetings. I had also read intensely on the subject.

I thoroughly enjoyed the project and need to thank Laura Whitton for the help she gave me in completing it. Unfortunately, at the same time, April/May 2017, I was also diagnosed with a very rare cancer in my sinuses which took ages to diagnose, so the MSc was a great distraction and kept me positive. I had a couple of small surgical procedures, but the chemo did not start until late June, by which time I had the bulk of my project completed and had only to make finishing touches to my write up. I soon learnt the link between the immune system and cognition where completing the finishing touches depended on my 3-week chemo cycle. As my white cell count dropped (days 7 -10) I found that my ability to focus also dropped and at the worst, I did not have the cognitive ability to boot up my computer and navigate to the correct document. Thankfully, I was back up to full cognitive ability by the end of the cycle and I got my MSc completed during this time. I thank Derek for his support during this period and the MSc coordinator, David Mothershill, and the course Director Gary O'Donohoe for their support during this time – along with my ever-present family and friends.

I was very pleased when Derek and Gary offered me an opportunity to continue my work on the genetics of healthy ageing as a PhD student. I started my PhD on the 10th of January 2018, the same day as I started a radiation regime at the Galway Clinic which was a lot milder than I expected. I did startle Derek on the first day when I arrived with mesh markings on my face from the radiation mask, but he handled it with ease as he did everything during my PhD. He was an excellent supervisor and mentor and has an uncanny ability to spot a mistake in a spread sheet. I would also like to thank Gary for his co-supervision. He gave me great advice and was appropriately challenging.

Almost four years on and my thesis is complete. I have met some wonderful people along the way, particularly all those in the Morris lab who helped me considerably, earlier on Laura W., Mairead and Mark and more recently Laura F., Rebecca, Aodán and Shane. I would also like to say a special thanks to Declan Bennett in the Seoighe lab (Maths) for helping me with coding.

I would also like to acknowledge all the members of the CogGene group who I have interacted with over the years, especially Laurena, who was a sounding board on many occasions. In addition, I would like to thank the wider NICOG group who expanded my knowledge on the broader areas of cognitive neuroscience and all those in the Centre for Chromosome Biology (CCB) and the wider biochemistry department.

I had great support from my family, in particular my youngest son Andy, who is a computer scientist and was able to answer those basic questions in my initial struggle with Linux and bash commands that I was too embarrassed to ask my colleagues. My eldest son, Tony, and his wife Aoife and three kids (my grandkids) were great moral support. My middle son, David, was in New Zealand during all my PhD but was great at long distance support and I did manage to take a break to see him down under in the middle of my research. My siblings were also great particularly my two sisters, Anto and Kate. And finally, I would like to thank Steve, my husband, for all his support during my period in NUIG and his constant supply of fresh oysters, fruit, and vegetables, ensuring I had the perfect diet to age healthily.

## List of Tables

Table 1.1: Longitudinal datasets used to study cognitive decline .....22

Table 2.1: Bioinformatic tools used in this thesis.....37

Table 2.2: Cognitive data in the UKB and its correlation with age.....42

Table 2.3: Comparison of the correlation of age with cognitive measures .....43

Table 2.4: Description of phenotype.....46

Table 2.5: Pathway loading and sample size (effective N) calculation .....50

Table 2.6: Publicly available datasets used for LDSR.....54

Table 2.7: Parameters used in FUMA.....60

Table 3.1: Correlation of cognitive tests with age .....64

Table 3.2 : Genetic loci identified by manual subtraction of EY/NonRes SNPs from EY+Res SNPs.....72

Table 3.3: Sample size for GWAS analysis by cohort.....74

Table 3.4: Replication of GWAS-by-Subtraction: Performed on two subsets of the UKB ....78

Table 3.5: Examining the results of effect sizes and P values for 26 SNPs in the contributing GWAS analysis using the independent significant SNPs in *Resilience.* .....81

Table 4.1: Thirteen genomic risk loci and their independent significant SNPs.....87

Table 4.2: Conditional analysis of each locus conditioning for the index SNP in EY+Res..102

Table 4.3: Summary of Fine Mapping.....103

Table 4.4: Genes significantly associated with *Resilience* .....119

Table 4.5: Mapped genes by all 4 strategies- position, eQTL, chromatin interaction mapping and GWGAS. .....121

Table 4.6: Expression of mapped genes for the 53 tissue samples analysed.....124

Table 4.7: Significant cell type associations across datasets. .....128

Table 4.8: MAGMA Gene-set Analysis .....129

Table 4.9: GWGAS P values for genes in the Wnt signalosome gene set.....130

Table 4.10: LD score regression analysis (LDSC) of cognitive, psychiatric, and neurological traits.....131

Table 4.11: LDSC of brain imaging phenotypes .....132

Table 4.12: *Resilience* SNPs overlapping in published research in GWAS catalog. .....135

Table 4.13: Genetic overlap of gene enrichment .....136

Table 4.14: Generalised Summary data based Mendelian Randomisation.....138

Table 5.1: Comparison of the thirteen index SNPs associated with *Resilience* in GWAS of RT .....141

Table 5.2: Genetic loci overlap between RT and *Resilience.* .....142

Table 5.3: Prioritized genes in RT and *Resilience.* .....144

Table 5.4: Comparison of GO terms between *Resilience* and RT. .....145

Table 5.5: Comparison of tissue enrichment of *Resilience* and RT.....146

Table 5.6: Comparison of LDSR between *Resilience* and RT for cognitive, psychiatric, and neurological phenotypes .....150

Table 5.7: Comparison of LDSR of brain volumes between *Resilience* and RT .....151

Table 5.8: Generalised Summary-data-based Mendelian Randomisation on RT.....152

Table 5.9: Comparison of tissue enrichment of *Resilience* with and without locus 3 .....155

Table 5.10: Comparison of GO terms.....156

Table 5.11: Comparison of Beta and P values between genetic loci associated with *Resilience* and the two HRS GWAS .....160

Table 5.12: Nominally significant SNPs in Recall .....161

Table 5.13: Genetic correlations *Resilience* and Recall and COGTOT.....162

Table 5.14: Comparison of Beta and P values between genetic loci associated with *Resilience* and Recall using 112 lead SNPs .....164

## List of Figures

Figure 1.1: Projected proportion of the population over 60 years of age in 2050 ..... 1

Figure 1.2: Change in general cognitive function ..... 3

Figure 1.3: Biological theories of cognitive decline ..... 6

Figure 1.4: Gene expression changes and master switch in age related cognitive decline ..... 7

Figure 1.5: A hypothetical cognitive reserve network ..... 12

Figure 1.6: Threshold model of cognitive reserve ..... 14

Figure 1.7: Reserve, maintenance, and compensation ..... 15

Figure 1.8: Effect of education on the adjusted grand index score at baseline across ages ..... 16

Figure 1.9: Structural equation modelling (SEM) of the relationships between cognitive abilities and general academic achievement in high school education ..... 20

Figure 1.10: Analytic model to decompose episodic memory in cognitive ageing ..... 21

Figure 1.11: Plot of lead SNPs from GWAS and MTAG of g and EA showing increases in significant findings with increasing participation ..... 26

Figure 1.12: Human hippocampal neurogenesis persists throughout ageing ..... 32

Figure 2.1: Normalising RT cognitive data. ..... 45

Figure 2.2: SEM of GWAS-by-subtraction ..... 49

Figure 2.3: Screenshot of dbGaP showing relevant HRS files ..... 57

Figure 3.1: GWAS of RT in UK biobank (n= 333,664) ..... 64

Figure 3.2: GWAS of ‘g’ in UKB (n=66,740) ..... 65

Figure 3.3: Creation of resilience phenotype. ..... 67

Figure 3.4: Manhattan plot of GWAS of EY+Res (n=156,008) ..... 68

Figure 3.5: Manhattan plot of GWAS of EY (n=165,000) ..... 69

Figure 3.6: Manhattan plot of GWAS of EY/NonRes (n=174,113) ..... 69

Figure 3.7: Overlap of mapped genes in EY+Res and EY/NonRes using BioVenn ..... 70

Figure 3.8: Manhattan plot of *Resilience* generated by subtraction of EY/NonRes SNPs from EY+Res SNPs ..... 71

Figure 3.9: Flow chart of study design. ..... 75

Figure 3.10:SEM of GWAS-by-subtraction. ..... 77

Figure 3.11: Manhattan plot of *discovery.Resilience* (a) and *replication.Resilience* (b) ..... 79

Figure 3.12: Q-Q plots of *Resilience* GWAS (a) and *EduYears* GWAS (b) ..... 80

Figure 3.13: Manhattan plots of *Resilience* GWAS (a) and *EduYears* GWAS (b). ..... 82

Figure 3.14: Heat map of genetic correlations ..... 83

Figure 4.1: Functional consequences of SNPs on genes ..... 86

Figure 4.2: Summary per genomic risk locus ..... 86

Figure 4.3: Locus 2 - Regional plot - 2:59987310-59988258 ..... 88

Figure 4.4: Locus 3 Regional plot - 3:49385350-50250837. ..... 89

Figure 4.5: Locus 4A Regional plot - 4:89455635-89612380. ..... 90

Figure 4.6: Locus 4B Regional plot - 4:106048360-106335951. ..... 91

Figure 4.7: Locus 5A Regional plot 5: 112015555-112176756. ..... 92

Figure 4.8: Locus 5B Regional plot 5: 139517197-139714690. ..... 93

Figure 4.9: Locus Regional plot – 6: 98274701-98450190. ..... 94

Figure 4.10: Locus 12A Regional plot – 12: 13414139-13417617. ..... 95

Figure 4.11: Locus 12B. Regional plot- 12: 49387955-49479968. ..... 96

Figure 4.12: Locus 12C Regional plot – 12: 111818487-112817847. ..... 97

Figure 4.13: Locus 13 Regional plot – 13:58250322 – 58796832. ..... 98

Figure 4.14: Locus 17 Regional plot- 17: 44040184-44852612. ..... 99

Figure 4.15: Locus 20 Regional plot – 20: 47511792-47914180. ..... 100

Figure 4.16: Chromosome 2 circos plot showing mapped genes by eQTL and chromatin interactions ..... 106

Figure 4.17: Chromosome 3 partial circos plot showing mapped genes by eQTL and chromatin interactions ..... 107

Figure 4.18: Chromosome 4 circos plot showing mapped genes by eQTL and chromatin interactions ..... 108

Figure 4.19: Chromosome 5 circos plot showing mapped genes by eQTL and chromatin interactions ..... 110

Figure 4.20: Chromosome 5 circos plot showing mapped genes by eQTL and chromatin interactions ..... 111

Figure 4.21: Partial circos plot showing locus 6 on chromosome 6 showing mapped genes by eQTL and chromatin interactions ..... 112

Figure 4.22: Chromosome 12 circos plot showing mapped genes by eQTL and chromatin interactions ..... 113

Figure 4.23: Chromosome 12 circos plot showing mapped genes by eQTL and chromatin interactions. ..... 114

Figure 4.24: Chromosome 13 circos plot showing mapped genes by eQTL and chromatin interactions ..... 115

Figure 4.25: Chromosome 17 partial circos plot showing mapped genes by eQTL and chromatin interactions ..... 116

Figure 4.26: Chromosome 20 circos plot showing mapped genes by eQTL and chromatin interactions ..... 117

Figure 4.27: Q-Q plot of the gene-based test computed by MAGMA ..... 118

Figure 4.28: Venn diagram of overlapping mapped genes ..... 120

Figure 4.29: MAGMA expression analysis across all tissue types. ..... 123

Figure 4.30: Gene-tissue expression analysis based on GTEx RNA-seq data. ..... 126

Figure 4.31: Independent cell type associations based on within-dataset conditional analyses. ..... 127

Figure 4.32: Genetic correlation with cognitive traits, psychiatric and brain disorders, and brain imaging ..... 133

Figure 5.1: Venn diagram of overlap between mapped genes of *Resilience* and RT ..... 142

Figure 5.2: Prioritised genes for RT and their overlap with prioritised genes for *Resilien*ce 143

Figure 5.3: Comparison of single cell enrichment ..... 149

Figure 5.4: Manhattan plot of Resilience minus Locus 3 (3: 49385350-50250837) ..... 153

Figure 5.5: Summary per genomic risk loci ..... 154

Figure 5.6: Comparison of single cell type enrichment ..... 155

Figure 5.7; Manhattan plots of COGTOT and Recall ..... 158

Figure 5.8: Q-Q plots of (a) COGTOT and (b) Recall ..... 159

Supplementary Tables

The following Supplementary Tables are referenced in this thesis and can be accessed in the Supplementary Tables in my paper (Fitzgerald et al., 2021)

| Supplementary Table Number | Title |
|---|---|
| 4 | Candidate SNPs |
| 7 | Fine mapping |
| 9 | Gene mapping of candidate SNPs in Resilience |
| 10 | Genome wide gene association study (GWGAS) |
| 12 | Expression of mapped genes sets in GTEX Tissue sample |
| 13 | Cell type associations across datasets |
| 14 | Conditional analysis between datasets |
| 15 | Magma Gene-set Analysis |
| 19 | Genetic overlap of gene enrichment |

**List of abbreviations**

| Abbreviation | Long form |
|---|---|
| AA | African American |
| ACAD | Acyl-coa dehydrogenase family member |
| ACADVL | Acyl-coa dehydrogenase very long chain |
| AD | Alzheimer's disease |
| ADGRA | Adhesion G Protein-Coupled Receptor |
| AKAP | A-kinase anchor proteins |
| ALDH | Aldehyde dehydrogenase |
| ALS | Amyotrophic lateral sclerosis |
| AMP | Ampicillin resistance gene |
| AMS | Amazon Marketing Services |
| AMT | Aminomethyltransferase |
| ANNOVAR | Annotate variation |
| ANOVA | Analysis of variation |
| APG | Adenomatous Polyposis Coli |
| APOE | Apolipoprotein e |
| ARFGAP | ADP ribosylation factor GTPase activating protein |
| ARFGEF | Brefeldin a-inhibited guanine nucleotide-exchange protein |
| ARHGAP | Rho GTPase activating protein |
| ATNX | Atrophin |
| ATXN | Ataxin |
| AUS | Australia |
| BA | Brodmann area |
| BIG | Brain imaging genetics |
| BIP | Bipolar disorder |
| BIRTHYR | Birth year |
| BRAP | Brca1 associated protein |
| BSN | Bassoon presynaptic cytomatrix |
| CA | Cornu ammonis |
| CADD | Combined annotation dependent depletion |
| CAMKV | Cam kinase like vesicle associated |
| CCB | Centre for chromosome biology |
| CDHR | Cadherin related family member |
| CHARGE | Cohorts for heart and aging research in genomic epidemiology |
| CKAP | Cytoskeleton associated protein |
| CNV | Copy number variant |
| COGENT | Cognitive genomics consortium |
| COGTOT | Cognition total |
| CR | Cognitive reserve |
| CRHR | Corticotropin releasing hormone receptor |
| CSE | Chromosome segregation |

| Abbreviation | Long form |
|---|---|
| CTD | Coats disease |
| CTNNB | Catenin (Cadherin-Associated Protein), Beta |
| CUX | Cut like homeobox |
| CVD | Cardiovascular disease |
| CVH | Cardiovascular health |
| CYSTM | Sulphate transport system permease protein |
| DA | Dopamine |
| DAG | Dystroglycan |
| DB | Data base |
| DDN | Dendrin |
| DDX | Dead-box helicase |
| DLG | Discs Large MAGUK Scaffold Protein |
| DMN | Default mode network |
| DNA | Deoxyribonucleic acid |
| DNMT | DNA methyl transferases |
| DYNC | Dynein, Cytoplasmic, Heavy Polypeptide |
| EA | Educational attainment |
| EF | Executive function |
| ELSA | English longitudinal study of ageing |
| ENSG | Ensembl stable id |
| EUR | European |
| EY | Education years |
| FAM | Family with sequence similarity |
| FC | Functional connectivity |
| FDR | False discovery rate |
| FDXR | Ferredoxin reductase |
| FUMA | Functional mapping and annotation of genetic associations |
| FUNC | Function |
| FZD | Frizzled Class Receptor |
| GABA | Gamma-Aminobutyric Acid Type A Receptor |
| GBS | GWAS-by-subtraction |
| GCTA | Genome-wide complex trait analysis |
| GM | Grey matter |
| GNAT | G Protein Subunit Alpha Transducin |
| GNC | Galway neuroscience centre |
| GO | Gene Ontology |
| GPX | Glutathione Peroxidase |
| GREML | Genome-based restricted maximum likelihood |
| GRIN | Glutamate Ionotropic Receptor NMDA |
| GSE | Genomic Spatial Event database |
| GSK | Glycogen Synthase Kinase |
| GSMR | Generalized Summary statistics-based Mendelian Randomization |
| GTE | Genotype tissue expression |

| Abbreviation | Long form |
|---|---|
| GW | Gestation week |
| GWAS | Genome wide association study |
| GWGAS | Genome wide gene association study |
| HBCHO | Hindbrain cholinergic neuron |
| HBGLU | Hindbrain glutaminergic neuron |
| HBINH | Hindbrain inhibitory neuron |
| HC | Highly constrained |
| HDAC | Histone deacetylases |
| HEIDI | Heterogeneity in dependent instrument |
| HERC | HECT And RLD Domain |
| HOXB | Homeobox B Cluster |
| HRS | Health and retirement study |
| HWE | Hardy–Weinberg equilibrium |
| IBM | International Business Machines |
| ICV | Intracranial volume |
| ID | Identification |
| IMAGEN | Imaging and Genetics |
| IMRC | Immediate work recall |
| INT | Intelligence |
| IP | Inositol Hexakisphosphate |
| IQ | Intelligence quotient |
| IT | Information technology |
| KANSL | KAT8 Regulatory NSL Complex Subunit 1 |
| KMT | Lysine Methyltransferase |
| LBC | Lothian birth control |
| LD | Linkage disequilibrium |
| LDSC | LD score regression analysis |
| LDSR | LD score regression analysis |
| LI | Loss of function mutation intolerance |
| LNK | Lymphocyte-Specific Adapter Protein |
| LOAD | Late onset Alzheimer’s disease |
| LRN | Local regulatory networks |
| LRP | LDL Receptor Related Protein |
| LRRK | Leucine Rich Repeat Kinase |
| MAF | Minor allele frequency |
| MAGMA | Multi-marker analysis of GenoMic annotation |
| MAP | Rush memory and aging project |
| MAPK | MAP kinase |
| MAPKAPK | MAPK Activated Protein Kinase |
| MAPT | Microtubule Associated Protein Tau |
| MCI | Mild cognitive impairment |
| MDD | Major depressive disorder |
| METAP | Methionyl Aminopeptidase |

| Abbreviation | Long form |
|---|---|
| MHC | Major histocompatibility complex |
| MIR | MicroRNA |
| MMP | Matrix Metallopeptidase |
| MON1A | MON1 homolog A |
| MR | Magnetic resonance |
| MRI | Magnetic resonance imaging |
| MS | Microsoft |
| MST | Macrophage Stimulating |
| MTAG | Multi-trait analysis of GWAS |
| NA | Nor adrenaline/ Not applicable |
| NEIL | NeuroEnhancement for Independent Lives |
| NEU | Neuroticism |
| NICN | Nicolin |
| NICOG | Center for Neuroimaging, Cognition and Cognitive Genomics |
| NSF | N-Ethylmaleimide Sensitive Factor |
| NUIG | National University of Ireland Galway |
| OLA | Obg Like ATPase |
| OR | Odds ratio |
| PACSIN | Protein kinase C And casein kinase substrate in neurons |
| PARAM | Parameter |
| PATH | Personality and total health |
| PBRM | Polybromo |
| PCA | Principal component analysis |
| PET | Positron emission topography |
| PFDN | Prefoldin Subunit |
| PGC | Psychiatric Genomics Consortium |
| PGS | Polygenic score |
| PK | Parkinson’s disease |
| PLINK | Putty Link |
| PRKAG | Protein Kinase AMP-Activated Non-Catalytic Subunit Gamma |
| PRS | Polygenic risk score |
| QC | Quality control |
| QTL | Quantitative trait locus |
| RAND | Research and development |
| RBM | RNA Binding Motif |
| RECK | Reversion Inducing Cysteine Rich Protein With Kazal Motifs |
| RHOA | Ras Homolog Family Member A |
| RNA | Ribonucleic acid |
| RNF | Ring Finger Protein |
| ROS | Religious order study |
| RP | long intergenic non-protein coding RNA |
| RSID | Reference SNP cluster ID |
| RT | Reaction time |

| Abbreviation | Long form |
|---|---|
| SAMD | Sterile alpha motif domain |
| SCHLYRS | School years |
| SCZ | Schizophrenia |
| SD | Standard deviation |
| SE | Standard error |
| SEM | Structural equation modelling |
| SEMA | Semaphorin |
| SEMF | Semaphorin F |
| SERGEF | Secretion Regulating Guanine Nucleotide Exchange Factor |
| SH | Src Homology |
| SIFT | Sorting intolerant from tolerant |
| SNP | Single-nucleotide polymorph |
| SNV | Single-nucleotide variation |
| SPATS | Spermatogenesis Associated Serine |
| SPSS | Statistics for the social sciences |
| STAC | Scaffolding Theory of aging and cognition |
| STAU | Staufen double-stranded RNA binding protein |
| STK | Stroke |
| SYS | Saguenay Youth Study |
| TCTA | T Cell Leukemia Translocation Altered |
| TEGLU | Excitatory neurons, hippocampus CA1 |
| TET | Ten-eleven translocation |
| TICS | Telephone interview for cognitive status |
| TMEM | Transmembrane protein |
| TOMM | Translocase of 0uter mitochondrial membrane |
| TNF | Tumor necrosis factor |
| TRAF | TNF receptor associated factor |
| TRAIP | TRAF interacting protein |
| UBA | Ubiquitin-associated domain |
| UCSC | University of California Santa Cruz |
| UK | United Kingdom |
| UKB | UK Biobank |
| URL | Uniform Resource Locator |
| US | United States |
| UTR | Untranslated region |
| VGLUT | Vesicular glutamate transporter |
| WHO | World Health Organisation |
| WM | White matter |
| WNT | Wingless-Type MMTV |
| YOE | Years of education |
| ZNF | Zinc Finger Protein |

**Abstract**

Cognitive decline is one of the most feared aspects of ageing leading to major health and social issues. Non-pathological or age-related cognitive decline leads to increased challenges in completing tasks that require information processing and memory, which in turn leads to a deleterious effect on an individual's enjoyment of and participation in life events. Cognitive resilience is our ability to withstand negative effects of stress and maintain cognitive functioning. Understanding the factors that contribute to resilience is becoming increasingly important given the ageing demographics of the world's population. There is a growing knowledge of how non-genetic factors such as cardiovascular health and social participation contribute to cognitive resilience; however, an understanding of the genetic contribution has been hampered by the lack of large datasets with genetic data and suitable longitudinal data on cognition. Chapter 1 explores the current understanding of cognitive genetics leading to our current knowledge of what constitutes cognitive resilience.

In chapter 2, I discuss the various bioinformatic tools and methods employed to create a cogntive resilience phenotype and to explore genetic variation associated with cognitive resilience within the UK Biobank (UKB). In Chapter 3, I discuss how in the absence of direct measurements of cognitive ability at distal timepoints we employed proxy phenotypes. We used number of years in education (education years (EY)) as a proxy phenotype for cognitive performance in early adulthood, following several previous studies. Current cognitive performance was determined based on reaction time (RT) as a measure of processing speed. This approach captured an average time span of 40 years between past and current cognitive performance in 330,097 individuals. A confounding factor in my analysis is that EY is highly polygenic and masked the genetics of resilience. To overcome this, I employed Genomics Structural Equation Modelling (GenomicSEM) to perform a GWAS-by-subtraction using two GWAS, one GWAS of EY and resilience and a second GWAS of EY but not resilience. Subtracting one from the other generated a GWAS of Resilience. Replication of this approach was shown using independent discovery and replication samples within UKB.

Chapter 4 outlines the results of functional analysis on the full UKB GWAS which show significant genetic correlation with a GWAS of cognitive change in the independent Health and Retirement Study (N=9,526; P=1.5x10-3). We found 13 independent genetic loci for Resilience. Functional analyses showed enrichment in several brain regions and involvement of specific cell types, including GABAergic neurons (P=6.59x10-8) and glutamatergic neurons (P=6.98x10-6) in the cortex. Gene-set analyses implicated the biological process

“neuron differentiation” (P=9.7x10-7) and the cellular component “synaptic part” (P=2.14x10-6). The cellular component “wnt signalosome” had a strong effect size (Beta=1.22, P=4.75x10-6). The role of Mendelian randomization analysis showed a causative effect of white matter volume on cognitive resilience.

In chapter 5, I discuss ad hoc testing to show that the genetic correlation between Resilience and RT is strong because this is an RT-based resilience phenotype. However, there are differences in the associated genes being detected. This phenotype enabled the identification of genetic differences between those individuals in the UKB who preserved or maintained their capability to process information and respond over a 40-year time period compared to individuals who showed diminishing processing speed. This chapter also explores the effect of the gene rich locus on chromosome 3 showing that is does not unduly influence the functional analysis. Ad hoc testing also shows limited overlap with a GWAS of declining cognitive ability in the Health and Retirement Study.

The discussion in chapter 6 summaries the findings of this thesis and highlights that this research is the first of its kind to explore the genetics of cognitive resilience in large data and opens the way for future investigations in the area to enhance the neurobiological understanding of resilience. It also proposes ways to advance the knowledge of the genetics of cognitive resilience going forward with the ultimate goal of discovering interventions and therapeutic compounds that will combat cognitive decline and improve quality of life for an ageing population.

# 1 Introduction

(Note this Introduction uses material gathered for my MSc in Clinical Neuroscience, a recent literature review (Fitzgerald, Morris, & Donohoe, 2020) and extracts from published research where I was the lead author (Fitzgerald et al., 2021) )

Improved life expectancy and declining birth rates has led to an increasing percentage of the population that is greater than 60 years of age. The average age of the world population is increasing and according to the WHO, is expected to increase overall from 12% to 22% by 2050 and will be up to 30% in the more developed countries (World Health, 2015) (Figure 1.1). The human ageing process has evolved, as in other species, to stabilize populations and ecosystems to mitigate for resource restrictions and predation (Mitteldorf & Sagan, 2016). It is an inherent part of our life cycle and understanding ageing processes is essential to implement policies that promote healthy ageing.

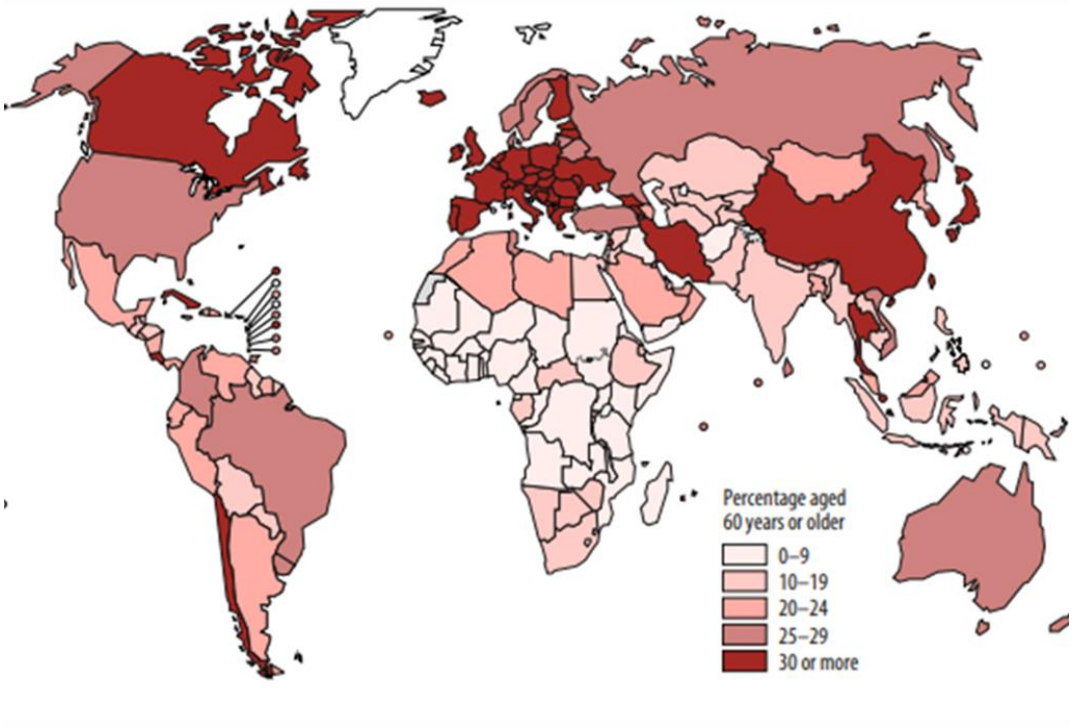



**Figure 1.1: Projected proportion of the population over 60 years of age in 2050**
WHO – Report on ageing and health 2015 (World Health, 2015)

While overall physical health is a core component of the ageing process, cognitive health is essential for normal functioning. Age-related cognitive decline leads to increased challenges

in completing tasks that require information processing and memory which in turn leads to a deleterious effect on the degree to which an individual can enjoy and participate in life events (Andrews, Das, Cherbuin, Anstey, & Easteal, 2016). Cognitive decline is one of the most feared aspects of ageing leading to major health and social issues and is associated with illness, dementia and death (Deary et al., 2009). However, individual rates of cognitive decline differ and cannot be explained by advancing age alone. Factors such as education level, fitness, diet, genetics, and overall life style influence rates of cognitive decline (Daffner, 2010). The concept of cognitive resilience has recently emerged to explain this variation.

## 1.1 Age-related cognitive decline

There are several theories as to what constitutes overall intelligence. However, with regards to the measurement of cognitive decline and resilience the concepts of crystallized and fluid intelligence are often used. Crystallized intelligence refers to the ability to use knowledge that has been accumulated throughout life, whereas fluid intelligence refers to an individual's ability to deal with new situations and problems (Brown, 2016).

Examining the constructs of crystallized and fluid intelligence shows that crystallized intelligence remains stable over the life span and can show a gradual improvement of 0.02 to 0.003 standard deviations (SD) per year in our sixties and seventies as it is based on the accumulation of life experiences. Conversely, fluid intelligence peaks in our thirties and declines at a rate of -0.02 SD per year (Harada, Natelson Love, & Triebel, 2013). Executive functioning, which is our capacity to direct our behaviour in a purposive, independent, appropriate and goal-related manner, declines with age, especially after the age of seventy and is affected by decline in speed of processing (Harada et al., 2013).

Prospective memory, which is remembering to perform an agreed task in the future, is generally poorer in older people under laboratory settings. However, in practice older people are often more reliable than their younger counterparts and it is proposed that this is due to an awareness of their limitations and the use of compensatory mechanisms. Furthermore, more recent studies have shown that older people perform better in the setting and regulation of goals which enhances prospective memory despite declining episodic memory (Brown, 2016).

Implicit memory, which is our automatic ability to perform a task such as riding a bike, is task dependent. New implicit memory learning is complicated by declining motor skills in older adults, but existing implicit memories are generally well preserved (Schacter, 2019). Attention is also affected by age, however, simple auditory attention span is only slightly affected whereas selective and divided attention demonstrate a more notable age effect (Harada et al., 2013).

### 1.1.1 Genetic and cellular processes involved in cognitive decline

Normal cognitive ageing is a process that happens during healthy ageing and is a result of the inability of the body to counteract various stressors, such as detoxification of free radicals or oxidative stress with resultant cellular damage (Daffner, 2010). Abnormal cognitive decline results from pathological processes such as tauopathies. It has been shown that decline seen in Alzheimer's disease (AD) is not an acceleration of the healthy ageing process but has a unique pathology of its own (Toepper, 2017) (Figure 1.2) .

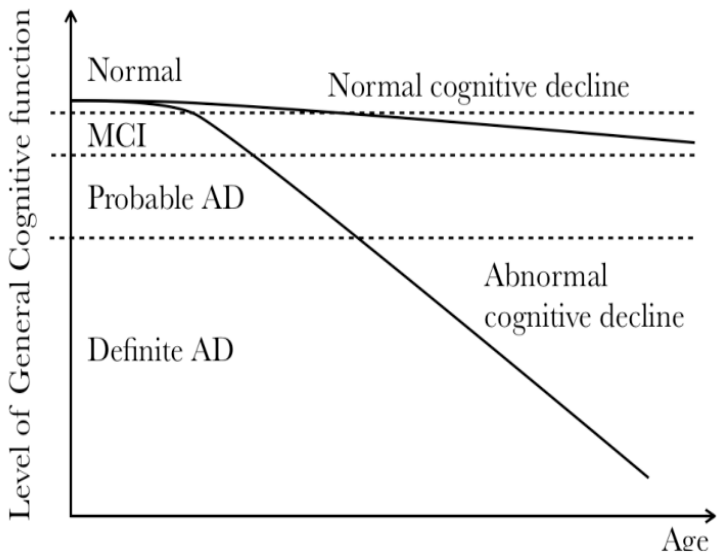



**Figure 1.2: Change in general cognitive function** Diagram shows the normal progression through age in contrast to the abnormal process where normal function is followed by mild cognitive impairment (MCI) and Alzheimer's disease (AD) (Gupta, Fua, Pautler, & Farber, 2013; Petersen et al., 2001)

However, the accumulation of altered proteins in the body that is associated with age-related pathologies such as Alzheimer's disease (AD) is also associated with non-pathological cognitive ageing. Normally these proteins are degraded by cellular proteases but during ageing there is either decreased elimination or increased production (Hipkiss, 2006; Nilsson & Tarnopolsky, 2019). These altered proteins may arise due to inaccurate synthesis by mitochondrial and cytoplasmic ribosomes, or damage inflicted by reactive nitrogen and oxygen species, unstable amino acid residues that spontaneously racemise, isomerise or deamidate and glycation (Hipkiss, 2006). This leads to apoptosis due to the build-up of breakdown products caused by activity of superoxide and hydrogen peroxide, which are generated in the mitochondria. As neuronal cells are not readily turned over or replaced, neurodegeneration occurs (Nilsson & Tarnopolsky, 2019).

Recently, proteome-wide association studies of cognitive decline found almost 600 proteins associated with cognitive trajectory, with neuronal mitochondrial activity, increased synaptic activity and decreased apoptosis and inflammation each found to be associated with cognitive resilience (Wingo et al., 2019).

Altered dopamine (DA) levels are associated with many neurodegenerative disorders and are also linked with non-pathological ageing. Various pharmaceuticals developed to control dopamine production are known to affect memory, for example, bromocriptine improves spatial working memory whereas haloperidol, which is a dopamine antagonist has the opposite effect (Baddeley, Eysenck, & Anderson, 2014). A meta-analysis shows that the reductions across the DA system are linked to a decrease in levels of DA transporters and receptors with age but the ability to synthesise DA is not affected. The average reduction in dopamine levels is 3.7%-14.0% per decade (Karrer, Josef, Mata, Morris, & Samanez-Larkin, 2017). The relationship of DA transporters to dopamine levels in age and cognition was studied using positron emission topography (PET), where a radioligand used to measure binding of DA to the putamen and caudate showed clear age-related losses and these were associated with a deterioration in executive function and episodic memory (Erixon-Lindroth et al., 2005). As well as mediating the effects of cognitive decline, DA was also shown to effect cognitive function where individual differences in binding, independent of age were shown to be associated with crystallized intelligence. However, a more recent PET study was unable to confirm the causal effect of dopamine receptor loss and age-related cognitive decline and cautions the interpreting of PET findings (Juarez et al., 2019).

Another biological pathway that has been implicated in cognitive change is noradrenergic signalling (for further details see Section 1.3.1.1).

Another biological factor that is increasingly associated with age-related cognitive decline is remodelling of the immune system. Over time, the efficacy of the immune system decreases leading to a susceptibility to inflammatory diseases, reduced vaccination response and increased vulnerability to infections (Aiello et al., 2019). Activated microglia which are the immune cells in the brain, produce the required inflammatory response. However, there is preclinical and clinical evidence that in ageing and diseased states the microglia, by genetic predisposition or by increasing sensitivity over time to previous pathology, produce a chronic response with multiple neurotoxic consequences leading to neurodegeneration and loss in brain function (Cunningham, 2013). A review of this relationship shows that chronic inflammatory conditions or immunosenscence, increases the susceptibility of older people to infection and increases vascular ageing causing neuroinflammation and cognitive decline (Tangestani Fard & Stough, 2019).

In addition, there is an association between replicative senescence, chronic inflammation, and telomere length. Telomeres are heterochromatic repeat regions at the ends of chromosomes, and the length of these regions is thought to be a biomarker for biological ageing (Zhang et al., 2016). The inflammatory response and oxidation are thought to accelerate the shortening of telomeres leading to eventual senescence and accelerated ageing. Leucocyte telomere length has been associated with cognitive capability and is a potential biomarker of cognitive ageing (Linghui et al., 2020), however, other data based on longitudinal studies did not show strong evidence for an association between decline in cognitive functioning and telomere length (Zhan et al., 2018). More work is needed to ascertain if telomere length is associated with cognitive decline (see Figure 1.3).

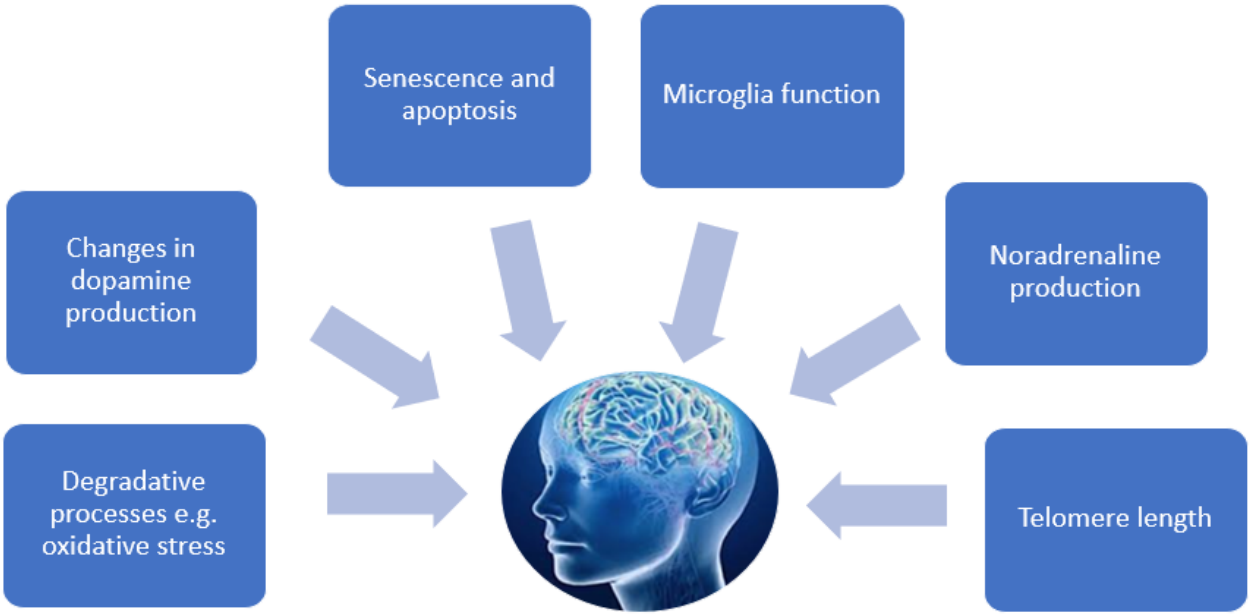



**Figure 1.3: Biological theories of cognitive decline**
Challenges to brain functioning leading to cognitive decline.

### 1.1.2 Epigenetic factors influencing cognitive decline

Alteration in lifestyle and environment modulates the effects of cognitive decline. The understanding of the epigenetic mechanisms triggered by these modulators through longitudinal studies is critical to our overall understanding of cognitive decline (Papenberg, Lindenberger, & Bäckman, 2015). Dysregulation of epigenetic mechanisms such as DNA methylation, microRNA-mediated gene regulation, nucleosome remodelling and changes in post-translational histone modifications have been associated with cognitive ageing and influence most of the brain functions including synaptic plasticity, memory and learning (Harman & Martín, 2020).

Master switches (molecules that that drive cognitive change) were identified in translational studies that control epigenome regulating gene expression involved in cognitive ageing. It appears that epigenetic modification is decreased in DNA methyl transferases 1 (DNMT1) and increased in histone deacetylases 2 (HDAC2) during the ageing process (Konar, Singh, & Thakur, 2016) (Figure 1.4).

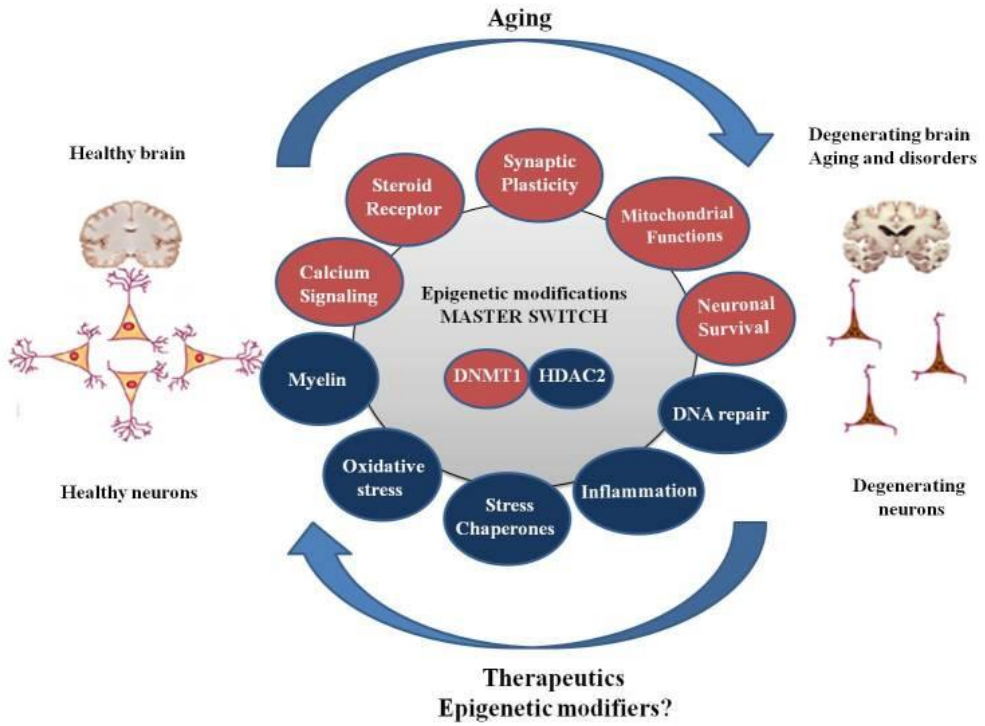



**Figure 1.4: Gene expression changes and master switch in age related cognitive decline**
Brain ageing accompanies alteration in expression (red circles represent downregulation; blue circles represent upregulation) of genes belonging to multiple pathways. Epigenetic modifications, particularly decrease in DNMT1 and increase in HDAC2 level, might be master regulators and accordingly epigenetic modifiers might prove ideal therapeutic targets (Konar et al., 2016).

### 1.1.3 Psychological factors affecting cognitive decline

A systematic review of relationships between cognitive decline and personality traits showed a consistent negative affect of neuroticism on cognitive performance in older adults and a positive relationship with conscientiousness. No other significant relationships were found with other personality traits (Koller, Hill, Mogle, & Bhang, 2019). It is possible that the relationship with conscientiousness is related to living a healthier lifestyle. Openness to new experiences leading to a diverse range of activities has a positive effect on cognitive ability during ageing (Jackson, Hill, Payne, Parisi, & Stine-Morrow, 2020). Vulnerability to stress is a trait of neuroticism and has been linked to poor cognitive performance in older adults (Manning, Chan, & Steffens, 2017).

Social interactions moderate cognitive decline. Social isolation has a negative effect on cognitive function (Evans et al., 2018). Loneliness, which is different from social isolation such that it is an emotional rather than a physical state, is also associated with cognitive decline (Levitin, 2020). Loneliness has been found to be a heritable trait and a poly genic risk

score for neuroticism was predictive of loneliness (Abdellaoui et al., 2018). Social interactions and educational attainment are associated with the trait of emotional intelligence. Both these parameters mediate the effects of declining emotional intelligence in older adults (Cabello, Navarro Bravo, Latorre, & Fernandez-Berrocal, 2014) . Both social isolation and loneliness are associated with reduced glutamate in the brain which is important for signal transmission (Levitin, 2020; Shao et al., 2015). Having a large, diverse and stimulating social network is associated with better cognitive function (Litwin & Stoeckel, 2015).

### 1.1.4 Environmental factors associated with cognitive decline

Educational attainment (EA) has a strong association with cognitive function and much of this is driven by a high correlation between EA and intelligence (Deary, Penke, & Johnson, 2010). However, environmental factors such as childhood health and socioeconomic parameters also affect EA independent of intelligence and have an association with cognitive outcomes in later life (Kobayashi et al., 2017).

Sleep is restorative and recent research has uncovered the repair processes and memory consolidations that happen during sleep. Quality and sleep patterns are associated with cognitive function in older adults. Long sleep latency (which is the time it takes to fall asleep) is associated with cognitive decline in healthy older adults, whereas long sleep duration and early sleep times are associated with normal cognitive function (Suh et al., 2018). The relationship of sleep duration with overall health is U shaped, in that significantly less than or greater than 8 hours sleep is associated with poorer performance. Average sleep period of less than six hours and greater than nine hours have been associated with poorer health outcomes (Fang et al., 2012; Levitin, 2020).

Cardiovascular health (CVH) is important to cognitive function and good midlife CVH is associated with preserved cognitive function in later life (González et al., 2018). Factors that negatively affect CVH such as obesity, diabetes and high blood pressure are also associated with cognitive decline (Leritz, McGlinchey, Kellison, Rudolph, & Milberg, 2011) . Cardiovascular disease (CVD) impairs the regulation of cerebral blood flow which results in reduced oxygen and nutrient supply to the brain, effecting neuronal processes (Vanherle, Matuskova, Don-Doncow, Uhl, & Meissner, 2020). Diet and exercise are therefore important to neurocognitive function and there is evidence to show that reversing the obesity trend has a

positive effect on cognitive health (Stillman, Weinstein, Marsland, Gianaros, & Erickson, 2017).

## 1.2 Theories of cognitive ageing

### 1.2.1 Processing speed and cognitive decline

Because processing speed is one of the strongest predictors of performance across cognitive tasks in older adults (Salthouse, 1996; Salthouse & Ferrer-Caja, 2003), it is the foundation of the reduced speed of processing hypothesis to explain decline in fluid cognitive processes (Salthouse, 1996). This theory proposes that older adults take longer to process information and the result of this slower processing leads to impairment in cognitive functions and information is not available for the next part of a task as quickly as with younger adults. It is proposed that superior intelligence is linked to faster processing speed and speed of higher order information processing explains about 80% of variance in cognitive ability (Schubert, Nunez, Hagemann, & Vandekerckhove, 2019). In a study using 1,800 adults ranging in age from 20 to 90 it was found that 70 to 80% of decline in processing speed was shared with declining reasoning ability (Scheiber, Chen, Kaufman, & Weiss, 2017).

Decline in processing speed had been found to be associated with cerebral small vessel disease and factors involved in the maintenance of cerebellar morphology (Eckert, Keren, Roberts, Calhoun, & Harris, 2010). In addition, better cognitive processing speed is associated with larger cerebral cortex volumes, lower levels of inflammatory markers and insulin and is mediated by physical exercise (Bott et al., 2017).

Frontal lobe and cerebellar grey matter volume predict variations in processing speed and research points to specific neural networks that undergo decline during ageing (Eckert, 2011). A systematic review of intra-individual variability during longitudinal assessment of processing speed as measured by reaction time and age-related cognitive decline have shown that poorer neuroanatomical integrity and greater behavioural variability are associated with lower white matter volumes and increased white matter hyperintensities. This could be explained by age-related dopamine reduction (Haynes, Bauermeister, & Bunce, 2017). Demyelination of white matter tracts is associated with ageing and it has been found that higher myelin content of white matter tracts results in faster processing speed (Chopra et al., 2018).

### 1.2.2 Other theories of cognitive decline

Another theory relates to working memory, where deficits are thought to be caused by the inability to deal with interference and it is proposed that the capacity to divide attention decreases with age (Baddeley et al., 2014). Our ability to suppress irrelevant information decreases as we age, resulting in impaired working memory performance (Samrani, Bäckman, & Persson, 2017).

In the 'less wiring more firing' hypotheses, it is proposed that a decrease in connectivity of certain networks is compensated for by increased activity in the remaining neurons (Daselaar et al., 2015). Rodent studies have shown that as the number of afferent neurons decrease, the synapses of the remaining neurons show higher synaptic potentials. This was also found in humans using event-related fMRI and diffusion weighted MRI where low executive function was linked to decreased white matter and more firing in the prefrontal cortex when subjected to a task and those low on memory scores showed the same results in the medial temporal lobes (Daselaar et al., 2015).

Compensation theory (Grady, 2012) proposes that older adults recruit more areas in the brain when performing cognitive tasks when compared to younger adults (attempted compensation). In some cases, this results in increased performance (successful compensation) and in others, has a negative affect (unsuccessful/ partial compensation). Grady has used fMRI studies to support her hypothesis.

The scaffolding theory of ageing and cognition (STAC) was proposed by Park and Reuter-Lorenz in 2008 (Park & Reuter-Lorenz, 2008) which argues that increased frontal activation in the ageing brain shows that the brain is adapting to declining neural structure and function through compensatory scaffolding to protect cognitive function.

## 1.3 Cognitive resilience/reserve

In the research literature the terms cognitive reserve is often used interchangeably with the term cognitive resilience, but as cognitive reserve is also used to describe one component of resilience, I will use the term cognitive resilience unless talking specifically about the subcomponent of cognitive reserve.

Cognitive resilience is the ability to withstand negative effects of stress on cognitive functioning (Staal, Bolton, Yaroush, & Bourne Jr, 2008). Cognitive reserve is the innate and

acquired buffer that dictates resilience. Recent theories in the understanding the individual differences in cognitive reserve have proposed that reserve has two components – brain reserve, which can loosely be described as hardware reserve or the structural properties of the brain and cognitive reserve which is the software or implementation of cognitive processes (Stern, 2012).

Individual resilience was first described by Katzman et al. in 1988 when examining some *post mortem* samples of brains that contained lesions associated with dementia, however, the subject did not display cognitive impairment before death (Katzman et al., 1988). Cognitive resilience is described as the buffer a person has against the effects of cognitive decline. This buffer has a strong genetic component and there is also evidence to show that this reserve can be altered by number of factors associated with a healthy lifestyle, such as aerobic exercise (Daffner, 2010). Structural brain alterations have been documented in rodent models where an exercise training regime resulted in increased hippocampus size and promotes neurogenesis through enhancing proliferation and survival of neurons in the dental gyrus to sustain a healthy brain (Vecchio et al., 2018). Exercise and physical activity is now understood to alter the human epigenome which can lead to enhanced cognitive health and overall health, thus improving quality of life in older adults (Rea, 2017). There is also evidence that other activities, such as meditation, reduce cognitive stress (Chan, Deng, Wu, & Yan, 2019) and it is proposed that mindfulness meditation may enhance cognitive reserve through activation of attentional function and indirectly by lowering stress and improving immune function (Malinowski & Shalamanova, 2017).

### 1.3.1 Theories of cognitive reserve

#### 1.3.1.1 Noradrenaline theory of cognitive resilience

Robertson proposed a theory of cognitive resilience involving the noradrenergic system where noradrenaline (norepinephrine) has a neuroprotective effect on brain function. Cognitive resilience is enhanced by environmental and psychological factors, and this stimulated noradrenaline which preserves working memory, which in turn stimulated factors such as attention, arousal, awareness and novelty which in turn increases cognitive resilience (Robertson, 2013, 2014) (see Figure 1.5). A recent review of research into the of the role of noradrenaline in cognition shows that loss of noradrenergic projections to the forebrain from the locus coeruleus is common in cognitive disorders (Holland, Robbins, & Rowe, 2021).

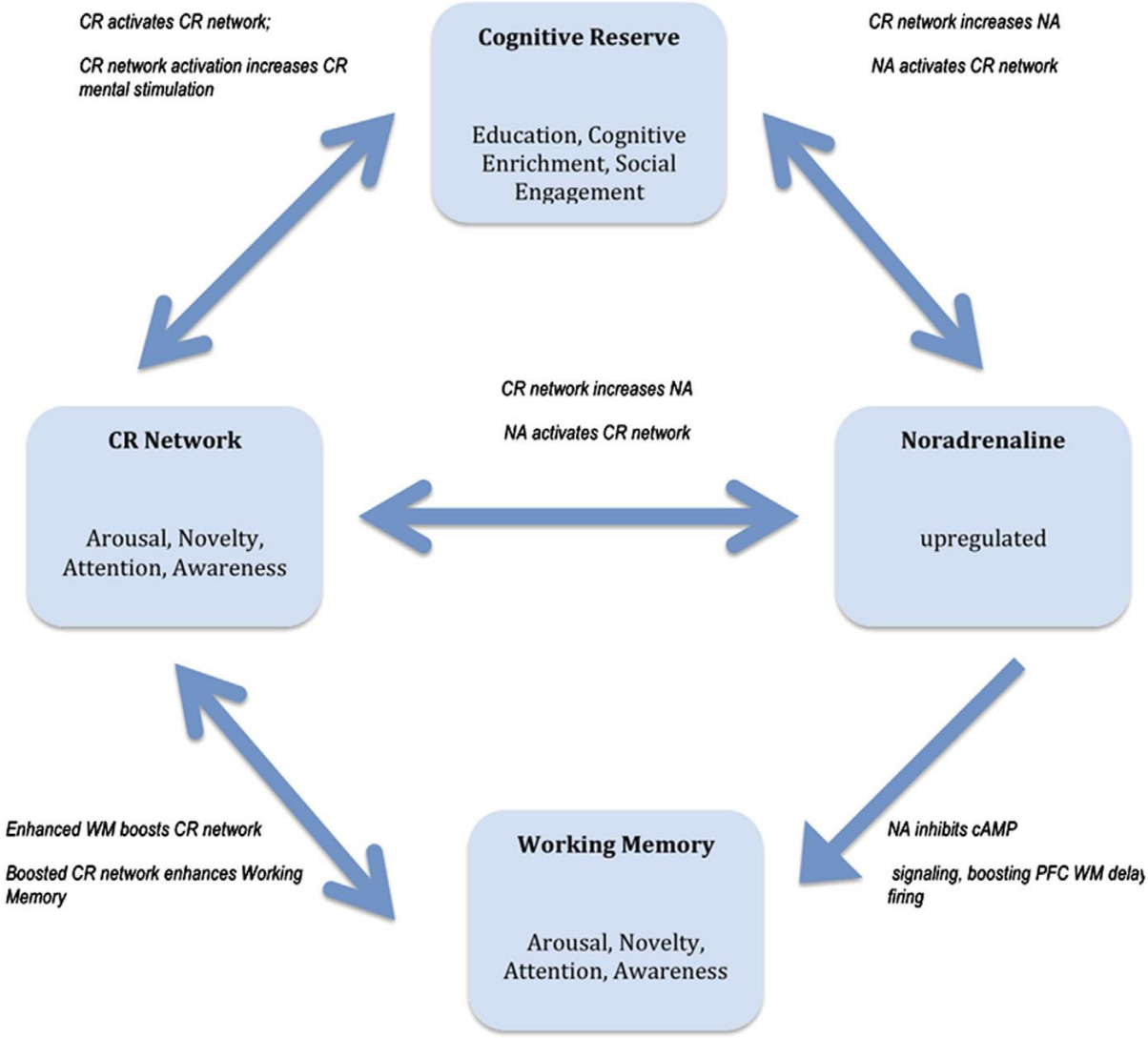



**Figure 1.5: A hypothetical cognitive reserve network**
The figure outlines the role of a cluster of networks, which hypothetically mediate between cognitive reserve variables such as education level on the one hand, and relatively protected cognitive function in later life on the other. The top right quadrant of the figure shows the CR-NA relationship. The top left quadrant of the figure illustrates the hypothetical relationship between CR and the hypothetical CR network. Education for instance should increase curiosity and hence exposure to novelty. Mental stimulation and social engagement will increase arousal, and with it, attention, and awareness. The bottom right quadrant shows the enhancement of working memory by NA inhibition of cyclic adenosine monophosphate (cAMP) signalling. Working memory enhancement may increase NA activity, but there is no direct evidence for this, hence this is the only unidirectional relationship. Finally, the bottom left quadrant illustrates the influence of enhanced WM on elements of the CR network, particularly attention and awareness. Abbreviations: cAMP, cyclic adenosine monophosphate; NA, noradrenergic; WM, white matter. (Robertson, 2014)

#### 1.3.1.2 Theory of brain and cognitive reserve

##### 1.3.1.2.1 Brain reserve

Brain reserve is seen as a passive component made up of brain volume and structural components and synaptic processes and as these decrease over time individuals with more brain reserve will express symptoms of impairment slower than those having less reserve (Medaglia, Pasqualetti, Hamilton, Thompson-Schill, & Bassett, 2017). Maximal brain volume using intracranial cavity measurements remains static in an individual and is associated with cognitive ability in later life and is proposed as a proxy for brain reserve (Adams et al., 2016; Royle et al., 2013). It is proposed that a structurally larger brain can tolerate more pathology prior to demonstrating cognitive decline (Stern, 2012). Cognitive changes in ageing are associated with changes in grey matter volume particularly in the temporal lobes (Fletcher et al., 2018). A meta-analysis of magnetic resonance imaging (MRI) shows an annual decline of 0.5% in whole brain volume at 60 years of age and this rate increases after 60 (Hedman, van Haren, Schnack, Kahn, & Hulshoff Pol, 2012). Diffusion weighted MRI has shown a decrease in white matter due to breakdown of the structural integrity of myelin in neurons of ageing brains. This results in a decline in communication in cortical networks and involves executive function, memory and perceptual speed (Madden et al., 2012). The effect of brain atrophy over time on an individual is related to their cognitive reserve and maintenance of this reserve mediates this effect (Bettcher et al., 2019). The static measurement of intracranial volume (ICV) has been found to be highly heritable and associated with cognitive function and supports ICV as a biomarker for brain reserve (Adams et al., 2016).

##### 1.3.1.2.2 Cognitive reserve

Cognitive reserve is seen as an active process involving the implementation and adaptation of cognitive processes. It measures the robustness of these processes against pathology and the ability to use alternative processes where necessary (Medaglia et al., 2017). Decreased connectivity has been found in ageing brains in multiple resting state networks, including the salience network, which directs our attention, and the default mode network (DMN). Resting state functional MRI (fMRI) could predict age and cognitive ability based on connectivity profiles, particularly those between the salience and visual networks and the salience and anterior part of the default mode network (DMN). Moreover, this connectivity was predictive of episodic memory and executive function performance (La Corte et al., 2016).

The threshold model of cognitive reserve (Figure 1.6) proposes that the combination of brain reserve and cognitive reserve result in different outcomes, such that and individual with low brain reserve and low cognitive reserve may exhibit symptoms of cognitive decline or impairment sooner that an individual with high brain and cognitive reserve (Medaglia et al., 2017; Stern, 2002) .

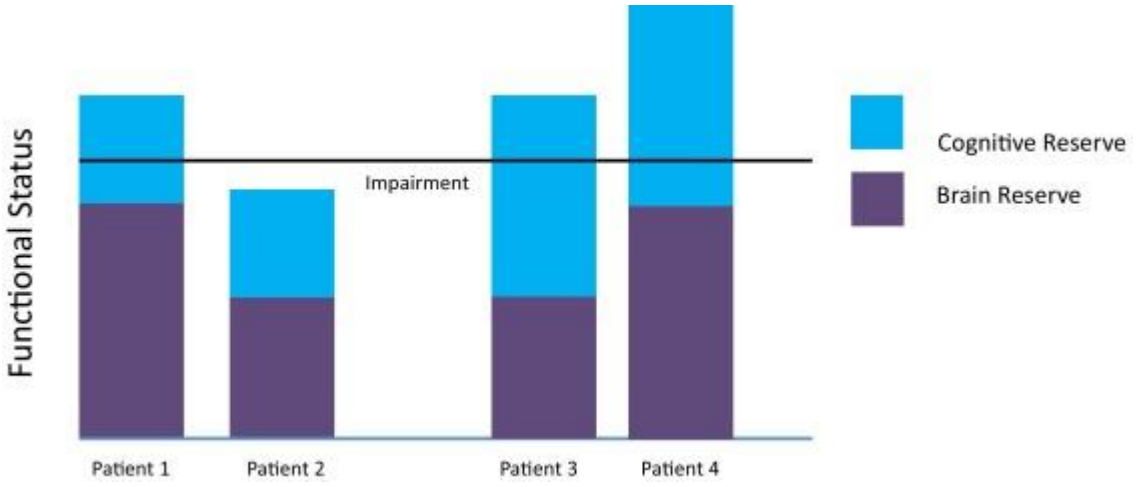



**Figure 1.6: Threshold model of cognitive reserve**
Brain and cognitive reserve are represented by measured quantities that cumulatively protect against disease. Patients with greater reserve remain above the impairment threshold following the onset of neuropathology. Patient 1 shows greater resilience to brain pathology than Patient 2 due to greater brain reserve with equivalent cognitive reserve. Patient 3 shows greater resilience to brain pathology than Patient 2 due to greater cognitive reserve with equivalent brain reserve. Patient 4 displays heightened neuroprotection due to the cumulative effects of (i) brain reserve equivalent in magnitude to that observed in Patient 1 and (ii) cognitive reserve equivalent in magnitude to that observed in Patient 3 (Medaglia et al., 2017).

However, others consider this model to be too simplistic in that resilience to cognitive decline is multifactorial and is a broader construct than cognitive and brain reserve. Other factors such as socio-emotional, physical, and spiritual components need to be considered. In addition, disease burden can overwhelm reserve and push it to a threshold or tipping point (Schwartz, Rapkin, & Healy, 2016) . Ongoing activities involving engagement in cognitive activities and general cognitive function are dynamic constructs and should be considered when measuring resilience (Malek-Ahmadi et al., 2017). Acquired reserve is composed of past and current reserve building activities and this is influenced by the personal characteristics of the individual as described by Schwartz et al., (Schwartz et al., 2016).

#### 1.3.1.3 Theory of cognitive reserve, brain reserve and maintenance

Yaakov Stern proposes a model of resilience of which there are three components, cognitive reserve, the passive component of brain reserve and a more active process of brain maintenance. This is based on recent research finding on neuroplasticity. Brain maintenance is the ability to maintain brain integrity through maintenance activities. In addition, cognitive reserve has two components, neural reserve and neural compensation (Stern, 2017; Stern et al., 2018). In a systematic review of super-agers (those over 80 with preserved episodic memory), preservation of the salience and default mode networks and strong functional connectivity were found. In addition, the anterior cingulate cortex was highlighted as a potential imaging biomarker for resilience. Brain maintenance, cognitive reserve, and brain reserve are complementary but independent.

#### 1.3.1.4 Theory of reserve, maintenance, and compensation

In a recent opinion piece by Cabeza and colleagues (Cabeza et al., 2018), a triad of the biological mechanisms of reserve, maintenance and compensation are proposed to control cognitive decline in healthy ageing. Individual differences in cognitive ageing are due to the genetic and environmental effects on these three components. Reserve is discussed in terms of brain reserve and cognitive reserve, and these are the resources that remain over and above that required for normal cognitive functioning and come into play when resources are depleted in later life. Maintenance off-sets neural decline by neural enhancement and is increasingly required as we age. Compensation is the addition of alternative neural resources when the demand is not met by existing processes (Figure 1.7).

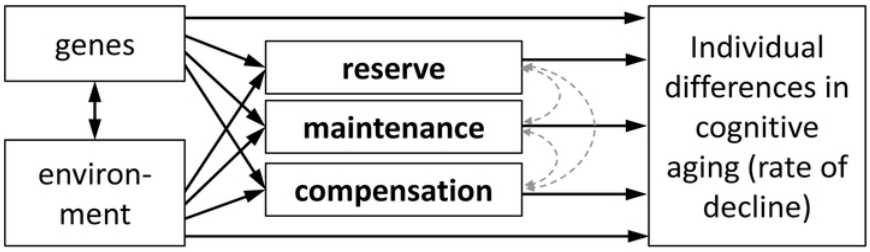



**Figure 1.7: Reserve, maintenance, and compensation**Individual differences in cognitive ageing have been attributed to the effects of three interacting mechanisms: reserve, maintenance, and compensation. These mechanisms are assumed to mediate some (but not all) of the effects of interacting genetic and environmental factors on cognitive ageing. (Cabeza et al., 2018).

### 1.3.2 Intelligence and resilience

Others propose that variation in the rate of cognitive decline can be explained by variation in intelligence. Longitudinal analysis in the Lothian Birth Cohort has shown that childhood intelligence has a protective effect on cognitive decline in later life (Cadar, Robitaille, Pattie, Deary, & Muniz-Terrera, 2020). Other studies show that while higher education reflects greater cognitive ability, the rates of change in that ability over time are consistent across all education levels, with those starting at a higher level simply having further to fall before they present with mild cognitive impairment (Guerra-Carrillo, Katovich, & Bunge, 2017; Lövdén, Fratiglioni, Glymour, Lindenberger, & Tucker-Drob, 2020) (Figure 1.8). The role of intelligence is confounded by the fact that higher intelligence is associated with healthier life styles, which has a protective effect on cognitive decline (Geary, 2019).

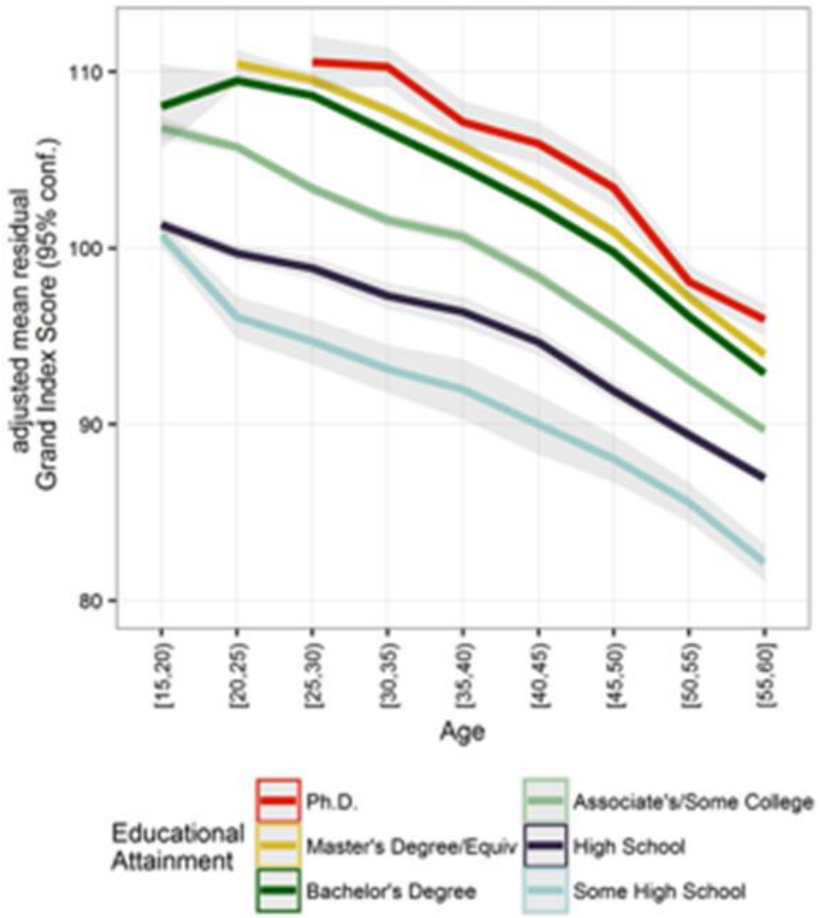



**Figure 1.8: Effect of education on the adjusted grand index score at baseline across ages (Guerra-Carrillo et al., 2017)**
Cognitive assessment results from Luminosity (n=196,288, age 15 -60) by educational attainment.

Given the multifactorial drivers that contribute to cognitive decline described in Figure 1-3, models that examine cognitive resilience will be complex. A summary of systematic reviews of strategies to operationalise cognitive reserve conclude that there is insufficient research to create a full cognitive reserve model (Harrison et al., 2015). The proxy measures of educational attainment and occupation have been shown to be associated with cognitive reserve. Socioeconomic status and premorbid IQ have shown inconsistent results, however, when combined with other factors including current cognitive activities they have been associated with reserve. Models have not included genetic indicators or diet and exercise.

### 1.3.3 Measuring cognitive resilience

To understand the complexity of phenotypic measurement of cognitive resilience, which requires repeated cognitive measures over long periods of time we must first examine the challenges in measuring cognition itself.

#### 1.3.3.1 Measuring cognition

‘General’ cognitive ability or ‘intelligence’ refers to our ability to reason, learn and solve problems and is measured based on performance on tests of processing speed, vocabulary size, abstract verbal and non-verbal reasoning, and visuospatial skills. These scores are aggregated to yield a general ability score or statistically reduced into a single factor or component referred to as Spearman’s ‘g’ (Lee et al., 2018; Plomin & von Stumm, 2018). Typically, a principal components analysis of individual subtests yields a single factor that explains ~50% of variance in measures used, reflecting the strong correlation usually observed between these cognitive tasks. Combining data from multiple sources shows that ‘g’ is a robust value, valid in both western and non-western countries (Deary et al., 2010; Lam, Hill, et al., 2019).

Notwithstanding the moderate correlations observed between many cognitive tasks, several measurement issues exist. These include low test/retest reliability for some aspects of cognition, (Johnson, Nijenhuis, & Bouchard, 2008; Warne & Burningham, 2019), a bewildering array of different measures of the same domain and even multiple versions of the same test, all of which complicates attempts to combine data from different groups to achieve the sample sizes required for genomic studies. Even where the same measures have been collected in very large population-based cohorts such as the UK Biobank (UKB), the use of

shorter cognitive tests within a larger battery of health relevant tasks have led to issues of task validity (De Schryver, Hughes, Rosseel, & De Houwer, 2016; Hedge, Powell, & Sumner, 2018).

Yet another issue of phenotypic complexity in large-scale studies relates to the use not of cognitive tests per se but to the use of proxy measures of cognitive ability. Given the need to combine different datasets to increase sample size to boost power for gene discovery and the lack of comprehensive cognitive data in these datasets, some readily available proxy phenotypes have been used including years of education (YOE) and educational attainment (EA). Based on samples of >70,000 English children, the correlation between EA and 'g' was observed to be 0.81 (Okbay et al., 2016).

Recent analyses in very large datasets have shown this correlation to be closer to 0.7 (Fawns-Ritchie & Deary, 2020). Cognition is defined as any measure of cognitive performance such as memory, processing speed, reasoning, acquisition of knowledge, attention, and executive function (Okbay et al., 2016). Measurement of intelligence through IQ testing has considerable similarity with these domains and the terms, cognitive function and intelligence tend to be used synonymously.

#### 1.3.3.2 Measuring cognitive resilience

Measuring cognitive resilience is a challenge in that it is only observed with a referent in that an individual's reserve or resilience is a measurement of a better-than-expected performance where the expected performance is based on prior knowledge (Schwartz et al., 2016). It cannot be measured at a single time point and given that the rate of cognitive decline in healthy ageing is slow, long periods of time are needed to determine differences in rates of change.

Both longitudinal and cross-sectional studies are used in research on cognitive change over time. In a longitudinal study, a cohort of people is tested at discrete intervals of a number of years over several decades, however, these studies present several problems in current research in that the acquisition of reliable data takes many years to accumulate and can be complicated by changes in cognitive measures and improvement in procedures. In addition, natural attrition, or loss due to the development of degenerative conditions, can decrease the

power of a study (Baddeley et al., 2014). Furthermore, practice effects through multiple testing of the same population may influence the results (Salthouse, 2015).

Cross-sectional design overcomes the problems with longitudinal studies in that different people are measured across age ranges and their performance is measured on unique occasions. The difficulty with this approach is that performance cannot be related to earlier or later data and therefore has limited use in the study of cognitive resilience. Another difficulty that affects both approaches is the cohort effect. This effect relates to substantial changes in lifestyles over the decades. A person currently in their twenties would, in general, have experienced better educational methods, better nutrition, and better health care than a current 80-year-old when they were the same age (Baddeley et al., 2014).

As cognitive decline is a slow process, repeated cognitive measures over several years are needed to access different rates in cognitive change over time. The difficulties in measuring cognition as explained above become even more complex. As cognitive measures have been improved and added to, trying to extrapolate different methodologies over time is difficult. In addition, when a cohort is measured at several intervals not all participants are available at each timepoint. Traditional statistical methods, such as analysis of variance (ANOVA) would eliminate participants with missing data thus curtailing the sample numbers further. To overcome this complex statistical method such as linear mixed modelling has been used in the limited studies on longitudinal data to date (Arpawong et al., 2017; Zhang & Pierce, 2014).

Most studies have measured cognition using the component Spearman's 'g' or one or more sub-components such as of processing speed, vocabulary size, abstract verbal and non-verbal reasoning, and visuospatial skills. Given that processing speed influences all other cognitive measures (see section 1.6.1) is it a reasonable measure to use on its own. In a study examining the link between academic achievement and cognition the authors propose a model where processing speed moderates academic achievement through its effects on other cognitive parameters. In this model, information processing speed is the key predictor of fluid intelligence, working memory, and number sense, which in turn contribute to individual differences in academic success (Tikhomirova, Malykh, & Malykh, 2020) (Figure 1.9).

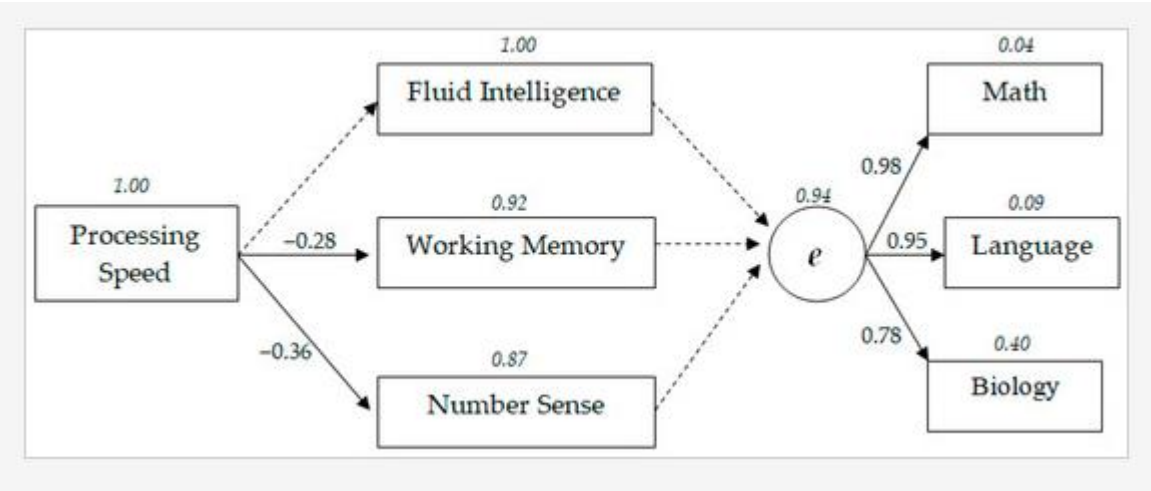



**Figure 1.9: Structural equation modelling (SEM) of the relationships between cognitive abilities and general academic achievement in high school education** "*e*" is the latent factor academic success. Dotted lines indicate nonsignificant relationships. (Tikhomirova et al., 2020)**.**

Given the multifactorial drivers that contribute to cognitive decline described in Figure 1-3, models that examine cognitive resilience will be complex. A summary of systematic reviews of strategies to operationalise cognitive reserve conclude that there is insufficient research to create a full cognitive reserve model (Harrison et al., 2015). The proxy measures of educational attainment and occupation have been shown to be associated with cognitive reserve. Socioeconomic status and premorbid IQ have shown inconsistent results, however, when combined with other factors including current cognitive activities, have been associated with reserve. Models have not included genetic indicators or diet and exercise.

Exploratory factor analysis arising for a longitudinal study of an ageing cohort in Tasmania led to the proposal of a two factors model comprising of current cognitive reserve (cCR) and prior cognitive reserve (pCR). pCR measures include prior intelligence, prior education, mental activities as young and middle-aged adults and midlife occupation whereas cCR measures included current measures of arithmetic ability, spelling ability and current IQ. The cCR factor structure was found to be longitudinally stable and had a positive association with further education (Ward, Summers, Saunders, & Vickers, 2015).

Others have used structural equation modelling to study cognitive resilience. Episodic memory was decomposed into three components, one predicted by a pathology measure derived from brain imaging, the second based on demographics and the third based on reserve (which is all the remaining variance). Given that the first two are known, the interaction of the latent variable of reserve with conditions such as mild cognitive impairment

can be examined. (Figure 1.10). This approach when used in a cross-sectional study shows that cognitive resilience mediates rates of cognitive decline towards impairment, rates of decline in executive function and rates of brain atrophy (Reed et al., 2011; Reed et al., 2010). Further research incorporating this model and using data from a longitudinal study found a great potential for its use with repeated measures to capture dynamic cognitive reserve (Zahodne et al., 2015). It was found that by incorporating function connectivity (FC) measures into the model, higher CR was associated with higher global efficiency, increased FC clustering and efficiency in the occipital lobes, CR was also associated with centrality and strength of the inferior temporal gyrus (Marques et al., 2016).

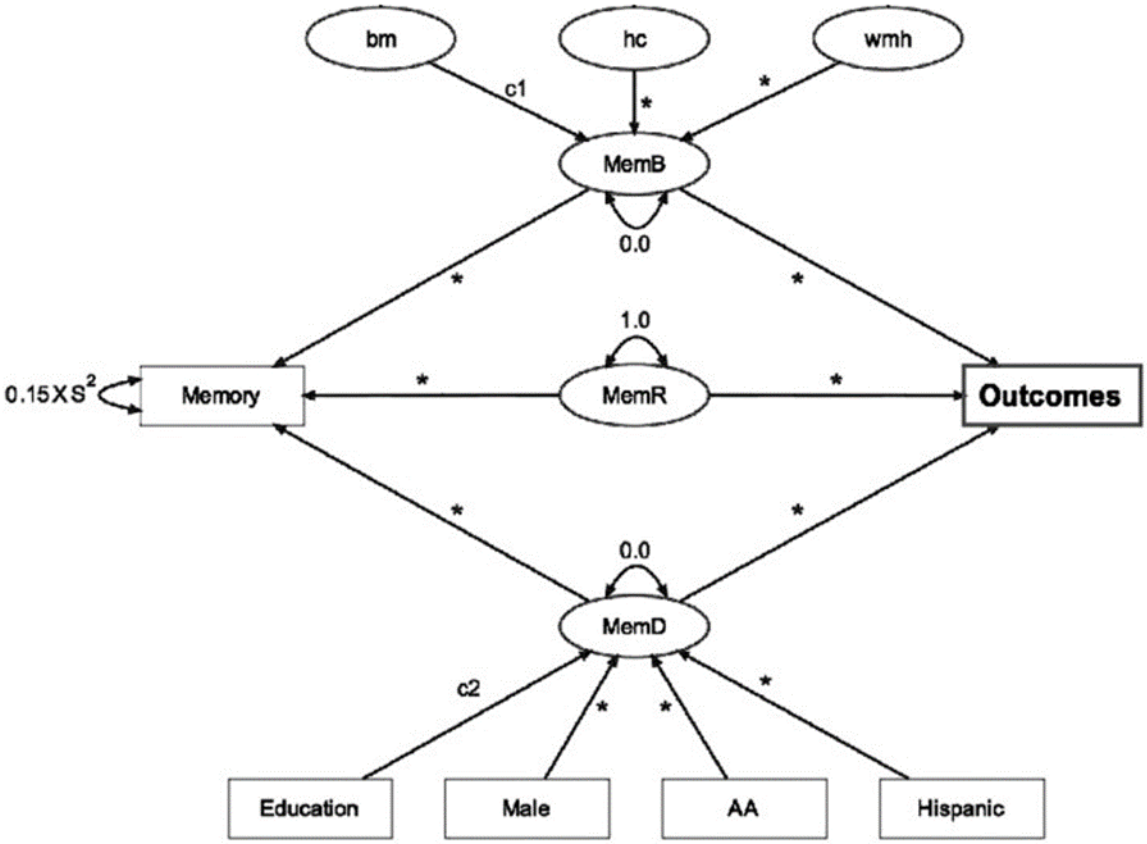



**Figure 1.10: Analytic model to decompose episodic memory in cognitive ageing**
Rectangles represent observed variables and ovals represent latent variables. Observed demographics, years of education, gender (female as reference) ethnicity (Caucasian as reference) African American (AA) and Hispanic and MRI variables of brain matter (bm), hippocampal volume (hc) and white matter hyperintensities (wmh) were allowed to correlate freely (paths not shown). Mem-B is a linear combination of the three MRI variables, with parameters representing regression coefficients of Mem-B on the three indicators. Mem-D is analogously related to the observed demographic variables, Mem-R is the component of episodic memory unrelated to demographic and and MRI variables. Freely estimated parameters are indicated by 'asterisk'. S2 refers to sample variance. c1 and c2 are scaling constants selected to set variances at 1.0 for the MemB and MemD latent variables (Reed et al., 2010).

### 1.3.4 Available datasets

#### 1.3.4.1 Datasets with longitudinal cognitive data

In the field of cognitive genetics, it has been acknowledged that the larger the dataset the better the outcome for genetic association studies. In recent years datasets with up to 1.3 million participants have been used in cognitive genomic studies (Lee et al., 2018). These studies are discussed further in section 1.5.1. Unfortunately, the generation of datasets large enough to explore cognitive genetics has only recently commenced and it will take time to accumulate data on cognitive change over time. There are very limited datasets available with genetic data on participants that have examined longitudinal measures in cognition in healthy ageing. Table 1.1 shows the studies that have been used to date. Findings from these studies are discussed in Section 1.4.2.2.

**Table 1.1: Longitudinal datasets used to study cognitive decline**

| Study | Acronym | Country | N | Date started | Waves |
|---|---|---|---|---|---|
| Health and retirement study | HRS | US | 20,000* | 1996 | 11 |
| English longitudinal study of ageing | ELSA | UK | 7,412 | 2002 | 8 |
| Personality and total health | PATH | AUS | 7,500 | 1999 | 4 |
| Religious order study | ROS | US | 750 | 1993 | 9 |
| Rush memory and aging project | MAP | US | 825 | 1997 | 9 |
| Lothian birth control 1921 | LBC1921 | Scotland | 550 | 1921 | 5** |
| Lothian birth control 1936 | LBC1936 | Scotland | 1,091 | 1936 | 5** |

*The participants are of different ethnicity (approximately 9,600 are Caucasian). ** Baseline testing was in the starting year but follow up waves of testing commenced in recent years. (n=number of participants with genotype data, Waves = number of times cogntive tests were repeated).

In the absence of longitudinal data, there is the possibly of incorporating proxy measures for past cognitive performance. The most common proxy measure for cognitive resilience is years of education but this static measure of cognitive reserve can be influenced by other

variables such as socioeconomic status that affect the risk of cognitive impairment through means other than affecting cognitive reserve (Reed et al., 2010).

## 1.4 Genetic component of cognitive resilience

To understand the genetics of cognitive resilience we first must examine the genetics of cognition itself.

### 1.4.1 Genetics of cognition

Long before the development of modern genomic methods, as far back as the early 1900's, the heritability of cognitive performance was recognized through twin and adoptive studies (Plomin & Deary, 2014). In a study of ~10,000 monozygotic and dizygotic twins, concordance in measures of intelligence was found to be 0.86 and 0.60 respectively (Plomin, 2001; Plomin & Spinath, 2004). Follow-up longitudinal twin research had further shown that heritability actually increases during childhood development; this is explained by genetic innovation in early childhood, whereby increasing numbers of genes become activated during cognitive development, thus amplifying the contribution of genetics over environment (Briley & Tucker-Drob, 2013). Given that the estimates of heritability of intelligence, estimated at 50% across the lifespan, it was originally assumed that it was only a matter of time until the key gene(s) involved in cognition were identified (Plomin & von Stumm, 2018; Ramus, 2006). However, the complex and highly polygenic nature of cognitive phenotypes is now well established, with literally hundreds of genes statistically associated with variation in cognitive function and implicating a wide variety of processes related to brain development and neuron to neuron communication (Lam, Hill, et al., 2019; Lee et al., 2018).

Early genetic studies of cognition focused on 'candidate' genes selected on the basis of their hypothesized biological importance to illness risk. However, a failure to replicate the findings from these studies, together with the emergence of genome-wide approaches to gene discovery in the past ten years have meant that a majority of recent discoveries in both cognitive and psychiatric genetics have come via genome-wide association studies (GWAS). A major initial challenge in adopting this approach was the limited sample sizes of available cohorts, which hindered identification of genome-wide significant results in early GWAS of cognitive phenotypes (Trampush et al., 2017). To boost power for genetic studies, several consortia were formed to pool sample resources to yield more significant outcomes. In 2015

the Cohorts for Heart and Ageing Research in Genomic Epidemiology (CHARGE) consortium combined data from 31 cohorts (n=53,949) and performed a meta-analysis of GWAS using a general cognitive factor derived from principal component analysis of several tests (Davies et al., 2015). This analysis identified three loci, on chromosomes 6, 14 and 19, as relevant to cognitive processes (Davies et al., 2015). A further analysis by the Cognitive Genomics Consortium (COGENT) combined 21 cohorts (n=35,298) and confirmed the findings of the CHARGE study as well as identifying two more significant loci on chromosomes 1 and 2 (Trampush et al., 2017). This study also compared the top SNPs from larger EA studies (n=164) and found 31 SNPs that were significantly associated with EA in other studies that were also nominally significant in this study; all had the same direction of effect showing a robust genetic correlation between EA and cognition.

The UKB project was initiated to generate a very large dataset based on the UK population where data was collected on over 500,000 people (Sudlow et al., 2015). Initially, genotypic data was released for ~150,000 individuals in May 2015 and was used in combination with existing data in a number of GWAS of cognition that confirmed previous findings and uncovered more associated loci (Davies et al., 2016; Hill et al., 2016; Okbay et al., 2016; Sniekers et al., 2017). The full dataset on >500,000 individuals was released in July 2017 and has proved a “game-changer” in GWAS of cognition function by facilitating studies with samples sizes of >100,000 individuals that have identified hundreds of independent associated loci. Study of the combined CHARGE, COGENT and UKB cognitive and genetic datasets (n=300,486 participants) have identified 146 genome-wide significant loci and 709 genes associated with general cognitive function (Davies et al., 2018). Associated genes show enriched expression in most brain regions with strongest signals in the cerebellum and cortex and *in silico* biological investigations of these genes points to processes such as neurogenesis, regulation of nervous system development and neuron differentiation being affected. A second study based on COGENT and UKB samples plus other samples (n= 267,867 participants) published around the same time, found a total of 205 loci (implicating 1,016 genes) to be associated with intelligence (Savage et al., 2018). Analysis of biological processes implicated by these associated genes found the pathways involving regulation of nervous system development, central nervous system, and neuron differentiation to be enriched for associated genes, plus regulation of synapse structure or activity was significantly enriched too. Beyond enriched expression of associated genes in multiple brain regions, single cell analysis identified the most enriched cell types for genes associated with

intelligence to be medium spiny neurons (striatum), CA1 pyramidal neurons (hippocampus) and pyramidal neurons (somatosensory cortex).

#### 1.4.1.1 Genome-wide association studies (GWAS) and cognition

In addition to data from publicly funded biobanks, commercial companies such as 23andMe, have also collaborated in cognitive genomic research (Eriksson et al., 2010). In the largest study to date on EA, Lee et al. combined data from 71 cohorts to yield a sample size of 1,131,881individuals, of which 365,538 samples were provided by 23andMe (Lee et al., 2018). This analysis identified 1,271 lead SNPs that were independently genome-wide significant, again demonstrating the positive correlation between sample sizes, and number of variants identified (see Figure 1.11). Lee et al. used multi-trait analysis of GWAS (MTAG), an approach that exploits the phenotypic and genetic correlations between different phenotypes (e.g., cognitive ones) to increase statistical power (Turley et al., 2018). By combining GWAS results from studies of EA, cognitive performance, and mathematical ability (for a total n=1,311,438), Lee et al were able to increase their number of genome-wide significant loci to 1,624 (n=1,311,438). Biological annotation analysis suggested that genes near to these SNPs are strongly enriched for expression in the central nervous system. These genes show elevated expression in the prenatal brain, where they are involved in many developmental processes, but also have high expression in the postnatal brain where genes were involved in nearly all levels of neuron-to-neuron communication and synaptic plasticity. Of note, while neurons were strongly enriched for EA-associated genes, astrocytes and oligodendrocytes were not, leading the authors to conclude that cognitive variation was not associated with genetic differences in myelin related axonal transmission speeds (Lee et al., 2018). This conclusion contrasted with findings from a MTAG study by Hill et al (Hill, Marioni, et al., 2018) that combined GWAS of intelligence (Sniekers et al., 2017) with EA (Okbay et al., 2016) (n=248,482) and identified 187 genetic loci associated with intelligence. Biological annotation analysis showed associated genes to be enriched in a number of processes including neurogenesis, synaptic plasticity, cell development and myelination, specifically oligodendrocyte differentiation (Hill, Marioni, et al., 2018). The disagreement between these two studies suggests a need for further studies to clarify whether the genetic architecture of cognition implicates white matter microstructure and oligodendrocytes function.

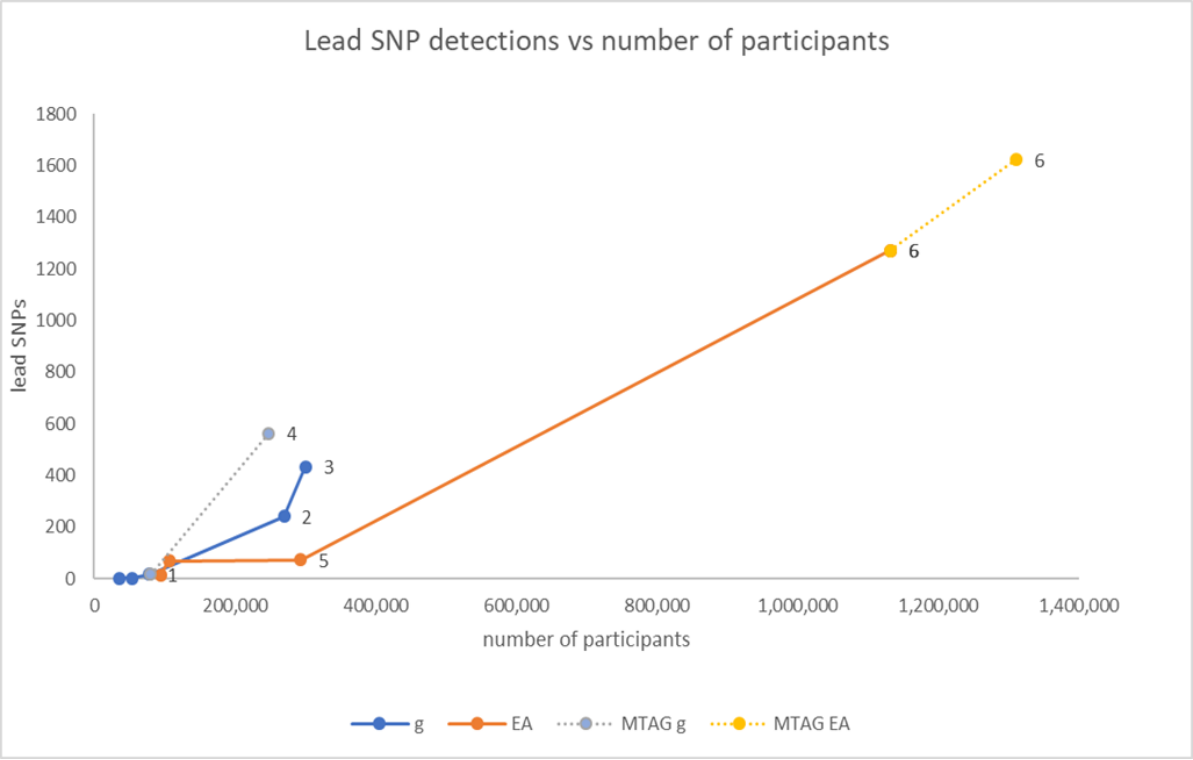



**Figure 1.11: Plot of lead SNPs from GWAS and MTAG of g and EA showing increases in significant findings with increasing participation**
The studies included for g are 2 lead SNPs (Trampush et al., 2017), 3 lead SNPs (Davies et al., 2015), (1) 18 lead SNPs (Sniekers et al., 2017), (2)242 lead SNPs (Savage et al., 2018), (5) 434 lead SNPs (Davies et al., 2018). MTAG analysis (Hill, Marioni, et al., 2018), combining data from Sniekers (1) and Okbay (5) results in 564 lead SNPs (4). EA includes 15 lead SNPs (Davies et al., 2016), 69 lead SNPs (Rietveld et al., 2014), (5) 74 lead SNPs (Okbay et al., 2016) and (6) 1271 lead SNPs (Lee et al., 2018). The MTAG results from Lee et al show an increase in lead SNP detection in EA from 1271 to 1624 lead SNPs (6).

#### 1.4.1.2 Polygenic scores and cognition

A polygenic score (PGS) or polygenic risk score (PRS) is a statistic measuring an individual's genetic 'loading' for variability in a trait (e.g. cognitive function) or risk of illness (e.g. schizophrenia) (Shafee et al., 2018). Using GWAS results, a PGS is a count of the number of common associated alleles carried by an individual, weighted by the strength of the allelic associations with the disorder or trait. PGS based on the GWAS above can explain 11–13% of the variance in educational attainment and 7–10% of the variance in cognitive performance in independent samples (Lee et al., 2018). Despite the major advances that these studies represent, this suggests that a significant gap remains between the overall heritability for cognition estimated from twin studies and SNP-based heritability for cognition (i.e. the contribution of common SNPs that can be analysed by GWAS), reported to

be 0.19 for general intelligence (Savage et al., 2018). This missing heritability is likely due to a variety of factors, including rare variants, gene x gene (GxG) interactions (epistasis) and gene x environment (GxE) interactions (Plomin & von Stumm, 2018).

#### 1.4.1.3 Rare variants

Copy number variants (CNVs) are structural variants that were originally described as >1 kbp sections of DNA that can be present in a human genome at a different copy number to the expected two copies in the reference genome. These can be deletions, duplications, inversions or other complex rearrangements, and can range in frequency, but it is those that are rare that have been of most interest in the study of complex phenotypes (Feuk, Carson, & Scherer, 2006; Kendall et al., 2017; Lee & Scherer, 2010). Recent technological advancement of comparative genomic hybridisation and high-throughput next generation sequencing has led to an improvement in the sensitivity of detection of CNVs resulting in the redefinition of their size to >50 bp (Nowakowska, 2017). An assembly-based approach to sequencing data from two haploid genomes identified over 460,000 variants from 2bp to 28kbp. Only 10% of these variants were detected in an analysis of the 1000 Genomes Project, highlighting that structural variants have been under-called and under-studied in human genomics (Huddleston et al., 2017). Structural variants contribute to genetic diversity (Chiang et al., 2017) and their important contribution to the genetic variability of cognition is now recognized (Feuk et al., 2006).

CNVs have been associated with disruption of cognitive development leading to intellectual disabilities and other neurodevelopmental disorders (Huguet et al., 2018). CNVs associated with these disorders may have incomplete penetrance in a population and apparently healthy adults may carry some of the CNVs associated with these disorders without displaying symptoms (Kendall et al., 2017). A study based on the reasonably homogenous Icelandic population showed that incomplete penetrance of pathogenic CNVs for autism and schizophrenia was associated with decreased cognitive performance in the healthy population and that individual CNVs affected different cognitive domains (Stefansson et al., 2014). Examination of non-pathogenic deletions based on children from the Saguenay Youth Study (n= 1,983) and the IMAGEN consortium (n= 2,090) found that non-pathogenic deletions were associated with decreased IQ and suggested that IQ was linked to haploinsufficiency of

most of the coding genome (Huguet et al., 2018; Pausova et al., 2017; Schumann et al., 2010).

Thirty-three CNVs associated with risk of neurodevelopmental disorders were examined for association with cognitive performance in the UKB (n= 420,247) using seven cognitive measures. Twenty-four of the 33 CNVS were associated with reduced cognitive performance in healthy carriers and these CNVs also showed an association with reduced educational attainment and income. In addition, all 12 of the CNVs associated with schizophrenia have been associated with reduced cognitive function in healthy adults (Kendall et al., 2019). In comparison to healthy non-carriers, healthy individuals who carried at least one of the 12 copy number variants associated with schizophrenia showed reduced brain volumes in the hippocampus, nucleus accumbens and thalamus, suggesting a mediation role for hippocampal and thalamic volumes in cognitive ability (Warland, Kendall, Rees, Kirov, & Caseras, 2019).

Disruptive (loss-of-function) and damaging (missense) rare and ultra-rare single nucleotide variants (SNVs) in highly constrained (HC) genes, i.e., genes under negative selection, are associated with neurocognitive disorders but are also found in the healthy population where they are associated with decreased EA. In a sample of 14,133 individuals, carrying either a disruptive or damaging SNV in a HC gene was associated on average with a reduction in years of education of 2.9-3.1 months (Ganna et al., 2016). Each additional disruptive SNV reduced the chance of going to college by on average 14%. This effect of ultra-rare disruptive and damaging SNVs on EA more than doubled when considering HC genes that are highly expressed in the brain.

In a novel approach to explaining the missing heritability in genetic studies on cognition, Hill et al, examined the high level of linkage disequilibrium found in members of the same family in the Generation Scotland family cohort (n=20,000) (Hill, Arslan, et al., 2018; Smith et al., 2006). This analysis using a tool based on a genome-based restricted maximum, GREML-KIN, measures both the variance explained by the genetic effects clustered in families and common SNPs and was replicated in unrelated individuals (Xia et al., 2016; Zaitlen et al., 2013) . Results showed that for general cognitive ability, genetic effects explained 54% of phenotypic variation, of which 31% was explained by pedigree-associated variants (which include rare variants, CNVs and structural variants) and 23% by common variants. These results are similar to heritability levels found in previous twin studies (Plomin & Deary,

2014). Overall, these findings show that most of the pedigree variants associated with cognition were rare with allele frequencies between 0.001 and 0.01 and current genotyping platforms do not sufficiently tag these variations.

#### 1.4.1.4 Gene by environment interactions

Hasan and Afzal (Hasan & Afzal, 2019) argue that to fully understand cognition, environmental effects need to be explored. Interplay between nature and nurture has been found through twin and adoption studies, with environment and genetics observed to co-vary in a manner whereby genetic make-up can determine environmental conditions. They propose that the study of candidate genes arising from next generation sequencing should include environmental parameters (Hasan & Afzal, 2019). While PGS can explain 10% of the variation in educational attainment some of this is indirect and is explained by passive gene-environment correlation where parents and other relatives provide a rearing environment that is associated with the parental genotype (Cheesman et al., 2019; Kong et al., 2018). A recent study shows that PGS for intelligence and EA had a 60% greater predictive value when tested between families as opposed to within families. This difference disappears when socio-economic class is controlled (Cheesman et al., 2020). In a further study of adopted individuals in the UKB (n=6311) it was found that PGS generated from mainly non-adoptive individuals was only 50% as predictive of YOE in adoptees when compared with non-adoptive individuals and conclude that parental influences affect YOE. It was also found that individuals who have a low PGS for YOE spent longer in education if adopted supporting the gene-environment correlations theory (Cheesman et al., 2020). These studies support the inclusion of environmental effects in genetic studies of cognition.

#### 1.4.1.5 Current developments in cognitive research

The comparison of the first release of whole exome sequencing data from the UKB on ~49,000 individuals and their previously imputed genetic data identified nearly four million coding SNPs and indels per individual, ~7 times higher than that observed in the imputed GWAS data. There was also a 10-fold increase in the identification of loss-of-function variants and loss-of-function variants were found in 97% of autosomal genes (Van Hout et al., 2019). A further release of exome data for ~200,000 lead to the examination of the association between protein-truncating variant gene burden and cognitive phenotypes. This study identified four novel genes associated with cognitive function (Chen et al., 2021) .

Whole exome sequencing of the remainder of the UKB, which is on-going, and subsequently whole genome sequencing will allow for new analysis of cognition phenotypes using rare genetic variants and may give new insights into the genomics of cognition.

According to Eichler, identifying all the genetic contribution is not just a matter of increasing sample size, as variants are being missed with short read datasets that are aligned to a single reference genome, even when using whole genome sequencing (Eichler, 2019). He argued that more meaningful results will be obtained by diversification of genomic data. Generic research to date on cognition (and other traits) has been almost exclusively confined to samples of individuals of European ancestry. Lee et al. found that their PGS for EA was far less predictive in an African American sample (Lee et al., 2018). Eichler proposed that the use of combinations of reference genomes from different populations, that are currently in production should in theory identify the majority of structural variants which have been untested in recent GWAS (Eichler, 2019; McCarthy et al., 2016; Stefansson et al., 2013). It also important that reference genomes contain representation for African populations to encompass the evolutionary influences on the genome (McClellan, Lehner, & King, 2017).

The use of whole genome sequencing, long-read and ultra-long-read sequencing technology coupled with the development of bioinformatic tools and the further extrapolation of the biological association of over 1000 lead SNPs identified by Lee et al. for EA and others should generate a great insight into cognitive processes. In addition, further development of tools and research approaches that gives us a greater insight into the interplay of the environment and genomics in healthy and psychiatric cohorts will add to our understanding of the critical biological pathways involved in neurocognition.

### 1.4.2 Genetics of cognitive resilience

#### 1.4.2.1 Overview of the genetics of cognitive resilience

It is now clear from recent genetic studies that cognition is highly polygenic. The genetics of cognitive resilience is even more complex as in addition we are examining rates of change in cognitive performance over time, and we depend on consistent and robust phenotypic measures of cognition over those long periods of time.

While research into the effects of environmental factors have shown the importance of cardiovascular health, social involvement and diet on healthy ageing, our assessment of the

understanding of the genomics involved in cognitive decline is hampered by the lack of strong cognitive measures coupled with large genetic datasets. As yet, we do not know whether cognitive decline is genetically influenced by genes associated with general intelligence or if genes that regulate other biological processes are involved.

The resource-modulation hypothesis proposes that during ageing, losses of neurochemical and neuroanatomical resources have a modulating effect on genes associated with cognitive function in that genetic effects become increasingly important and genes that appear to have a weak effect with cognitive performance in young adults may have a stronger association in older adults (Lindenberger et al., 2008). In addition, there is growing evidence that nonadditive genetic variance becomes increasingly important with age where the influence of life course dynamics plays an important role (Reynolds & Finkel, 2015). Longitudinal twin studies have shown that as the rearing environment is more distal, more variance in genetic effects is seen.

Genetic variation accounts for 40 to 50% of cognitive performance of older adults and 24% of the variability of cognitive change over the life span (Davies et al., 2018; Deary et al., 2012). Some studies show an association between genetic variants and age-related cognitive decline, yet they only explained a fraction of the phenotypic variability. In addition, many of the studies failed to replicate due to difference in cognitive measurements and other methodological issues and lack of control of participant characteristics (Andrews et al., 2016).

Interestingly, recent research has shown that neurogenesis occurs in the dental gyrus of the adult hippocampus into the 8th decade of life despite declines in neural plasticity, angiogenesis, and quiescent stem cell pools (Figure 1.12). Healthy individuals without neurodegenerative conditions show preserved neurogenesis. The authors propose that individual resilience leads to variation in rates of neurogenesis and differing rates in cognitive decline (Boldrini et al., 2018). However, there are still many unresolved questions on the role that adult neurogenesis plays in brain repair, neuroplasticity and overall hippocampal function (Kuhn, Toda, & Gage, 2018).

A meta-analysis of studies on cognitive decline concluded that major improvements were needed in research methods, in particular the use of standardized procedures across studies (Plassman, Williams, Burke, Holsinger, & Benjamin, 2010).

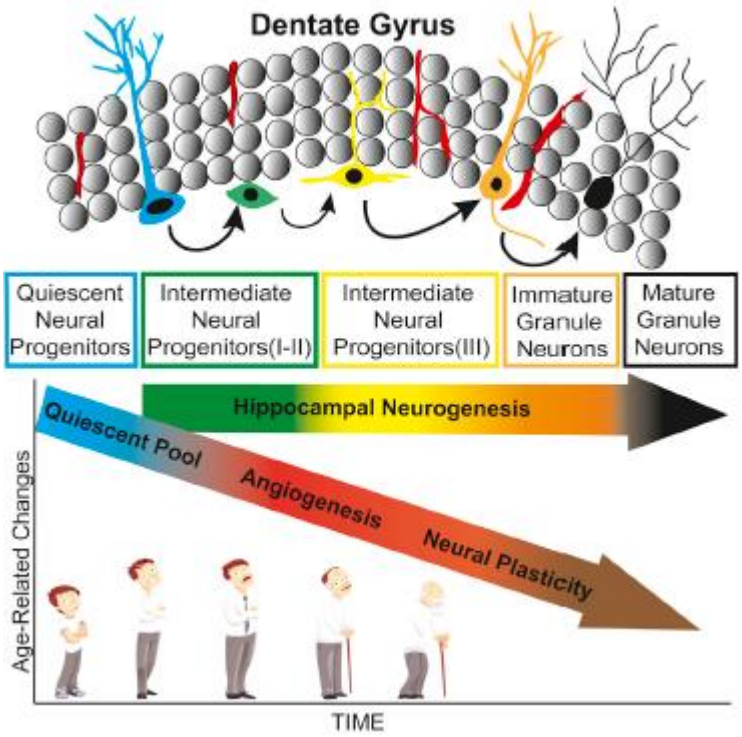



**Figure 1.12: Human hippocampal neurogenesis persists throughout ageing**
Persistent adult neurogenesis in humans into the eighth decade of life, despite declines in quiescent stem cell pools, angiogenesis, and neuroplasticity. Over a 65-year age span, proliferating neural progenitors, immature and mature granule neurons, glia, and dentate gyrus volume were unchanged (Boldrini et al., 2018).

#### 1.4.2.2 Genetic studies to date

There are a few modest datasets that have both genetic data and longitudinal data. The most prominent being the US Health and Retirement Study (HRS) (Sonnega et al., 2014) and its sister UK dataset , the English Longitudinal Study of Ageing (ELSA) (Steptoe, Breeze, Banks, & Nazroo, 2013). The HRS dataset consists of approximately 21,000 participants who have been assessed by consistent cognitive tests every two years from 1996. The level of genotypic data on these individuals is increasing over time. A GWAS was performed in 2014 on 5765 participants of European extraction using total cognition to assess cogntive decline. Associations were found with two loci and chromosome 19, mapping to an Apolipoprotein E (*APOE*) intron and at Translocase of outer mitochondrial membrane 40 homolog (*TOMM40*) intron. A second GWAS published in 2017 (n= 7,486) on cognitive change using the construct of delayed recall found significant associations between one variant, rs2075650, in

*TOMM40*, and conditioning analysis indicates the change in cognitive function is driven by *APOE*. This finding was replicated in ELSA (n=6,898) (Arpawong et al., 2017).

Another genetic study used participants from the US Religious Order Study (n=749) and confirmatory testing in 3 other cohorts (n=717 to 825) only found markers for *APOE* associated with rate of cognitive decline (De Jager et al., 2012).

Both *APOE* and *TOMM40* are associated with late onset Alzheimer's disease (LOAD) and have not been associated with normal cognitive decline (Zhang & Pierce, 2014).

The Personality and Total Health (PATH) is an Australian cohort of European extraction (n= 1570) and was used to examine the relationship of loci identified in a previously mentioned GWAS (Davies et al., 2015). Of the 3 loci identified by Davies one was in the *TOMM40* region but the other two *MIR211*-rs10457441 on chromosome 6 and *AKAP*6-rs17522122 on chromosome 14 were not associated with Alzheimer's disease and were examined for association with cognitive decline and cognitive change in baseline and longitudinal data accumulated over 12 years in the PATH cohort. Using linear mixed models both SNPs were tested for association with perceptual speed, reaction time, working memory, episodic memory, and vocabulary. *AKAP6* was associated with baseline performance across multiple domains but not with cognitive change. *MIR2113,* on the other hand was associated with memory decline over time (Andrews, Das, Anstey, & Easteal, 2017).

Other cohorts that have been monitored both phenotypically and genotypically for cogntive decline are the Lothian birth cohorts from 1921 (LBC1921, n= 550) and 1936 (LBC1936, n= 1,091). Both cohorts were tested for cognitive performance at the age of 11. The surviving members of the LBC1921 who have a mean age of 79 at baseline were tested 5 times up until the age of 92. The LBC1936 have been tested five times so far, from the age of 70 to 82 (Taylor, Pattie, & Deary, 2018). There has been a comprehensive analysis of available the genotypic data on these cohorts (Corley, Cox, & Deary, 2018). Various candidate genes highlighted in other studies have been tested within these cohorts but apart from the *APOE* e4 allele no other gene associations with age -related cognitive decline or cognitive function were found. PGS analysis identified several conditions associated with lower cognitive performance but increased cognitive decline was associated with a PGS for schizophrenia alone (Corley et al., 2018). The complexity of these processes suggests a highly polygenic

genetic contribution to cognitive decline and suitable datasets are needed to examine the genes and biological pathways involved.

A recent study was carried out using the Lothian Birth Control cohort of 1935 at four different time points between the age of 70 and 79 to measure the association of changes in 'g' with fourteen robustly generated PGS. These PGS included EA, grip strength, schizophrenia, Alzheimer's disease, and other health related PGS. The researchers conclude that the predictive power of PGS in not yet sensitive enough to explain the variance in cognitive decline (Ritchie et al., 2019).

Local regulatory networks (LRNs) are produced by combining data on genetic variants and multi-omics data that infers mechanisms that regulate expression of certain genes in the ageing brain. Multi-omics data from the ROS and the Rush Memory and Aging Project (MAP) cohort (n= 413) was used to generate LRNs which were then related to measures of cognitive decline. This process identified a number of neuronal genes that are predicted to control cognitive decline – the most prominent of these being *STAU1* and *SEMA3F* (Tasaki et al., 2018).

## 1.5 Limitation of research on genetics of cognitive resilience to date

In comparison to the growing genetic understanding of cognition, the understanding of the genetics of cognitive resilience is in its infancy. The studies performed to date have relied on data from very small cohorts and these cohorts do not have the power to study the effect of ageing on a highly polygenic trait of cognition. In addition, much of the focus on cognitive decline had been on neurodegenerative disorders with a particular focus on dementia, and as a result the study of cognitive decline in healthy ageing has been neglected. The conundrum in studying the genetics of cognitive resilience is that while large datasets such as the UKB and 23and Me are emerging, they do not have longitudinal measures on cognition and indeed, it will take several years for this data to accumulate. In chapter 6, I discuss at length the type of study I would design given unlimited resources. We need studies designed specifically to examine the genetics of cognitive resilience. These studies should also collect environmental data that effects healthy ageing to allow the use of modelling to examine gene/environment interactions. In the meantime, alternative strategies such as the use of proxy phenotypes for past cognitive performance and available cognitive phenotypes should be explored.

## 1.6 Project aims

It is very important to understand cognitive resilience in health ageing for a number of reasons, including (a) the growing burden on society driven by the gradual shift to an ageing population and the need to plan government strategies to manage this change, (b) to have a good quality of life for as long as possible at the later end of life to sustain independent living and (c) to reduce the fear we all have of diminishing cognitive performance as we age. While there is a general understanding of the environmental factors that contribute to cognitive resilience therefore protecting against cognitive decline, the understanding of the genetic contribution to resilience is in its infancy. Due to a lack of suitable genetic and longitudinal data the aim of this thesis is to investigate alternative ways to obtain genetic information on cognitive resilience using large datasets that do not have direct longitudinal data on cognition.

The central hypothesise guiding my research is that genetic variation in the population bestows enhanced cognitive resilience on certain individuals and this derives from associated genetic variants strengthening the biological processes involved in neuronal activity relevant to cognition.

Based on this central hypothesise, my research explored the genetic basis of cognitive resilience in the UKB using proxy measures to estimate past cogntive performance and current cognitive performance, so as to address the following research questions:

a) Can individual genetic variants associated with cognitive resilience be identified in the UKB?
b) Can these results be replicated within the UKB using independent discovery and replication samples?
c) Does functional analysis using the GWAS output highlight specific brain regions, cell types, biological processes and pathways that are enriched for genes associated with cogntive resilience?
d) Can existing longitudinal datasets be used to confirm these findings?
e) Do the findings support the contribution of brain reserve, cognitive reserve, and brain maintenance to cognitive resilience?
f) Is there more to superior cognitive resilience than superior intelligence?

Firstly, in chapter 3, I explore suitable phenotypes within the UKB to use in a GWAS of cogntive resilience and isolate those SNPs associated with resilience independent of other influences. I then examine the mapped genes associated with these SNPs and link them to biological processes. I then explore the use of structural equation modelling to produce a full GWAS of cognitive resilience. I show that this method can be replicated by first using a discovery sample and repeating the work in a replication sample. In Chapter 4, I describe how I combine my findings to perform a functional analysis of the full GWAS. In chapter 5, I perform *ad hoc* analysis to satisfy potential questions on the findings of the functional analysis and I explore the use of a longitudinal dataset (the HRS dataset) to confirm findings. Finally , in chapter 6, I summarise my findings and explore the potential for future research .

# 2 Materials and Methods

**Table 2.1: Bioinformatic tools used in this thesis.**

| Acronym | Description | Version | Purpose | Web link |
|---|---|---|---|---|
| **BIG40** | Oxford Brain Imaging Genetics Server - BIG40 | 26/03/2021 | Imaging phenotypes for UKB | https: //open.win.ox.ac.uk/ukbiobank/big40/ |
| **BioVenn** | Comparison and visualization of biological lists using area-proportional Venn diagrams | Current | Venn diagram | https://www.biovenn.nl/ |
| **ConsensusPathDB human** | Integrates interaction networks in Homo sapiens | 34 | Overrepresentation analysis of gene sets (data mining) | http://cpdb.molgen.mpg.de/ |
| **dbGaP** | Database of Genotypes and Phenotypes | Current | Downloading HRS data | https://www.ncbi.nlm.nih.gov/gap |
| **dbSNP** | Database of Short Genetic Variation | 2019 | Data mining | https://www.ncbi.nlm.nih.gov/snp |
| **EVP** | Ensembl Variant Effect Predictor | 104 - May 2021 | Data mining | https://www.ensembl.org/info/docs/tools/vep |
| **FINEMAP** | Efficient variable selection using summary data from genome-wide association studies | 01:01 | Fine mapping of SNPs | http://www.christianbenner.com/ |

| Acronym | Description | Version | Purpose | Web link |
|---|---|---|---|---|
| **FUMA** | Functional Mapping and Annotation (FUMA) | 1.3.6a | Functional analysis | https://fuma.ctglab.nl/ |
| **Galaxy** | Web-based platform for data intensive biomedical research | Current | Manipulation of large datasets | https://usegalaxy.org |
| **GCTA -GSMR** | Tool for Genome-wide Complex Trait Analysis | v1.93.2beta | Analysis of mendelian randomisation | https: //cnsgenomics.com/software/gcta/#GSMR |
| **GenomicSEM** | R-package for structural equation modelling based on GWAS summary data | 0.0.2e | package to perform GWAS-by-Subtraction and LDSR | https: //github.com/MichelNivard/GenomicSEM/wiki |
| **GitHub** | GBS sample size (N effective) calculation | N/A | Calculation of sample size after GBS | https: //github.com/PerlineDemange/non-cognitive/blob/master/GenomicSEM/Cholesky%20model/Calculation_samplesize.R |
| **GitHub** | Repository for code | current | Code used in thesis | https://github.com/joanfitz5/cog.res |
| **GWAS Atlas** | Atlas of GWAS Summary Statistics | 3:20191115 | Source of public summary statistics and comparing GWAS outputs | https://atlas.ctglab.nl/ |
| **GWAS Catalog** | The NHGRI-EBI Catalog of human genome-wide association studies | 19/05/2021 | Source of public summary statistics | https: //www.ebi.ac.uk/gwas/ |

| Acronym | Description | Version | Purpose | Web link |
|---|---|---|---|---|
| **GWAS-by-subtraction** | A tutorial on how to perform GWAS-by-subtraction in GenomicSEM | 14/01/2020 | Method used to perform GBS | https: //rpubs.com/MichelNivard/565885 |
| **Haploreg** | A tool for exploring annotations of the noncoding genome at variants on haplotype blocks | 4.1 | Obtaining proxy SNPs to compare HRS and UKB datasets | https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php |
| **Ldlink** | A tool to interrogate linkage disequilibrium in population groups. | 5.1 | Exploring relationships between SNPs | https://ldlink.nci.nih.gov/ |
| **Linux** | Operation system | 7 | Creation, analysis and storge of files through the bash command line | https://www.centos.org/ |
| **Locus Zoom** | Tools to provide fast visualization of GWAS results | 0/13 | Data mining | https://my.locuszoom.org/ |
| **MAGMA** | Gene analysis and generalized gene-set analysis of GWAS data | 1.08 | Performing conditional analysis on genset enrichment data | https://ctg.cncr.nl/software/magma |
| **Mathcracker** | Mathematical calculations | 2 | Simple sign test | https://mathcracker.com/sign-test |

| Acronym | Description | Version | Purpose | Web link |
|---|---|---|---|---|
| **Plink** | Whole genome association analysis toolset | 1.9 | Performing preliminary GWAS and creating LD files for FINEMAP | https://www.cog-genomics.org/plink/1.9 |
| **Plink** | | 2 | Performing GWAS and working with Pgen files | www.cog-genomics.org/plink/2.0 |
| **R** | R is a language and environment for statistical computing and graphics | 3.6.1 | Running GenomicSEM , LDSR, and other statistical analysis | https://www.r-project.org/ |
| **SPSS** | Statistics for the social sciences | 24 | Analysis of the phenotypic data in UKB and HRS | https://www.ibm.com/analytics/spss-statistics-software |
| **UKB** | UK Biobank | May-21 | Source for UKB data | http: //biobank.ndph.ox.ac.uk |
| **Venn diagram** | Calculate the intersection(s) of list of elements | 3 | Venn diagram used to study prioritised genes (>3 datasets) | https://www.vandepeerlab.org/?q=tools/venn-diagrams |

## 2.1 UK Biobank

### 2.1.1 Participants

The UKB is prospective study to examine a range of conditions/traits in middle age to older adults. A dataset of 502,620 community dwelling participants between the ages of 37 and 73, recruited from all over the UK in the period of 2006 to 2010. Sampling at baseline included physical and cognitive measures, completion of lifestyle questionnaires and blood, urine, and saliva samples. Some participants performed follow up in person and web-based cognitive tests. Imaging sampling of participants is ongoing and currently data is available for approximately 50,000 individuals. More detail on the study is available from the UKB (Sudlow et al., 2015). We obtained permission to access both the phenotypic and genetic data under project # 23739.

### 2.1.2 Genetic data

Bycroft et al. describe the processes used by the UKB to genotype DNA extracted from blood samples collected from participants. Genotyping was carried out by Affymetrix Research Services Laboratory using the UK BiLEVE Axiom and Biosystems UK Biobank Axiom arrays. Quality control parameters were applied prior to imputation. Imputation was performed using the Haplotype Reference Consortium data and the merged UK10K and 100 Genomes phase 3 reference panels using the IMPUTE 4 programme (Bycroft et al., 2018).

Encripted files were downloaded from UK Biobank using Aspera. Files were decrypted using the EgaDemoClient application. During our in-house quality control of the imputed data, samples were restricted to those of European descent using 1000 Genomes data and PCA. UKB directly genotyped files were merged with the 1000 genome project vcf files, and the SNPs used for PCA by UKB (identified in UKB supplied marker QC file) were extracted using Plink2 and the --approx option to minimise memory requirements. The multi-mean of the 1000 Genomes CEU samples was calculated and UKB samples with a Mahalanobis distance < 6 SD from this multi-mean were identified as being of European ancestry and were retained.

We excluded related samples using UKB supplied relatedness files which lists pairs of individuals related up to a third degree. Subjects with more than 10 relatives were removed followed by one individual from each pair until no related subjects remained. We also removed samples with discordant sex information, chromosomal aneuploidies, high

missingness/heterozygosity, retracted consent and missing phenotype or covariate data. The final sample size used in this analysis was 333,664 participants.

Using Plink, Imputed variants were converted to hard calls at a certainty threshold of 0.9. SNPs were excluded if their proportion of missing genotypes exceeded 2%, minor allele frequency (MAF) was less than 1%, or Hardy–Weinberg equilibrium (HWE) was lower than $1\times10^{-6}$. Duplicate SNPs were removed resulting in 8,378,152 variants for use in our final analysis.

### 2.1.3 Cognitive data

Participants were tested using several cognitive tests which are described fully in the UKB. The types of tests and the method of collection and reliability are described elsewhere (Fawns-Ritchie & Deary, 2020; Lyall et al., 2016). A summary of the tests used in this thesis are listed in Table 2.2. Correlation analysis of these tests with age of participants in the UKB was performed using SPSS V.24. A brief description of these tests will follow.

**Table 2.2: Cognitive data in the UKB and its correlation with age**

| Test | N | r |
|---|---|---|
| Initial tests 2007/2008 | | |
| Verbal Numerical reasoning | 165,486 | 0.05* |
| Reaction Time | 496,776 | 0.27* |
| | | |
| Follow up 2014+ | | |
| Verbal Numerical reasoning | 21,204 | 0.04* |
| Reaction Time | 21,689 | 0.25* |
| | | |
| | | |
| Online tests (2014/2015) | | |
| Trail Making (#1) Online | 104,052 | 0.27* |
| Trail Making (#2) Online | 104,050 | 0.34* |
| Symbol Digit Substitution Online | 118,490 | 0.43* |
| Verbal Numerical reasoning Online | 123,665 | 0.12* |
| Numeric Memory Online | 111,086 | 0.13* |
| Principle Component analysis factor (g) | 111,039 | 0.39* |

Note: *P<.01, n=number of participants, R = correlation with age

#### 2.1.3.1 Verbal-numerical reasoning test (fluid intelligence)

At total of 165,450 participants were tested at baseline (UKB ref 210016-0.0). This assessment consisted of 13 multiple choice scenarios of ranging complexities designed to test both verbal and numerical skills. Participants were scored on the number of correct answers given in two minutes. It was also administered at two follow up visits (n= 20,113 and n=21,204) and a similar test with 14 questions was administered online (n=123,665). We found that this test only showed a small correlation with age (r = -0.05, $P < 0.01$) and there was no decline seen at the two subsequent time points with mean scores of 5.98, 6.59 and 6.73, respectively. The improvement in scores over time may be due to practice effects (Lyall et al., 2016).

#### 2.1.3.2 Reaction time (RT)

A total of 496,790 participants were tested at baseline for reaction time (UKB ref 20023-0.0) and at two follow up intervals (n=20,257 and n=21,689). This test consisted of matching pairs of cards. The participants were presented with pairs of cards that were either identical or different. The participant was required to acknowledge matching pairs by pushing a button as quickly as possible. RT had a moderate correlation with age (r = 0.27, $P < 0.201$).

#### 2.1.3.3 Other baseline tests

Other cognitive tests were performed at baseline (Table 2.3). These include numerical memory, visual pairs matching and prospective memory. These parameters were not considered in my analysis as along with fluid intelligence, they had a low correlation with age.

**Table 2.3: Comparison of the correlation of age with cognitive measures**

| UKB Ref | Variable | N | Correlation (P<.01) |
|---|---|---|---|
| **20016-0.0** | Fluid intelligence score | 165,477 | -0.05 |
| **20023-0.0** | Reaction Time | 496,713 | -0.27 |
| **4282-0.0** | Numeric Memory | 51,811 | -0.08 |
| **399-0.1** | Visual Mem Pairs Matching | 497,926 | -0.10 |
| **4292-0.0** | Prospective Memory | 171,569 | -0.10 |

#### 2.1.3.4 Web based tests:

- **Verbal-numerical reasoning test** (fluid intelligence) which was similar to that described in section 2.1.1.2.1 with an additional question.
- **Symbol digit substitution test** which measures processing speed where participants are shown a key which paired symbols with numbers and participants were measured in their ability to correctly match symbols to digits as quickly as possible. They were scored on the number of correct symbol-digit matches in 60 seconds.
- **Trail making** measures executive function and consists of two parts. In part 1, participants must arrange a series of 25 numbers on a screen in numerical order and in part 2, had to switch between numbers and letters to arrange them in the correct order. They were measured on the time taken to complete the task.
- **Numerical memory** measures working memory and a participant is required to remember and reverse a sequence of numbers starting with two numbers and increasing until a participant had two wrong sequences or reached 12 digits.

### 2.1.4 Creation of cognitive phenotypes

#### 2.1.4.1 Current cognitive performance

We selected RT to represent current cognitive performance as it had a good correlation with age and data was available on most participants (n=331,495). Using SPSS, RT was adjusted for age to improve normality (Davies et al., 2016), the natural log of corrected RT was computed (Figure 2.1).

(a) (b)

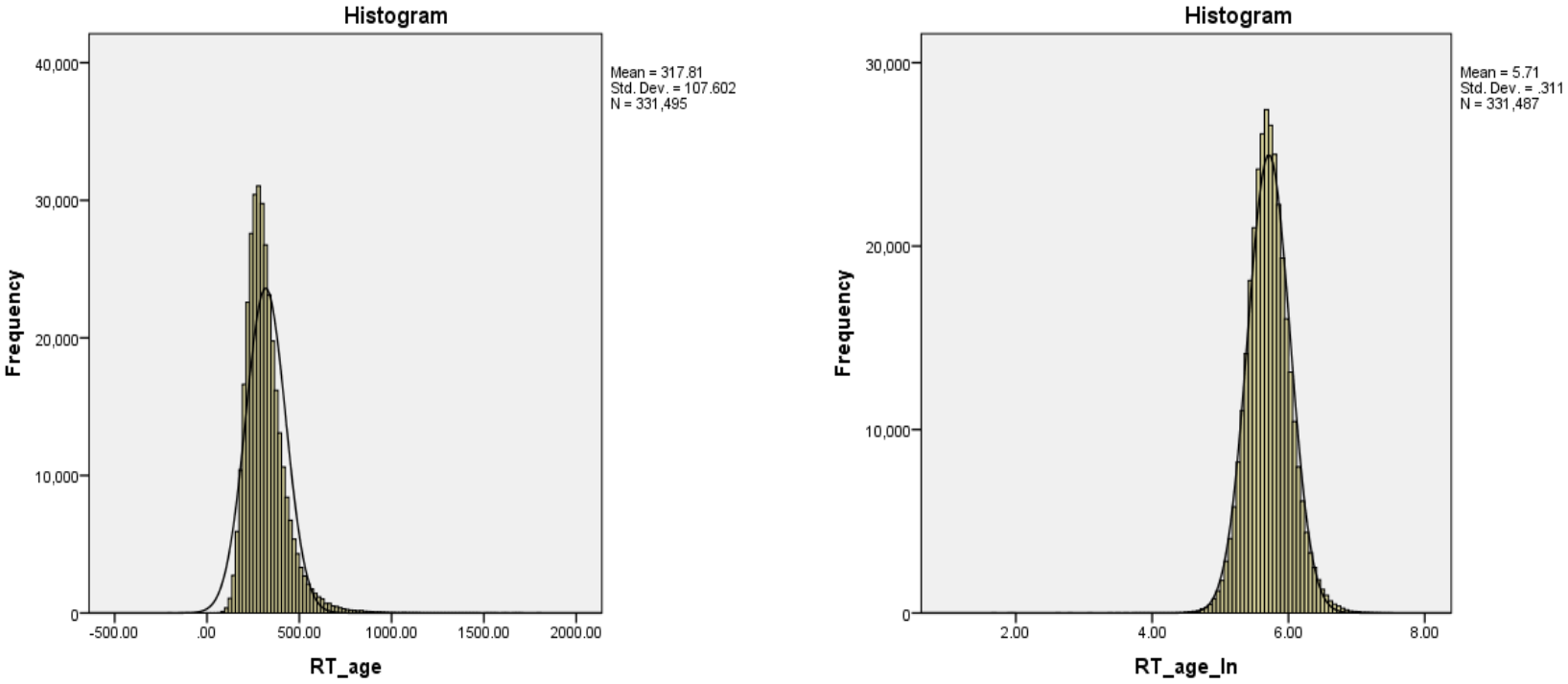



**Figure 2.1: Normalising RT cognitive data.**
(a) RT data before log transformation (b) RT data after log transformation. The x axis shows the RT score, and the y axis is the frequency of that score occurring in the sample.

A binary RT variable was created using the mean value (m = 5.71). Those with a value less than or equal to the mean were considered to have faster than average processing speed/RT (quicker to react) and those above the mean were considered to have slower than average processing speed/RT.

A second variable to measure current cognitive performance combining the web-based tests was created using principal component analysis in SPSS. This created a generalised measure of cognition or 'g'. I subtracted trail 1 from trail 2 to remove the motor speed component from the test (trail making/EF) and then performed a dimension reduction using the online tests of fluid intelligence, symbol digit substitution, numerical memory and trail making/EF. I then examined the correlation of this new PCA variable with age and it had a moderated correlation (r = 0.39, $P < 0.01$) and corrected the variable for age.

#### 2.1.4.2 Past cognitive performance

Given the lack of longitudinal data, an alternative approach was to use proxy phenotypes. For past cognitive performance we examined the use of educational attainment/years in education. Educational attainment is available for 332,089 individuals in UKB that met our genotypic QC requirements. In the dataset, age completed full-time education was recorded for participants who did not go to college but not for those who attended higher education. We therefore assigned a default score of 20 to those who attended college and created a binary phenotype using less than or equal to age 17 to divide participants into two categories – above average and below average education years (EY).

At total of 330,098 individuals had measurements for EY and RT and genetic data and these made up the final sample (Table 2.4).

**Table 2.4: Description of phenotype**

| Variable | N | Minimum | Maximum | Mean | Std. Dev |
|---|---|---|---|---|---|
| **Female** | 179,737 | | | | |
| **Male** | 153,927 | | | | |
| **Age at baseline sampling** | 333,664 | 38 | 72 | 56.85 | 8.01 |
| **Age completed FT education** | 332,089 | 0 | 35 | 17.63 | 2.92 |
| **Reaction time (RT)** | 331,495 | 79 | 1985 | 554.81 | 112.68 |
| **Log RT corrected** | 331,487 | 2.66 | 7.5 | 5.71 | 0.311 |
| **Valid N** | 330,098 | | | | |

Using these two binary variables – above or below average EY and faster or slower RT – I created four group of participants. One of these groups demonstrated high resilience and these were our cases for our first "EY+Res" GWAS who had below average EY previously and faster than average RT now. A second group demonstrated low resilience or cognitive decline, and these were our controls for that GWAS who had above average EY previously and slower than average RT now. The two remaining groups of UKB samples displayed consistent cognitive performance over time. Here our cases for our second "EY/NonRes" GWAS had below average EY previously and slower than average RT now (below average cognition over time) and our controls had above average EY previously and faster than average RT now (above average cognition over time).

### 2.1.5 Performing GWAS with UKB

Initially Plink1.9 was used to perform preliminary genetic studies but Plink 2.0 (Chang et al., 2015) was used to perform all genetic studies used in the final analysis. Logistic regression was used to perform the case/control analysis using --glm (Hill et al., 2017). Covariates used were age, sex, test centre, genotype array and first 8 PCA supplied by UKB. All codes used in Plink are available in my GitHub page at https://github.com/joanfitz5/cog.res. All analysis was performed on the server housed by the School of Mathematics at NUIG.

### 2.1.6 GWAS-by subtraction (GBS)

To extract those SNPs that were associated with resilience only, we used Genomics Structural Equation Modelling (GenomicSEM ) (Grotzinger et al., 2019). There are several processing steps that need to be performed to enable the summary statistics to be processed through GenomicSEM and these are described in the original paper by Grotzinger et al and accompanying tutorials (Grotzinger et al., 2019; Nivard, 2019). Following closely the process use by Demange et al (Demange et al., 2021), we defined a Cholesky model (Figure 2.2: SEM of GWAS-by-subtraction) as follows using the summary statistics from the EY+Res and EY/NonRes GWASs. Both EY+Res and EY/NonRes were regressed on a latent factor, which captured the shared genetic variance in EY (hereafter "*EduYears"*). EY+Res was further regressed on a second latent factor capturing the variance in EY+Res independent of EY/NonRes, hereafter "*Resilience*". Genetic variance in *Resilience* was independent of genetic variance in *EduYears* ($r_g = 0$) as the *Resilience* factor represents residual genetic variation in our EY+Res phenotype that is not accounted for by the *EduYears* factor. These two latent variables, *Resilience* and *EduYears* were then regressed on each SNP in the

original GWASs (EY+Res and EY/NonRes) resulting in new GWAS summary statistics for both *Resilience* and *EduYears* (Figure 2.2). To calculate the path loadings for λEduYears – EY+Res and λResilience – EY+Res, the model was run without the SNPs. Coding used to perform GBS are available in my GitHub page at https://github.com/joanfitz5/cog.res.

### 2.1.7 Calculation of sample size after GBS

Running the analysis through GBS alters the sample size and it is necessary to calculate the new value for downstream analysis. To calculate sample size or effective N (Neff) of the *Resilience* GWAS for discovery, replication and full analyses, we followed the procedure specified in GenomicSEM (Grotzinger et al., 2019; Mallard et al., 2020) and by Demange et al (Table 2.1). To do this we needed to determine path loading for the models used in the three analyses as the path loading differs with different sample sizes. We trimmed our data to only include SNPs with a MAF of >0.10 and <0.40 as low and high MAF can bias the result. The analysis was performed in R 3.6.1 on the math server at NUIG. Output of this analysis and the calculations of sample size is in Table 2.5.

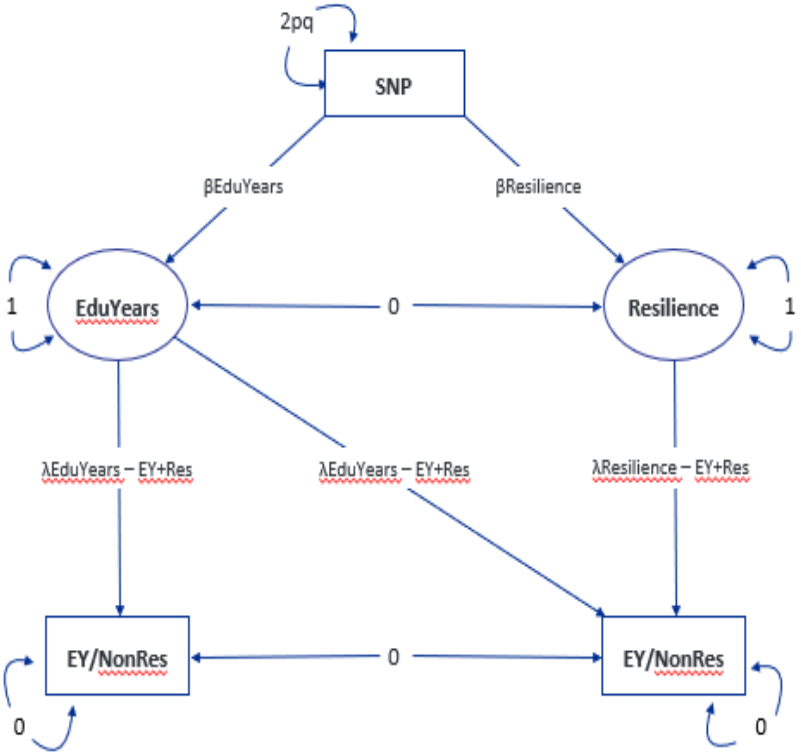



**Figure 2.2: SEM of GWAS-by-subtraction**
The observed variables are the GWAS EY+Res and EY/NonRes and SNP and the latent variables (unknown) are Resilience and EduYears. There are two pathways for the SNPs analysis in this model to EY+Res – the first is through EduYears to EY+Res and EY/NonRes and incorporates the genetic effects of the variables used in the phenotype. The other path is through Resilience to EY+Res and measures the genetic effect of resilience independent of EduYears. To calculate the model, the genetic covariances between EY+Res and EY/NonRes and Resilience and EduYears are set to 0 and the variances of EY+Res and EY/NonRes are also set to 0. The variance is therefore explained by the latent factors. The SNP value is calculated as 2pq from allele frequencies of the 1000 Genome phase 3 data where p is the reference allele and q the alternative allele.

**Table 2.5: Pathway loading and sample size (effective N) calculation**

| **Sample** | **Pathway** | | | | | | **Effective N (Neff)** | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | res =~ EY+Res | | EduYears =~ EY+Res | | EduYears =~ EY/NonRes | | | | | |
| | est | P value | est | P value | est | P value | Min | Median | Mean | Max |
| **Discovery** | 0.4129 | < 5e-300 | 0.2862 | 2.76E-118 | 0.5036 | < 5e-300 | 84,137 | 88,684 | 88,607 | 88,796 |
| **Replication** | 0.4743 | 2.28E-63 | 0.2133 | 1.04E-12 | 0.5228 | 7.04E-132 | 22,618 | 25,786 | 25,706 | 25,926 |
| **Full** | 0.4220 | < 5e-300 | 0.2758 | 7.04E-132 | 0.5081 | < 5e-300 | 105,876 | 111,396 | 111,316 | 111,513 |

### 2.1.8 Identification of genomic loci associated with resilience

**FUMA**: Manhattan plots of GWAS outputs from original phenotypes and GBS outputs were generated in FUMA v 1.3.6 (Watanabe, Taskesen, van Bochoven, & Posthuma, 2017) using a P-value setting of $< 5 \times 10^{-8}$ for genome-wide significant SNPs. We used an LD $r^2$ setting of 0.6 and the 1000G phase 3 European reference panel to identify independent lead SNPs and an additional $r^2$ setting of 0.1 to identify lead SNPs and a maximum distance for LD blocks of 250 kb to separate findings into separate genetic loci. Conditional analysis was performed where there was more than one independent significant SNP within 1000 kb distance using --condition command in Plink 1.9 (Chang et al., 2015), which adds a SNP as a covariate in GWAS analysis. Setting used in FUMA are showed in Table 2.7 at the end of this chapter.

**FINEMAP v 1.4** (Benner et al., 2016) was used to investigate causal SNPs by analysing the relationship between the candidate GWAS SNPs generated in FUMA and LD data. LD files were generated in plink 1.9 using the --r square spaces command – for example -

"./plink --bfile cogres --chr 6 --extract 6.fuma --r square spaces --out 6f"

Results of SNPs listed by Bayes Factor for each locus were examined as well as the configuration files generated by FINEMAP to examine for causal SNPs sets. The maximum number of SNPs in a set was fixed at 3.

Coding used to perform Fine mapping is available in my GitHub page at https://github.com/joanfitz5/cog.res.

### 2.1.9 Function analysis of GWAS output

We used FUMA v 1.3.6 (Watanabe et al., 2017) to perform functional analysis. We used the default settings as described in the Tutorial section of the website and in previous publications (Jansen et al., 2019; Savage et al., 2018) . The parameters used are shown in Table 2.7 at the end of this chapter. FUMA analysis of *Resilience* is published and can be viewed publicly in FUMA as ID:171. We used the calculated effective sample size of 111,316 (Neff) for the analysis of the *Resilience* output to examine the functional consequences of SNPs on genes, Combined Annotation Dependent Depletion (CADD) scores, chromatin states and Regulome DB analysis.

### 2.1.10 Mapping SNPs to genes

Gene-mapping was performed in FUMA using three strategies: (a) *Positional mapping* which mapped SNPs to genes based on their genomic location within a 10 kb window of known gene boundaries. (b) *Expression quantitative trait (eQTL) mapping* which aligned cis-eQTL SNPs to genes whose expression they affected, selecting information from tissue types in 4 datasets in FUMA (PsychENCODE (Wang et al., 2018), BIOS QTL (Bonder et al., 2017), Blood eQTL (Westra et al., 2013), and GTEx 8 (Battle, Brown, Engelhardt, & Montgomery, 2017)). (c) *Chromatin interaction mapping* using the 3D DNA to DNA interactions mapped SNPs to genes.

Gene-set analyses: The GENE2FUNC function within FUMA examines enrichment of mapped genes using hypergeometic tests of 9,494 gene-sets form GTEx (Carithers et al., 2015) , MSigDB (Liberzon et al., 2015) and GWAS catalog (Buniello et al., 2019).

#### 2.1.10.1 MAGMA gene-based analysis

FUMA computes a gene-based genome-wide association analysis (GWGAS) from the SNP-based P-value from the GWAS. A total of 18,879 protein coding genes containing a minimum of one GWAS SNP were used in this analysis and were used to test for association with 53 tissue types obtained from GTEx (Consortium, 2015). Associations were Bonferroni corrected for multiple testing with $P < 0.05/18{,}879 = 2.648 \times 10^{-6}$.

We further explored the sets of associated genes in cell type specificity analyses with scRNA-seq in FUMA (Watanabe, Umićević Mirkov, de Leeuw, van den Heuvel, & Posthuma, 2019) using the following datasets: GSE104276 Human Prefrontal cortex per ages (Zhong et al., 2018), GSE67835 Human Cortex (Darmanis et al., 2015) and Linnarsson Mouse Brain Atlas (Zeisel et al., 2018). We analysed significant cell types across datasets, independent cell type associations based on within-dataset conditional analyses and pair-wise cross-datasets conditional analyses.

Pathway enrichment analysis was performed on curated gene sets and Gene Ontology (GO) terms from Msigbd v 7.0 (Ashburner et al., 2000) terms using the full distribution of SNP P-values from the Resilience GWAS.

### 2.1.11 Comparison with published traits

LD score regression (LDSR) analysis was performed using the LDSC function within GenomicSEM (Grotzinger et al., 2019) to examine the genetic correlation between *Resilience* with other phenotypes. Various sources were used to obtain summary statistics from GWAS of published research in psychiatry, brain imaging, and other traits of interest (see Table 2.6). Summary statistic files generated during GBS were used for *Resilience*, *EduYears*, *EY+Res* and *EY/NonRes* in the LDSR. Associations were Bonferroni corrected for multiple testing with $P < 0.05/21 = 2.88 \times 10^{-3}$.

**Table 2.6: Publicly available datasets used for LDSR**

| Trait | Source | Link | Author | Year | N |
|---|---|---|---|---|---|
| ***Cognitive*** | | | | | |
| **Intelligence** | GWAS catalog | ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/SavageJE_29942086_GCST006250/SavageJansen_IntMeta_sumstats.zip | Savage et al | 2018 | 269,867 |
| **Reaction time** | GWAS catalog | http://www.psy.ed.ac.uk/ccace/downloads/Davies_NC_2018.zip | Davies et al | 2018 | 330,069 |
| **Educational attainment** | GWAS atlas | https://www.dropbox.com/s/ho58e9jmytmpaf8/GWAS_EA_excl23andMe.txt?dl=1 | Lee et al | 2018 | 766,345 |
| ***Psychiatric and Neurological*** | | | | | |
| **Amyotrophic lateral sclerosis** | GWAS catalog | ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/vanRheenenW_27455348_GCST004692/harmonised/27455348-GCST004692-EFO_0000253.h.tsv.gz | Van Rheenan et al | 2016 | 36,052 |
| **Alzheimer's disease** | GWAS catalog | ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/JansenIE_30617256_GCST007320/AD_sumstats_Jansenetal_2019sept.txt.gz | Jansen et al | 2019 | 455,258 |
| **Unipolar Depression** | GWAS catalog | ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/NagelM_29942085_GCST006475/sumstats_depressed_affect_ctg_format.txt.gz | Nagel et al | 2018 | 358,000 |
| **Schizophrenia** | GWAS atlas | http://walters.psycm.cf.ac.uk/clozuk_pgc2.meta.sumstats.txt.gz | Pardinas et al | 2018 | 105,318 |

| Trait | Source | Link | Author | Year | N |
|---|---|---|---|---|---|
| **Bipolar disorder** | PGC | https://www.med.unc.edu/pgc/download-results/bip/?choice=Bipolar+Disorder+%28BIP%29Bipolar+Disorder+%28BIP%29 | Stahl et al | 2018 | 35,802 |
| **Parkinson's disease** | GWAS atlas | https://drive.google.com/open?id=1FZ9UL99LAqyWnyNBxxlx6qOUlfAnublN | Nalls et al | 2019 | 482,730 |
| **Stroke** | GWAS catalog | ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/MalikR_29531354_GCST005843/harmonised/29531354-GCST005843-HP_0002140.h.tsv.gz | Malik et al | 2018 | 520,000 |
| **Neuroticism** | CCAGE (PGC) | http://www.psy.ed.ac.uk/ccace/downloads/Luciano_2017.zip | Luciano et al | 2017 | 329,000 |

### 2.1.12 Mendelian randomisation

Mendelian randomisation was performed using Generalized Summary statistics-based Mendelian Randomization (Zhu et al., 2018) using the GCTA tool v1.93.2 beta (Yang, Lee, Goddard, & Visscher, 2011). The procedure examines credible causal associations between different traits based on GWAS outputs and requires non-overlapping samples. This restricted our analysis because most of the traits examined by LDSC contained UKB participants. However, the sample used for the *discovery.Resilience* GWAS (section 1.2.1) does not contain individuals that have brain imaging data within the UKB so we used this cohort to examine unidirectional and bidirectional causal associations between *Resilience* and brain imaging phenotypes that showed significant correlations with *Resilience* using LDSC. We used a HEIDI-outlier p-value of 0.01 for outlier detection analysis. Given the low level of independent significant SNPs in the *discovery.Resilience* GWAS and the imaging GWAS, we reduced the default minimum level of significant SNPs from 10 to 8. For the disorders of ALS, bipolar disorder and schizophrenia we used the full *Resilience* GWAS and ran the analysis at the default setting of a minimum of 10. Associations were Bonferroni corrected for multiple testing with $P < 0.05/12 = 4.23 \times 10^{-3}$.

## 2.2 Health and Retirement Study (HRS)

The HRS is a longitudinal study of adults aged 50 years or older in households in the United States. The study commenced in 1992 and participants were interviewed at baseline and every two subsequent years. I applied for and was given access to the phenotypic data through the HRS website. To obtain the genetic data we applied for and were given approval through dbGaP as project 18937.

### 2.2.1 Genetic data

Genetic data was downloaded from the portal on dbGaP (see Figure 2.3). Data was downloaded using IBM Aspera connect and was decrypted using the srs tool kit (V2.9.4). The imputes files were presented per chromosome as probability files (gprob.gz) and the quality metrics were supplied with the genotype data.

Plink 2 has the option with gprob of creating binary files (bfiles) or keeping the probability information in the form of pfiles. Preliminary research showed that there was little difference in the output, so I proceeded to use pfiles. Plink 2 does not have the ability currently to merge

pfiles so each chromosome was processed through GWAS separately and the output combined.

Variants were screened by applying quality control filters (geno 0.02, MAF 0.001, info score 0.9 and HWE 0.000001) and removing duplicates for each chromosome.

In addition, many of the HRS SNPs were in an older kgp format and needed to be converted to RSID. This was done using a list of common SNPs per chromosome by RSID and position from the UCSC browser (Kent et al., 2002).

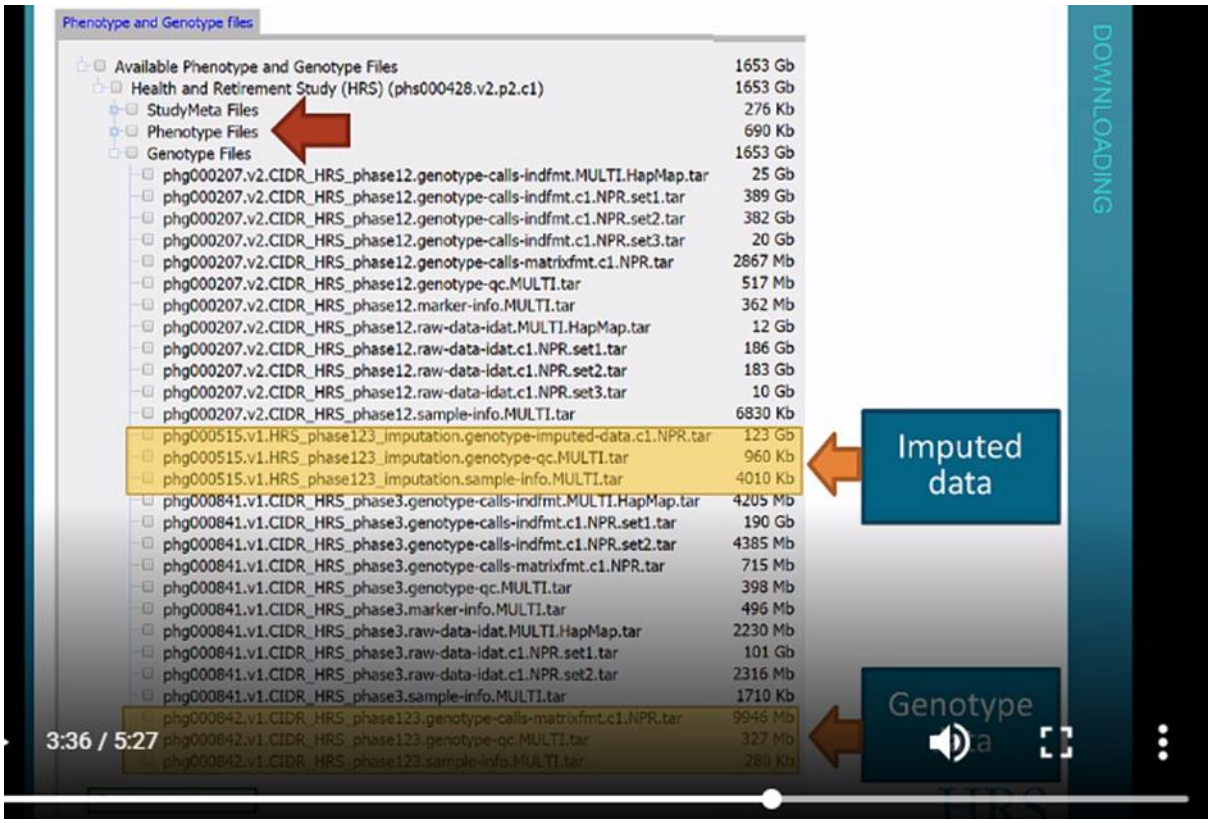



**Figure 2.3: Screenshot of dbGaP showing relevant HRS files**

### 2.2.2 Cognitive phenotypes

Several cognitive tests were administered by interviewers, either by phone of face-to-face and these are described in detail on the HRS website in the following document:

http://hrsonline.isr.umich.edu/sitedocs/userg/dr-006.pdf

The HRS dataset does not contain a reaction time or processing speed measure so after examining the various measures, I selected two cognitive variables to explore further, these were Total Cognitive performance (COGTOT) and Immediate Word recall (Recall).

#### 2.2.2.1 Total Cognitive performance

Using the RAND HRS Longitudinal file 2014 (v.2) I extracted data on total cognitive performance (RxCOGTOT) that was based on the Telephone Interview for Cognitive Status (TICS) (Brandt, Spencer, & Folstein, 1988), which was validated for use to screen for cognitive performance (Welsh, Breitner, & Magruder-Habib, 1993). This is a 27-point test which includes a 10-word immediate and delayed recall test (0–20 points) that measures episodic memory, a serial 7s test to measures working memory (0–5 points), and a backwards-counting test that measures mental processing speed (0–2 points). Scores range from 0 to 27, with lower scores indicating poorer cognitive performance. Cognitive data were collected at each wave of data collection.

I used COGTOT data from 1998 to 2014 as the testing format did not stabilise until 1998. Data on 38,183 participants were available at nine timepoints. In order to diminish the effect of APOE/TOMM40 locus associated with Alzheimer's disease (Mise et al., 2017) on the results , I examined the removal of people demonstrating dementia from the results and I found that a minimum limit of 9 on the cognitive score can be justified (Crimmins, Kim, Langa, & Weir, 2011; Dassel & Carr, 2014; Lievre, Alley, & Crimmins, 2008). I removed any person who had a value below 9, which resulted in the elimination of 929 individuals. After also eliminating participants with missing data and who were not non-Hispanic white/Caucasian, and less than 2 time points of cognitive data, a total of 13,010 participants remained. When matched with the genotypic data, there were 5,345 individuals available for GWAS.

#### 2.2.2.2 Recall

To increase the number of participants with two data points I looked at direct phenotypes in the consisted of the interviewer reading a randomised list of 10 nouns to the respondent from one of four lists, and afterwards asking the respondent to recall as many words as possible. A different list was used for the same respondent for four time points to exclude practice effects.

I had data on immediate recall for 38,183 participants at a total of eleven time points. 15,620 participants had genetic data, and this was reduced further (accounting for ethnicity, missing data, or less than 2 data points) and my final sample size was 9,526 individuals.

### 2.2.3 Creating longitudinal phenotypes with mixed effect modelling

Given the multiple times points in the longitudinal cognitive data in HRS and the fact that many participants would have missed individual time points, it is not possible to measure the effect of change using analysis of variance methods as participants with any missing data would be eliminated. It was therefore necessary to use linear mixed modelling. Following the procedure in SPSS as specified by Andy Field (Field, 2013) page 849, I used the covariates of gender, birth year, education and time of test. I used a step wise approach to add linear, quadratic, and cubic polynomials for time and assessed the effect on the -2 restricted log likelihood results. This showed that timepoint and quadratic time point influence the model for COGTOT. Cubic polynomial time point influenced the model for Recall, so it was included in the model.

### 2.2.4 GWAS of cognitive change over time

Using the two outputs from the linear mixed modelling analysis for COGTOT and Recall, I created two pheno files and ran two GWAS using the first six PCA components as covariates as this is recommended in the genotype QC report (Faul, Smith, & Zhao, 2014). The command used was as follows:

“./plink2 --pfile 22 --glm hide-covar cols=+a1freq --pheno recall.pheno --covar HRS6PCA --covar-variance-standardize --out 22”

GWAS analysis was performed per chromosome and outputs combined into a full GWAS and exported to FUMA for analysis.

## 2.3 Other Bioinformatic tools used

See Table 2.1 for a list of the tool used during my PhD. At the start of my research, I used SPSS and MS Excel for data manipulation and for large datasets I used Galaxy.org to perform sort, extract and join files. During my PhD, I gradually shifted to using server-based R for statistical analysis and learned to use bash commands on the server to do data manipulation.

**Table 2.7: Parameters used in FUMA**

| Parameter | Value |
|---|---|
| **created_at** | 2020-09-30 15:55:59 |
| **title** | res_sept |
| **FUMA** | v1.3.6 |
| **MAGMA** | v1.07 |
| **GWAScatalog** | e96_r2019-09-24 |
| **ANNOVAR** | 2017-07-17 |
| **gwasfile** | all.res.txt.gz |
| **becol** | beta |
| **secol** | se |
| **addleadSNPs** | 1 |
| **N** | 115463 |
| **exMHC** | 1 |
| **MHCopt** | annot |
| **ensembl** | v92 |
| **genetype** | all |
| **leadP** | 5.00E-08 |
| **gwasP** | 1.00E-05 |
| **r2** | 0.6 |
| **r2_2** | 0.1 |
| **refpanel** | 1KG/Phase3 |
| **pop** | EUR |
| **MAF** | 0 |
| **refSNPs** | 1 |
| **mergeDist** | 250 |
| **[magma]** | |
| **magma** | 1 |
| **magma_window** | 0 |
| **magma_exp** | GTEx/v8/gtex_v8_ts_avg_log2TPM:GTEx/v8/gtex_v8_ts_general_avg_log2TPM |
| **posMap** | 1 |
| **posMapWindowSize** | 10 |
| **posMapCADDth** | 0 |
| **[eqtlMap]** | |
| **eqtlMap** | 1 |
| **eqtlMaptss** | PsychENCODE/PsychENCODE_eQTLs.txt.gz:BloodeQTL/BloodeQTL.txt.gz:BIOSQTL/BIOS_eQTL_geneLevel.txt.gz:GTEx/v8/Adipose_Subcutaneous.txt.gz:GTEx/v8/Adipose_Visceral_Omentum.txt.gz:GTEx/v8/Adrenal_Gland.txt.gz:GTEx/v8/Cells_EBV-transformed_lymphocytes.txt.gz:GTEx/v8/Whole_Blood.txt.gz:GTEx/v8/Artery_Aorta.txt.gz:GTEx/v8/Artery_Coronary.txt.gz:GTEx/v8/ |

| Parameter | Value |
|---|---|
| | Artery_Tibial.txt.gz:GTEx/v8/Brain_Amygdala.txt.gz:GTEx/v8/Brain_Anterior_cingulate_cortex_BA24.txt.gz:GTEx/v8/Brain_Caudate_basal_ganglia.txt.gz:GTEx/v8/Brain_Cerebellar_Hemisphere.txt.gz:GTEx/v8/Brain_Cerebellum.txt.gz:GTEx/v8/Brain_Cortex.txt.gz:GTEx/v8/Brain_Frontal_Cortex_BA9.txt.gz:GTEx/v8/Brain_Hippocampus.txt.gz:GTEx/v8/Brain_Hypothalamus.txt.gz:GTEx/v8/Brain_Nucleus_accumbens_basal_ganglia.txt.gz:GTEx/v8/Brain_Putamen_basal_ganglia.txt.gz:GTEx/v8/Brain_Spinal_cord_cervical_c-1.txt.gz:GTEx/v8/Brain_Substantia_nigra.txt.gz:GTEx/v8/Breast_Mammary_Tissue.txt.gz:GTEx/v8/Colon_Sigmoid.txt.gz:GTEx/v8/Colon_Transverse.txt.gz:GTEx/v8/Esophagus_Gastroesophageal_Junction.txt.gz:GTEx/v8/Esophagus_Mucosa.txt.gz:GTEx/v8/Esophagus_Muscularis.txt.gz:GTEx/v8/Heart_Atrial_Appendage.txt.gz:GTEx/v8/Heart_Left_Ventricle.txt.gz:GTEx/v8/Kidney_Cortex.txt.gz:GTEx/v8/Liver.txt.gz:GTEx/v8/Lung.txt.gz:GTEx/v8/Muscle_Skeletal.txt.gz:GTEx/v8/Nerve_Tibial.txt.gz:GTEx/v8/Ovary.txt.gz:GTEx/v8/Pancreas.txt.gz:GTEx/v8/Pituitary.txt.gz:GTEx/v8/Prostate.txt.gz:GTEx/v8/Minor_Salivary_Gland.txt.gz:GTEx/v8/Cells_Cultured_fibroblasts.txt.gz:GTEx/v8/Skin_Not_Sun_Exposed_Suprapubic.txt.gz:GTEx/v8/Skin_Sun_Exposed_Lower_leg.txt.gz:GTEx/v8/Small_Intestine_Terminal_Ileum.txt.gz:GTEx/v8/Spleen.txt.gz:GTEx/v8/Stomach.txt.gz:GTEx/v8/Testis.txt.gz:GTEx/v8/Thyroid.txt.gz:GTEx/v8/Uterus.txt.gz:GTEx/v8/Vagina.txt.gz |
| **eqtlMapSig** | 1 |
| **eqtlMapP** | 1 |
| **eqtlMapCADDth** | 0 |
| **[ciMap]** | |
| **ciMap** | 1 |
| **ciMapBuiltin** | EP/PsychENCODE/EP_links_oneway.txt.gz:HiC/PsychENCODE/Promoter_anchored_loops.txt.gz:EP/FANTOM5/EP_correlation_cell_type_oneway.txt.gz:EP/FANTOM5/EP_correlation_organ_oneway.txt.gz:HiC/Giusti-Rodriguez_et_al_2019/Adult_Cortex.txt.gz:HiC/Giusti-Rodriguez_et_al_2019/Fetal_Cortex.txt.gz:HiC/GSE87112/Adrenal.txt.gz:HiC/GSE87112/Aorta.txt.gz:HiC/GSE87112/Bladder.txt.gz:HiC/GSE87112/Dorsolateral_Prefrontal_Cortex.txt.gz:HiC/GSE87112/Hippocampus.txt.gz:HiC/GSE87112/Left_Ventricle.txt.gz:HiC/GSE87112/Liver.txt.gz:HiC/GSE87112/Lung.txt.gz:HiC/GSE87112/Ovary.txt.gz:HiC/GSE87112/Pancreas.txt.gz:HiC/GSE87112/Psoas.txt.gz:HiC/GSE87112/Right_Ventricle.txt.gz:HiC/GSE87112/Small_Bowel.txt.gz:HiC/GSE87112/Spleen.txt.gz:HiC/GSE87112/GM12878.txt.gz:HiC/GSE87112/IMR90.txt.gz:HiC/GSE87112/Mesenchymal_Stem_Cell.txt.gz:HiC/GSE87112/Mesendoderm.txt.gz:HiC/GSE87112/Neural_Progenitor_Cell.txt.gz:HiC/GSE87112/Trophoblast-like_Cell.txt.gz:HiC/GSE87112/hESC.txt.gz |
| **ciMapFileN** | 0 |

| Parameter | Value |
|---|---|
| **ciMapFiles** | NA |
| **ciMapFDR** | 1.00E-06 |
| **ciMapPromWindow** | 250-500 |
| **ciMapRoadmap** | E080:E029:E030:E031:E032:E033:E034:E035:E036:E037:E038:E039:E040:E041:E042:E043:E044:E045:E046:E047:E048:E050:E051:E062:E053:E054:E067:E068:E069:E070:E071:E072:E073:E074:E081:E082:E027:E028:E001:E002:E003:E008:E014:E015:E016:E024:E004:E005:E006:E007:E009:E010:E011:E012:E013:E023:E025:E063:E075:E076:E106:E077:E078:E079:E084:E085:E109:E101:E102:E103:E092:E094:E110:E111:E083:E095:E104:E105:E086:E066:E017:E088:E096:E052:E089:E100:E107:E108:E090:E097:E087:E098:E091:E099:E055:E056:E057:E058:E059:E061:E113:E026:E049:E093:E112:E065:E018:E019:E020:E021:E022 |
| **ciMapEnhFilt** | 1 |
| **ciMapPromFilt** | 1 |
| **ciMapCADDth** | 0 |

Note: any parameter marked N/A was deleted from this table.

# 3 Investigation of resilience in the UK biobank

## 3.1 Introduction

The UKB has genetic and cognitive data for over 500,000 individuals and is described in detail in chapter 2. In this chapter, I describe how I examined the available cognitive data to see which variables I could use to examine the genetic component of cognitive resilience in healthy ageing. I will explain the creation of a longitudinal measure of resilience using a proxy phenotype to measure cognitive performance in early adulthood (education years) and current cognitive measures to examine performance approximately 40 years later. We realised a confounding factor in this approach in that EY is highly heritability with a polygenic nature (Lee et al., 2018) that can mask the genetics of resilience. I therefore explore methods to extract the genetic variants associated with resilience alone.

## 3.2 Selection of Cognitive variables in the UKB

Due to lack of longitudinal data in the UK biobank, I needed to select a cogntive measure from the dataset to represent current cognitive performance and a proxy measure to represent past cognitive performance.

### 3.2.1 Current cognitive performance

Table 3.1 lists the cognitive variables in the UKB showing the number of participants assessed and the correlation of the variable to age. Participants undertook a wide range of cognitive tests (see **Error! Reference source not found.**). The types of tests and reliability are described elsewhere (Fawns-Ritchie & Deary, 2020; Lyall et al., 2018).

Of the cognitive tests performed at baseline, reaction time (RT) had the highest correlation with increasing age, where older age was associated with slower reaction time ($r = 0.27$, $P < 0.01$). In addition, RT was tested on nearly every individual in the dataset. When combined with the genetic data, there were 333,664 samples. A GWAS of this phenotype processed through FUMA showed 26 associated genetic loci mapping to 542 genes (Figure 3.1).

**Table 3.1: Correlation of cognitive tests with age**

| Test | n | r |
|---|---|---|
| **Initial 2007/2008** | | |
| Verbal Numerical reasoning | 165,486 | 0.05* |
| Reaction Time | 496,776 | 0.27* |
| Numeric Memory | 51,811 | 0.08* |
| Visual Pairs matching | 118,547 | 0.01* |
| Prospective memory | 171,569 | 0.01* |
| | | |
| **Online tests (2014/2015)** | | |
| Trail Making (#1) Online | 104,052 | 0.27* |
| Trail Making (#2) Online | 104,050 | 0.34* |
| Symbol Digit Substitution Online | 118,490 | 0.43* |
| Verbal Numerical reasoning Online | 123,665 | 0.12* |
| Numeric Memory Online | 111,086 | 0.13* |
| Principle Component analysis factor | 111,039 | 0.39* |

*Correlation is significant at the 0.01 level.

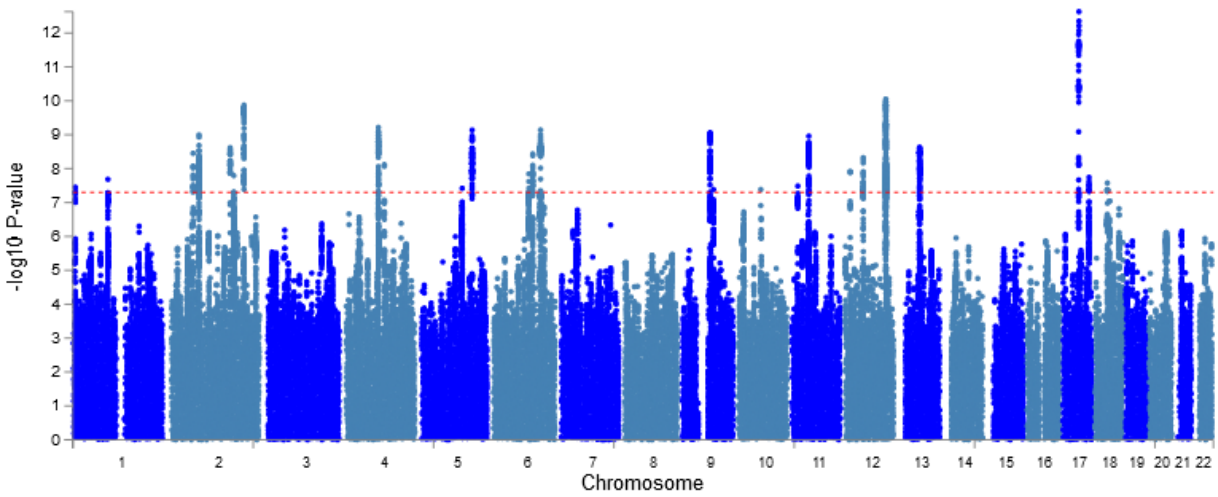



**Figure 3.1: GWAS of RT in UK biobank (n= 333,664)**
The Y axis shows the -log10 transformed P-values of each SNP from the GWAS. The x axis shows the base pair position along the chromosomes. The dotted read line shows the Bonferroni corrected P-value (P<5.0E-8).

In addition, some cognitive tests were performed online. I considered combining these to produce a measure of 'g' which measures overall intelligence using principal component analysis (PCA) in SPSS (see 2.1.4). The resultant variable showed a moderate correlation with age (r = 0.39, *P* <0.01) but was limited to 111,039 individuals. When combined with

genetic data, there were 66,740 participants in the dataset. Running a GWAS in plink 1.9 using this phenotype found just one significant locus on chromosome 16 (Figure 3.2).

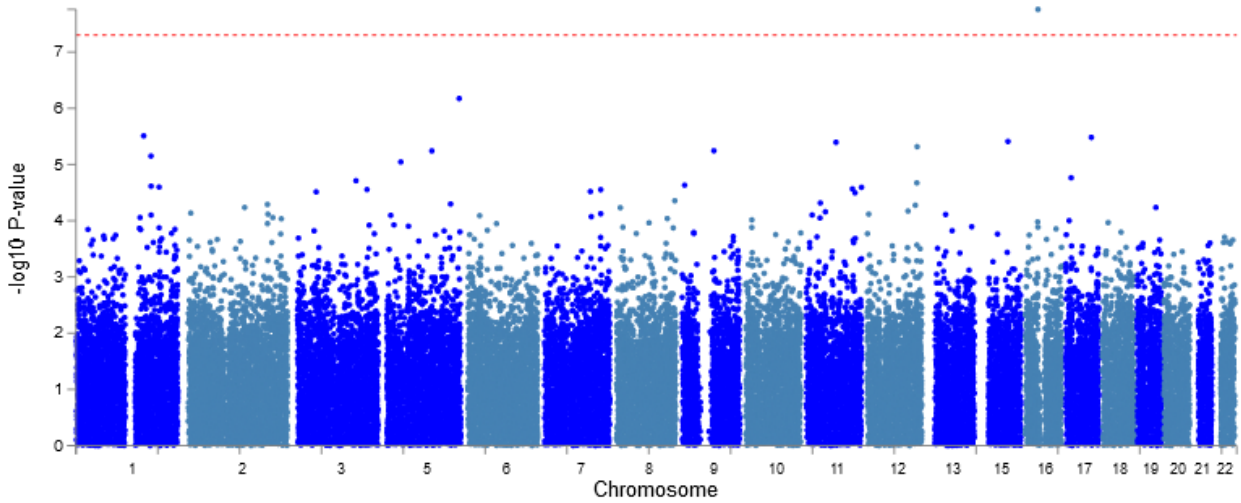



**Figure 3.2: GWAS of 'g' in UKB (n=66,740)**
The Y axis shows the -log10 transformed P-values of each SNP from the GWAS. The x axis shows the base pair position along the chromosomes. The dotted read line shows the Bonferroni corrected P-value (P<5.0E-8).

Given the stronger signal of RT vs 'g' and the association of processing speed with cognitive ageing (see 1.2.1), I decided to move forward using RT as my current cognitive measure.

### 3.2.2 Proxy measures

As described in 3.1, we first used RT, reflecting an individual's processing speed, as a measure of current cognitive performance. Processing speed is a key component, and predictor, of cognitive ability (Deary, 2013; Schubert et al., 2019). In the absence of a direct measure of processing speed at an earlier timepoint, we used academic achievement measured by number of years in education (education years (EY)) as a proxy phenotype for cognitive performance in early adulthood, following several previous studies (Davies et al., 2016; Plomin & von Stumm, 2018; Rietveld et al., 2014). Individual differences in processing speed are important in the relationship between executive functioning and academic performance (Gordon, Smith-Spark, Newton, & Henry, 2018). This approach captures an average time span of 40 years between past and current cognitive performance in UKB.

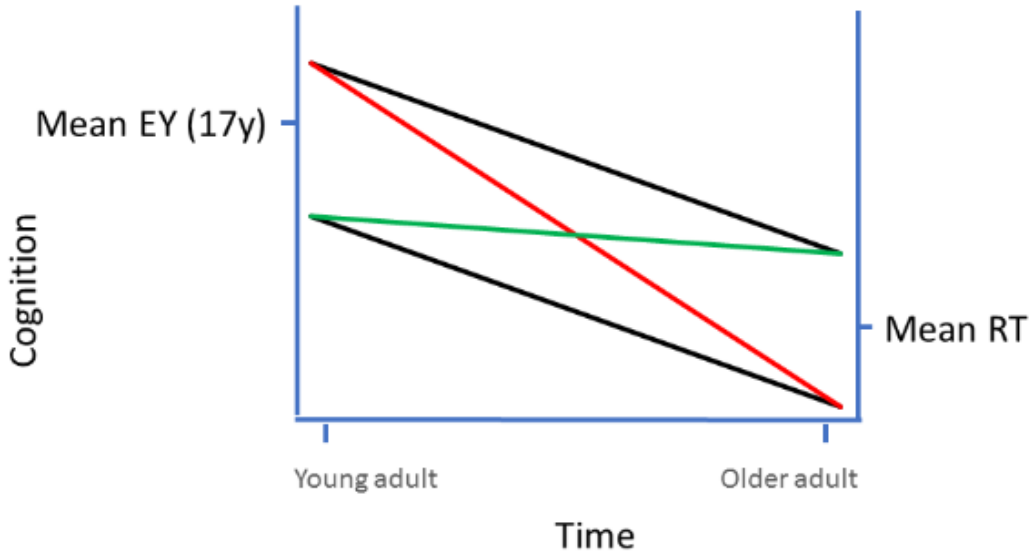
Educational attainment or EY have both been used in the past as a proxy for cognition. Using a combination of both variables within the UKB (see 2.1.4), I created an education years

(EY) variable. When matched with genotypic data there were 332,087 individuals available for further analysis.

### 3.2.3 Generation of the resilience phenotype

There were 330,098 individuals that have both EY and RT data. I created binary variables for both variables as described in 2.1.4. I used EY as a proxy phenotype measuring past cognitive performance. Processing speed as measured by RT was chosen as an indicator of current performance given its strong correlation with age and the fact that data was available on most participants in UKB. We created a binary variable for each measure by using the average score within the dataset to split the participants into similarly sized groups. EY was split into above and below average based on participants completing greater than or equal to 17 years, or less than 17 years in education. RT was corrected for age and normalised using a log 10 transformation and was split into faster and slower based on participants having a processing speed better or worse than the mean value, such that faster RT speeds reflected better processing speed, and thus cognitive performance, than slower RT. Using these two binary variables – above or below average EY and faster or slower RT – we created four groups of participants (Figure 3.3). One of these groups demonstrated high resilience and these were our cases for GWAS who had below average EY previously and faster than average RT now. A second group demonstrated low resilience or cognitive decline, and these were our controls for GWAS who had above average EY previously and slower than average RT now.

(a)




(b)

| Edu years | Processing speed | Group | N (330,089) |
|---|---|---|---|
| >17 | ↑ mean RT | Norm-high | 85,358 |
| <17 | ↓ mean RT | Norm -low | 88,728 |
| >17 | ↓ mean RT | Low resilience | 76,825 |
| <17 | ↑ mean RT | High resilience | 79,186 |

**Figure 3.3: Creation of resilience phenotype.**
Using the two binary variables – above or below average EY and faster or slower RT – we created four groups of participants. (a) Black lines show groups that behaved as expected. Green line is those that demonstrated resilience and red showed low resilience. (b) The green group demonstrated high resilience, and these were our cases for GWAS who had below average EY previously and faster than average RT now. A red group demonstrated low resilience or cognitive decline, and these were our controls for GWAS who had above average EY previously and slower than average RT now. The black groups were our cases and controls for those that did not demonstrate either high or low resilience.

### 3.2.4 GWAS of Resilience groups

To examine the genetics of resilience I performed a GWAS in Plink 2 (see 2.1.5) using the group that demonstrated high resilience and these were my cases for our first "Resilience" GWAS who had below average EY previously and faster than average RT now. A second group demonstrated low resilience or cognitive decline, and these were my controls for that GWAS who had above average EY previously and slower than average RT now. The two remaining groups of UKB samples displayed consistent cognitive performance over time. A Manhattan plot of this GWAS was generated in FUMA (Figure 3.4).

However, when I compare this GWAS with a GWAS of EY I noted very similar results (Figure 3.5). When I ran a correlation analysis, I found a very high correlation between Resilience and EY ($r_g$ =-0.9, $P = 1 \times 10^{-10}$). A confounding factor in this strategy is that EY is highly heritability with a polygenic nature (Lee et al., 2018) that can mask the genetics of resilience. I therefore recognised that our Resilience GWAS had a stong EY component and from there I refered to it as the "EY+Res" GWAS.

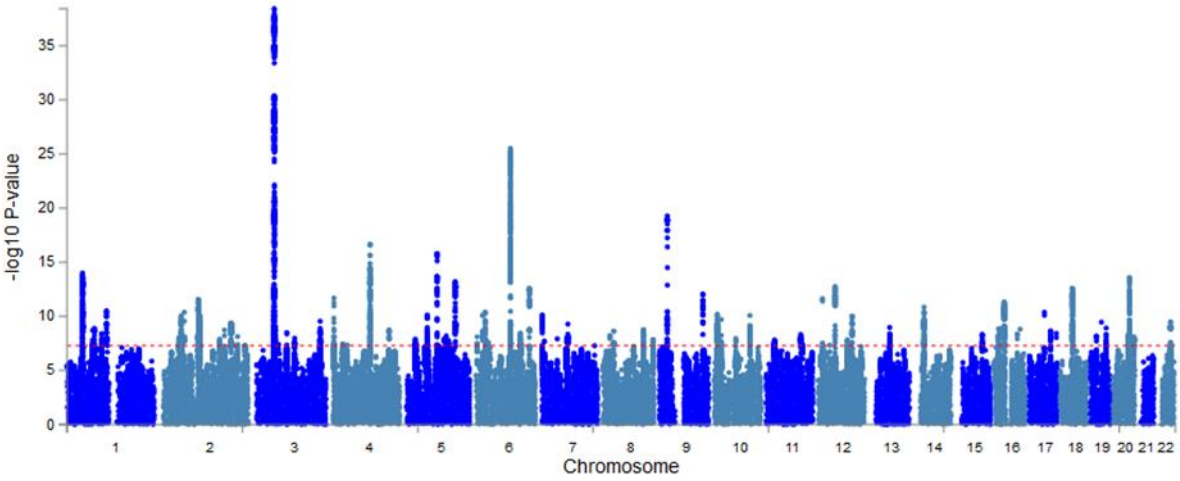



**Figure 3.4: Manhattan plot of GWAS of EY+Res (n=156,008)**
The y-axis shows the -log10 transformed P-values of each SNP from the GWAS. The x-axis shows the base pair position along the chromosomes. The dotted read line shows the Bonferroni corrected P-value (P<5.0E-8).

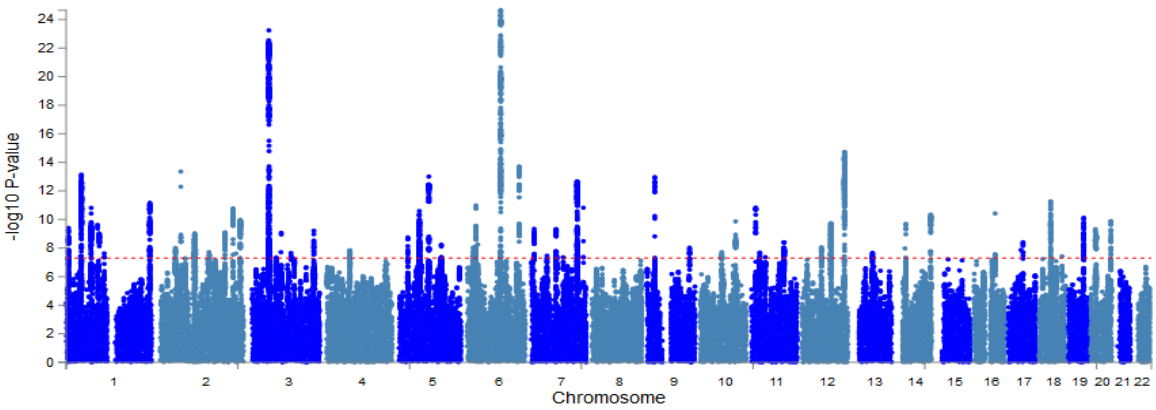



**Figure 3.5: Manhattan plot of GWAS of EY (n=165,000)**
The Y axis shows the -log10 transformed P-values of each SNP from the GWAS. The x axis shows the base pair position along the chromosomes. The dotted read line shows the Bonferroni corrected P-value (P<5.0E-8).

To identify those SNPs that were associated with resilience alone I needed to remove those SNPs that were associated with the EY component of the phenotype. To identify the EY component I performed a second GWAS using the two remaining groups of UKB samples that displayed consistent (i.e., unchanging) performance over time. The first of these groups consisted of those with below average EY previously and slower than average RT now (i.e., consistently below average performance over time); the second group consisted of those who showed above average EY previously and faster than average RT now (i.e., consistently above average performance over time). I named this GWAS "EY/NonRes" because it identified SNPs associated with EY but not resilience (Figure 3.6).

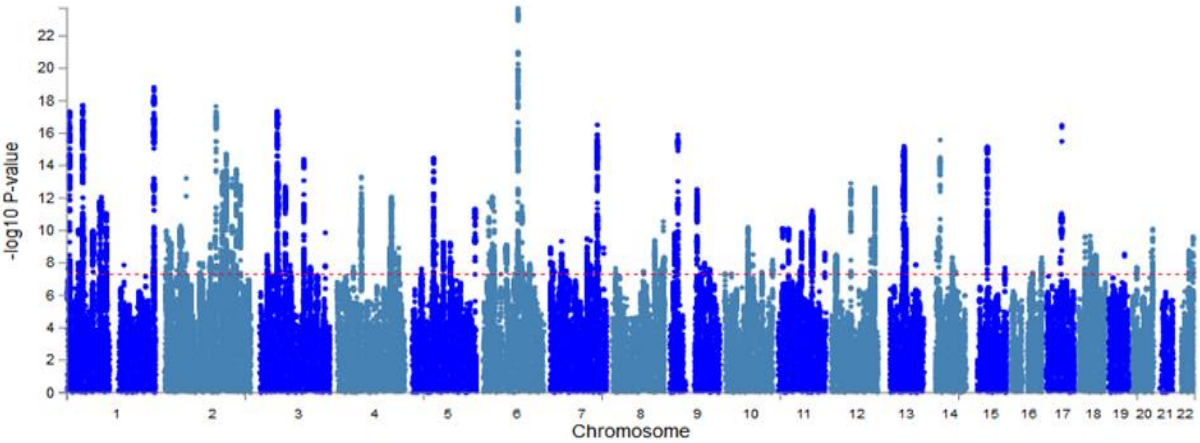



**Figure 3.6: Manhattan plot of GWAS of EY/NonRes (n=174,113)**
The y-axis shows the -log10 transformed P-values of each SNP from the GWAS. The x-axis shows the base pair position along the chromosomes. The dotted read line shows the Bonferroni corrected P-value (P<5.0E-8).

### 3.2.5 Extraction of Res from EY+Res

To identify the genetic component of resilience alone it was necessary to look at the difference between EY+Res and EY/NonRes. In other words, perform the following:

EY+Res minus EY/NonRes = *Resilience*

Initially I examined how we could perform this subtraction using the analysis of both GWAS in FUMA.

#### 3.2.5.1 Functional level

I extracted the mapped genes from FUMA using the FUMA default parameters for EY+Res and EY/NonRes. EY+Res had 320 mapped genes and EY/NonRes had 423 mapped genes. A total of 107 genes were common to both GWAS (Figure 3.7). I also examined overlap between GO terms generated in FUMA and through processing non overlapped genes through consensus path DB.

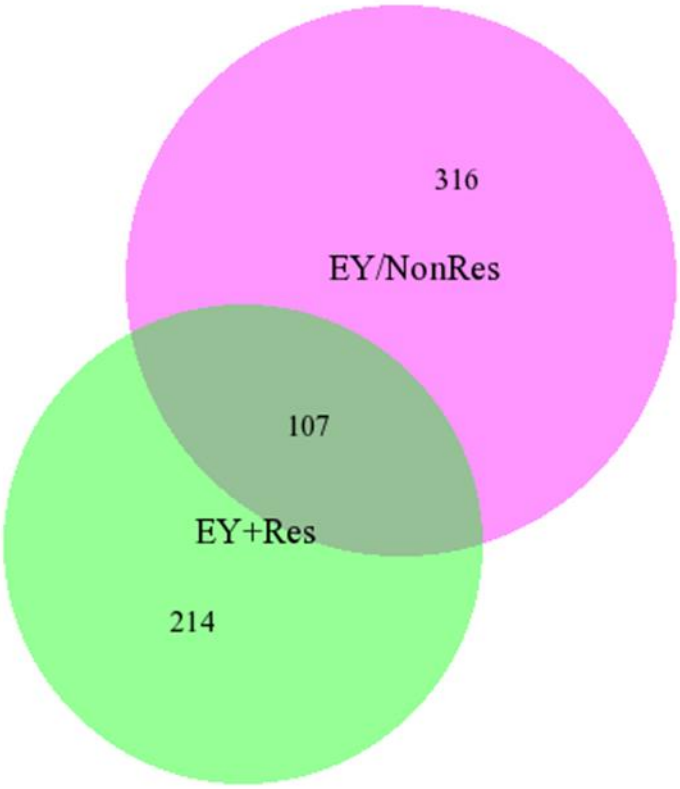



**Figure 3.7: Overlap of mapped genes in EY+Res and EY/NonRes using BioVenn**

#### 3.2.5.2 SNP level

To identify those SNPs associated with resilience alone I subtracted all SNPs with a nominally significant P value (P<.05) in EY/NonRes (n=1102,873) from significant SNPs (P<0.5 x $10^{-8}$) in EY+Res (n=5013). This results in a list of 1,431 SNPs that were significantly associated with *Resilience* with the interference from low/high EY removed. This SNP list was submitted to FUMA for analysis. As there are no SNPs with a P value greater than 0.5 x $10^{-8}$, the output is a partial Manhattan plot (Figure 3.8) These SNPs are clumped together into 50 loci (Table 3.2)

.

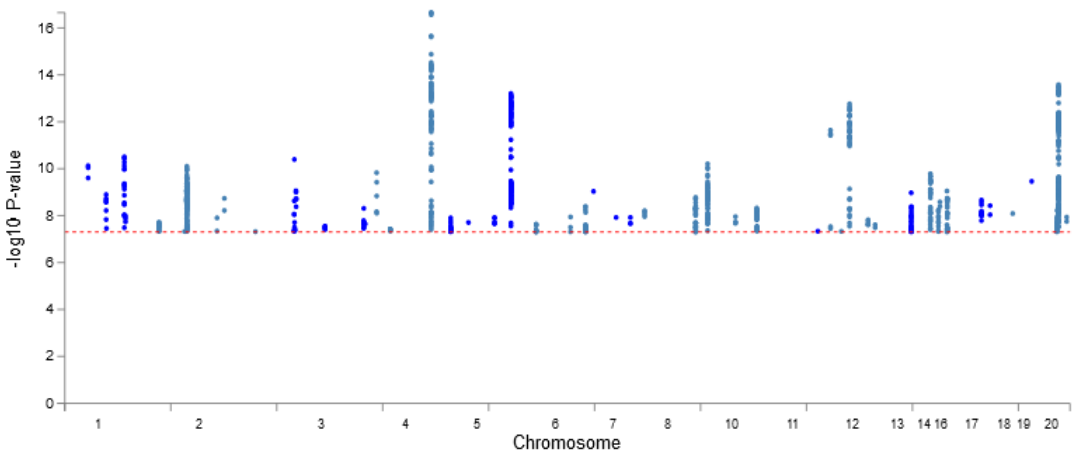



**Figure 3.8: Manhattan plot of *Resilience* generated by subtraction of EY/NonRes SNPs from EY+Res SNPs**
The y-axis shows the -log10 transformed P-values of each SNP from the GWAS. The x-axis shows the base pair position along the chromosomes. The dotted read line shows the Bonferroni corrected P-value (P<5.0E-8).

**Table 3.2 : Genetic loci identified by manual subtraction of EY/NonRes SNPs from EY+Res SNPs.**

| Locus | rsID | chr | position | p | start | end | SNPs |
|---|---|---|---|---|---|---|---|
| **1** | rs2494994 | 1 | 43924166 | 7.45E-11 | 43760236 | 43949718 | 3 |
| **2** | rs2647499 | 1 | 77949806 | 1.27E-09 | 77919310 | 78687156 | 8 |
| **3** | rs197374 | 1 | 112289983 | 3.13E-11 | 112273485 | 112328245 | 20 |
| **4** | rs74404380 | 1 | 114488628 | 1.02E-08 | 114467664 | 114506786 | 3 |
| **5** | rs1400101 | 2 | 51896988 | 1.89E-08 | 51816988 | 51936388 | 15 |
| **6** | rs778177929 | 2 | 100050426 | 4.75E-08 | 99573471 | 100109251 | 1 |
| **7** | rs9308868 | 2 | 104361420 | 7.90E-11 | 104056454 | 104479862 | 272 |
| **8** | rs13010673 | 2 | 161353564 | 1.25E-08 | 161080841 | 161367714 | 2 |
| **9** | rs34372833 | 2 | 175200014 | 1.84E-09 | 175199869 | 175200903 | 2 |
| **10** | rs2675964 | 2 | 233793676 | 4.80E-08 | 233698457 | 233807585 | 1 |
| **11** | rs17225749 | 3 | 50131140 | 4.03E-11 | 48793504 | 51549024 | 12 |
| **12** | rs77321042 | 3 | 53179543 | 8.74E-10 | 52492601 | 53556816 | 5 |
| **13** | rs113013592 | 3 | 108014239 | 2.83E-08 | 107864673 | 108031094 | 6 |
| **14** | rs13061117 | 3 | 181186466 | 4.89E-09 | 180524952 | 181205593 | 11 |
| **15** | rs843370 | 3 | 183875822 | 2.27E-08 | 183861243 | 183922076 | 1 |
| **16** | rs112261906 | 4 | 3050001 | 1.48E-10 | 2935618 | 3385176 | 5 |
| **17** | rs10014342 | 4 | 28751389 | 3.74E-08 | 28733631 | 28753519 | 7 |
| **18** | rs2454202 | 4 | 106205280 | 2.22E-17 | 106048291 | 106429390 | 90 |
| **19** | rs7718191 | 5 | 25895568 | 1.25E-08 | 25892558 | 26046971 | 28 |
| **20** | rs7717864 | 5 | 59486768 | 1.96E-08 | 59281617 | 59562579 | 1 |
| **21** | rs2914685 | 5 | 108774745 | 1.21E-08 | 108641351 | 108881782 | 5 |
| **22** | rs13186198 | 5 | 139652057 | 6.20E-14 | 139515394 | 139714837 | 193 |
| **23** | rs9469529 | 6 | 33587443 | 2.29E-08 | 33573994 | 33587443 | 9 |
| **24** | rs4839936 | 6 | 98489654 | 1.14E-08 | 98349189 | 98532334 | 3 |
| **25** | rs9398803 | 6 | 126683594 | 4.13E-09 | 126659043 | 127080700 | 22 |
| **26** | rs4719416 | 7 | 2106608 | 9.26E-10 | 1906466 | 2110850 | 1 |
| **27** | rs3735478 | 7 | 44800176 | 1.19E-08 | 44765133 | 44800176 | 1 |
| **28** | rs2944826 | 7 | 71792250 | 1.20E-08 | 71683050 | 71852951 | 3 |
| **29** | rs4392917 | 8 | 19680635 | 6.16E-09 | 19627586 | 19680979 | 4 |
| **30** | rs78126454 | 8 | 115967342 | 1.66E-09 | 115869157 | 116043658 | 35 |
| **31** | rs7903084 | 10 | 10861098 | 6.20E-11 | 10630010 | 11140234 | 64 |
| **32** | rs11599801 | 10 | 63447289 | 1.11E-08 | 63425243 | 63524979 | 4 |
| **33** | rs79550344 | 10 | 103613098 | 4.73E-09 | 103379885 | 104198528 | 41 |
| **34** | rs10895704 | 11 | 104395746 | 4.62E-08 | 104320214 | 104477159 | 1 |
| **35** | rs10772644 | 12 | 13417617 | 2.29E-12 | 13414139 | 13463954 | 5 |
| **36** | rs16921317 | 12 | 33979400 | 4.70E-08 | 33548001 | 34674292 | 1 |
| **37** | rs6580699 | 12 | 49478812 | 1.72E-13 | 49387955 | 49479968 | 38 |
| **38** | rs35638842 | 12 | 84304740 | 1.52E-08 | 84211940 | 84539271 | 5 |
| **39** | rs12831306 | 12 | 97664962 | 2.45E-08 | 97653869 | 97683185 | 2 |
| **40** | rs9569733 | 13 | 58316612 | 1.06E-09 | 58250322 | 58828747 | 25 |
| **41** | rs8019612 | 14 | 29822421 | 1.63E-10 | 29737713 | 29854122 | 40 |

| Locus | rsID | chr | position | p | start | end | SNPs |
|---|---|---|---|---|---|---|---|
| **42** | rs11647572 | 16 | 12211417 | 5.88E-09 | 12159290 | 12219411 | 12 |
| **43** | rs11644601 | 16 | 15172118 | 2.61E-09 | 14961204 | 15255751 | 2 |
| **44** | rs2650494 | 16 | 28318440 | 8.95E-10 | 28299132 | 29001460 | 24 |
| **45** | rs6504165 | 17 | 61630076 | 2.17E-09 | 61606190 | 61747965 | 12 |
| **46** | rs870681 | 17 | 77773876 | 3.76E-09 | 77773876 | 77778226 | 2 |
| **47** | rs117623407 | 19 | 32204489 | 3.46E-10 | 32200518 | 32208909 | 1 |
| **48** | rs6073984 | 20 | 44630653 | 5.97E-09 | 44447625 | 44656140 | 28 |
| **49** | rs779402116 | 20 | 47530801 | 2.62E-14 | 47508077 | 47933479 | 337 |
| **50** | rs61746505 | 20 | 62839710 | 1.17E-08 | 62836637 | 62841415 | 2 |

### 3.2.6 Extraction of Resilience using SEM

In section 3.2.5, I described how we extracted data on *Resilience* by a manual subtraction that left us with a partial GWAS. This method was very limited in that I could not examine a full GWAS of *Resilience* and perform full functional analysis as with other published traits. In addition, this method is based on P values only and does not take odds ratios into account. It also eliminates all SNPs in *Resilience* that have a nominal association in EY/NonRes but may be highly significant in EY+Res. To overcome this I employed a recently published method which uses Genomics Structural Equation Modelling (GenomicSEM) (Grotzinger et al., 2019) to perform a GWAS-by-subtraction (Demange et al., 2021) using two GWAS, one which captured genetic variants associated with EY and resilience (EY + Res as described in 3.2.4) and a second which captured genetic variants associated with EY but not resilience (EY/NonRes). Subtracting one from the other generated two new GWAS, one capturing EY and the other capturing what I wanted - the genetics of a processing speed-based cognitive resilience phenotype.

#### 3.2.6.1 Initial phenotype development

Figure 3.9 shows an overview of the analysis steps using the initial steps already outlined in Section 3.2.3 and a detailed description of the process used to generate the resilience variable is included in the Materials and Methods. Given the multi-step method proposed in this analysis, I sought to confirm findings using our method in an independent sample. Therefore, I divided the UKB into discovery (n=266,543; 81% of participants) and replication (n=63,554; 19% of participants) samples (Figure 3.9a). Sample sizes used for analysis are shown in Table 3.3.

**Table 3.3: Sample size for GWAS analysis by cohort**

| **Sample** | **Replication n=63,554** | | **Discovery n=266,543** | | **Full n=330,097** | |
|---|---|---|---|---|---|---|
| | EY+Res | EY/NonRes | EY+Res | EY/NonRes | EY+Res | EY/NonRes |
| Case | 13,893 | 14,980 | 65,293 | 73,749 | 79,186 | 88,728 |
| Control | 16,145 | 18,536 | 60,679 | 66,822 | 76,825 | 85,358 |
| Total | 30,038 | 33,516 | 125,972 | 140,571 | 156,011 | 174,086 |

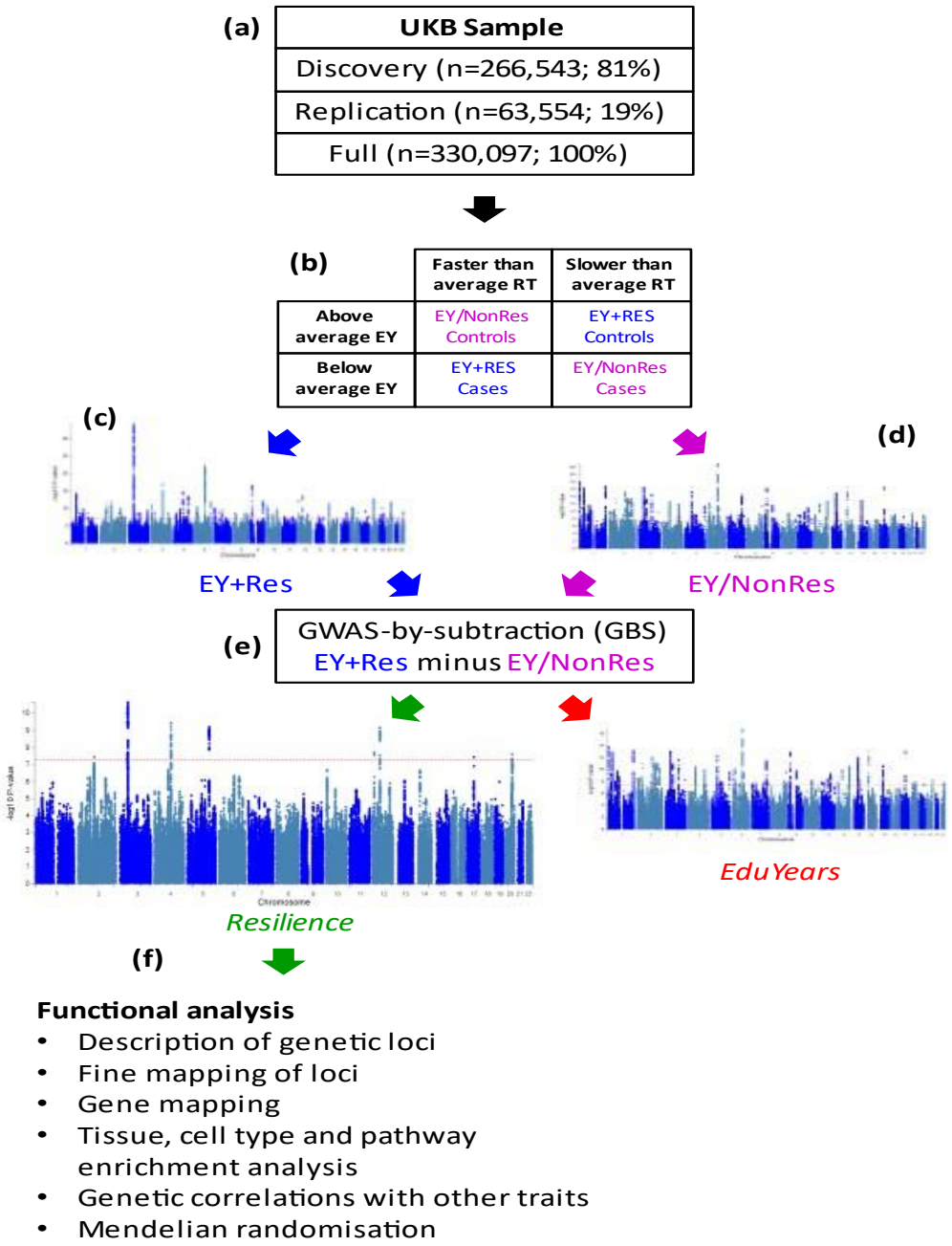



**Figure 3.9: Flow chart of study design.**
(a) The available UKB samples were split into Discovery (81%) and Replication (19%) samples. Following successful replication analysis, the Full sample was also put through the analysis pipeline. (b) For Discovery, Replication or Full, samples were assigned to one of four categories based on their EY and RT measures. (c) EY+Res cases and controls were analysed in a GWAS. (d) EY/NonRes cases and controls were analysed in a GWAS. (e) GBS used to subtract the genetic signals for EY/NonRes from EY+Res to result in a Resilience GWAS and an EduYears GWAS. (f) Resilience GWAS functionally analysed to identify associated SNPs and genes, and enriched tissues, cell types and pathways, identify genetic correlations with other traits and explore causal relationships between resilience and other traits using Mendelian randomisation.

I used GWAS-by-subtraction (GBS) (Demange et al., 2021) to subtract the results of EY/NonRes from EY+Res to leave SNP associations with resilience. This method uses GenomicSEM (Grotzinger et al., 2019) to integrate GWAS into structural equation modelling. Following the process described by Demange et al (Demange et al., 2021), I defined a Cholesky model using the summary statistics from the EY+Res and EY/NonRes GWASs. Both EY+Res and EY/NonRes were regressed on a latent factor, which captured the shared genetic variance in EY (hereafter "*EduYears"*). EY+Res was further regressed on a second latent factor capturing the variance in EY+Res independent of EY/NonRes, hereafter "*Resilience*". Genetic variance in *Resilience* was independent of genetic variance in *EduYears* ($r_g = 0$) as the *Resilience* factor represents residual genetic variation in our EY+Res phenotype that is not accounted for by the *EduYears* factor. These two latent variables, *Resilience* and *EduYears* were then regressed on each SNP in the original GWASs (EY+Res and EY/NonRes) resulting in new GWAS summary statistics for both *Resilience* and *EduYears* (Figure 3.9e). Figure 3.10 shows the Cholesky model with the pathway loadings for the full GWAS. The loading for the different sample sizes is described in Table 2.5.

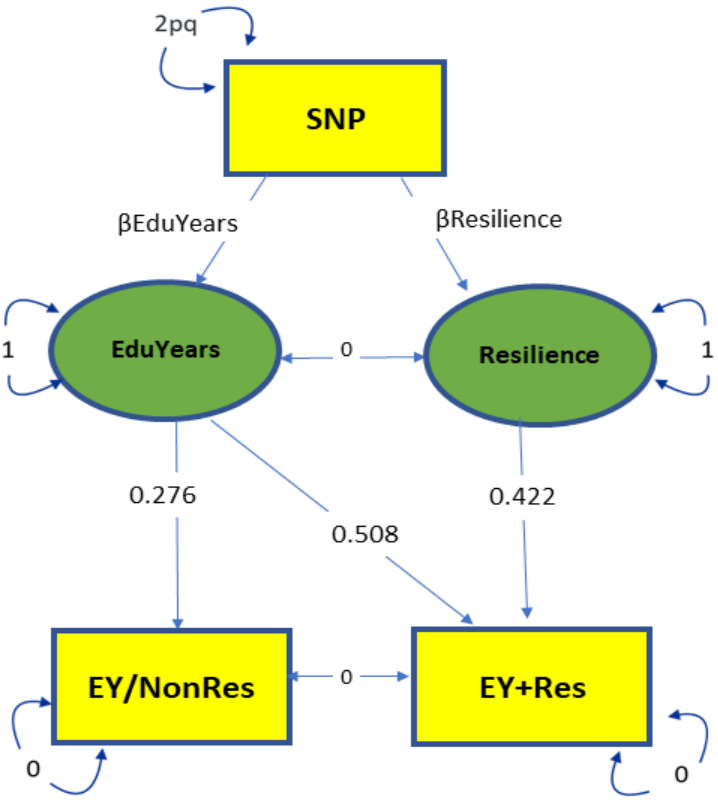



**Figure 3.10:SEM of GWAS-by-subtraction.**
The observed variables are the GWAS EY+Res and EY/NonRes and SNP and the latent variables (unknown) are *Resilience* and *EduYears*. There are two pathways for the SNPs analysis in this model to EY+Res – the first is through *EduYears* to EY+Res and EY/NonRes and incorporates the genetic effects of the variables used in the phenotype. The other path is through *Resilience* to EY+Res and measures the genetic effect of resilience independent of *EduYears*. To calculate the model, the genetic covariances between EY+Res and EY/NonRes and *Resilience* and *EduYears* are set to 0 and the variances of EY+Res and EY/NonRes are also set to 0. The variance is therefore explained by the latent factors. The SNP value is calculated as 2pq from allele frequencies of the 1000 Genome phase 3 data where p is the reference allele and q the alternative allele. This figure shows the pathway loadings for the full sample. See Figure 2.2 for the model without specific loadings.

### 3.2.7 Discovery and replication analysis

For the discovery sample, I performed the two initial GWASs (discovery.EY+Res and discovery.EY/NonRes) and then performed GBS on both sets of samples resulting in *discovery.Resilience* GWAS results and *discovery.EduYears* GWAS results. I repeated this for the replication sample to produce *replication.Resilience* GWAS results and *replication.EduYears* GWAS results. Comparison of the *discovery.Resilience* GWAS with the *replication.Resilience* GWAS by LDSR analysis (Grotzinger et al., 2019) showed

extremely high correlation between the two datasets ($r_g$ = 0.964, $P$ = 4.45 x $10^{-44}$). The *discovery.Resilience* GWAS was then processed through FUMA v 1.3.6 (Watanabe et al., 2017) and ten independent genome-wide significant SNPs were identified. When compared to the *replication.Resilience* GWAS, there was a consistent direction of effect for all ten SNPs (Binomial sign test, $P$ = 9.77 x $10^{-4}$). Five of the ten SNPs were significant after Bonferroni multiple test correction ($P$ < 0.005). Thus, we demonstrated that we could replicate genetic associations with *Resilience* in an independent sample. Results for the ten-independent genome-wide significant SNPs and their replication analysis are in Table 3.4.

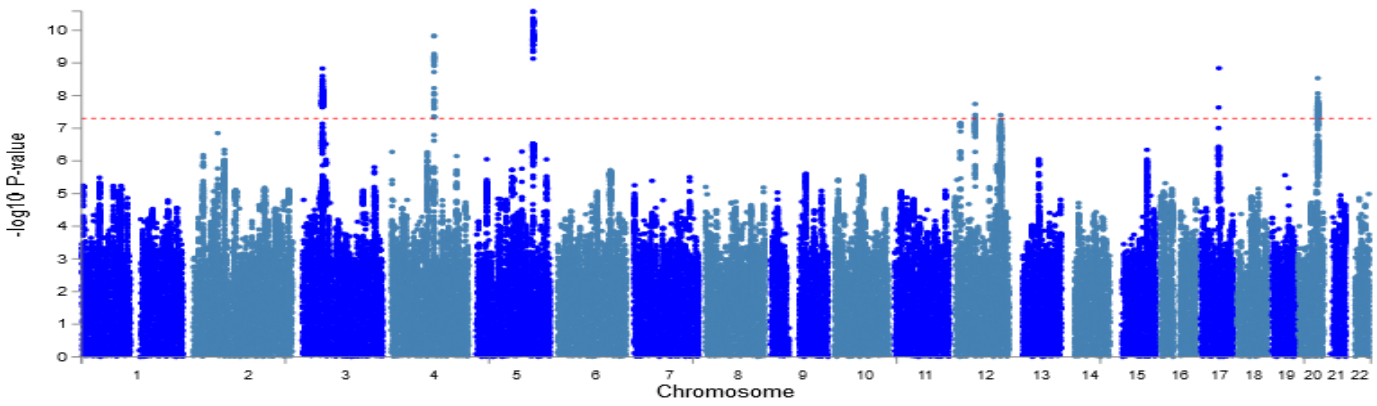
**Table 3.4: Replication of GWAS-by-Subtraction: Performed on two subsets of the UKB**

| | | | | | **Discovery [a]Neff =88,607** | | **Replication [a]Neff=25,706** | |
|---|---|---|---|---|---|---|---|---|
| **SNP** | **Chr** | **Maf** | **A1** | **A2** | **[b]Beta** | **P Value** | **[b]Beta** | **[c]P Value** |
| rs1043357 | 20 | 0.41 | C | A | 0.068 | 8.49E-09 | 0.044 | 2.20E-02 |
| rs1054442 | 12 | 0.37 | C | A | -0.068 | 1.81E-08 | -0.058 | **3.08E-03** |
| rs11065967 | 12 | 0.21 | G | T | 0.080 | 3.93E-08 | 0.046 | 4.69E-02 |
| rs1347143 | 5 | 0.44 | A | G | 0.079 | 2.59E-11 | 0.018 | 3.23E-01 |
| rs2189234 | 4 | 0.38 | T | G | 0.077 | 1.46E-10 | 0.056 | **3.95E-03** |
| rs2352974 | 3 | 0.49 | T | C | 0.071 | 1.48E-09 | 0.110 | **3.32E-08** |
| rs2426132 | 20 | 0.46 | C | G | -0.070 | 2.93E-09 | -0.029 | 1.18E-01 |
| rs2624824 | 3 | 0.48 | C | T | -0.068 | 6.82E-09 | -0.104 | **1.43E-07** |
| rs62074125 | 17 | 0.25 | C | A | -0.082 | 1.44E-09 | -0.055 | 1.16E-02 |
| rs7225002 | 17 | 0.42 | G | A | 0.066 | 2.30E-08 | 0.060 | **1.91E-03** |

Notes: [a]Neff = effective N which is the sample size adjusted for GWAS-By-Subtraction, [b]Beta = effect size of the A1 (minor) allele, [c]highlighted in bold if significant after Bonferroni correction (P<0.005).

Manhattan plots for the discovery and replication samples are shown in Figure 3.11.

(a)




(b)

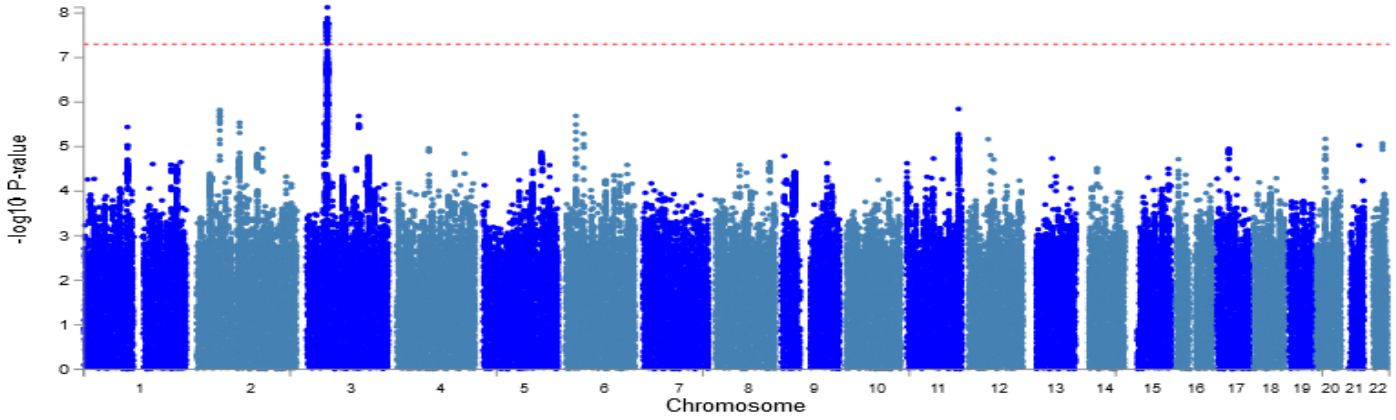



**Figure 3.11: Manhattan plot of *discovery.Resilience* (a) and *replication.Resilience* (b)**
The y-axis shows the -log10 transformed P-values of each SNP from the GWAS. The x-axis shows the base pair position along the chromosomes. The dotted read line shows the Bonferroni corrected P-value (P<5.0E-8).

### 3.2.8 Analysis of the full sample

Next, we combined both the discovery and replication samples to run an analysis on the full sample (n=330,097). This resulted in initial EY+Res and EY/NonRes GWASs and following GBS, *Resilience* GWAS results and *EduYears* GWAS results. SNP based heritability estimate analysis showed a $h^2$ value of 0.13 (SE = .006) for *Resilience*. For comparison in similarly sized samples, we also ran GWASs of EY and RT using participants randomly selected from UKB (EY, n=82,000 above average EY cases and n=81,999 below average EY controls; RT, n=82,000 faster than average RT cases and n=82,000 slower than average RT controls). These comparisons are shown in Table 3.5. This comparison shows that in similar sized samples, that one lead SNPs in *Resilience* is significant in EY (rs2352974, Chr 3, $P < 3.00 \times 10^{-21}$) and a separate one in RT (rs62074125. Chr 17, $P < 7.54 \times 10^{-09}$).

A Quantile-quantile (Q-Q) plot of *Resilience* and *EduYears*, on the full sample is shown in Figure 3.12. Manhattan plots are available in Figure 3.13. The other five GWAS (EY+Res, EY/NonRes, EY and RT) have been shown previously in this chapter.

a b

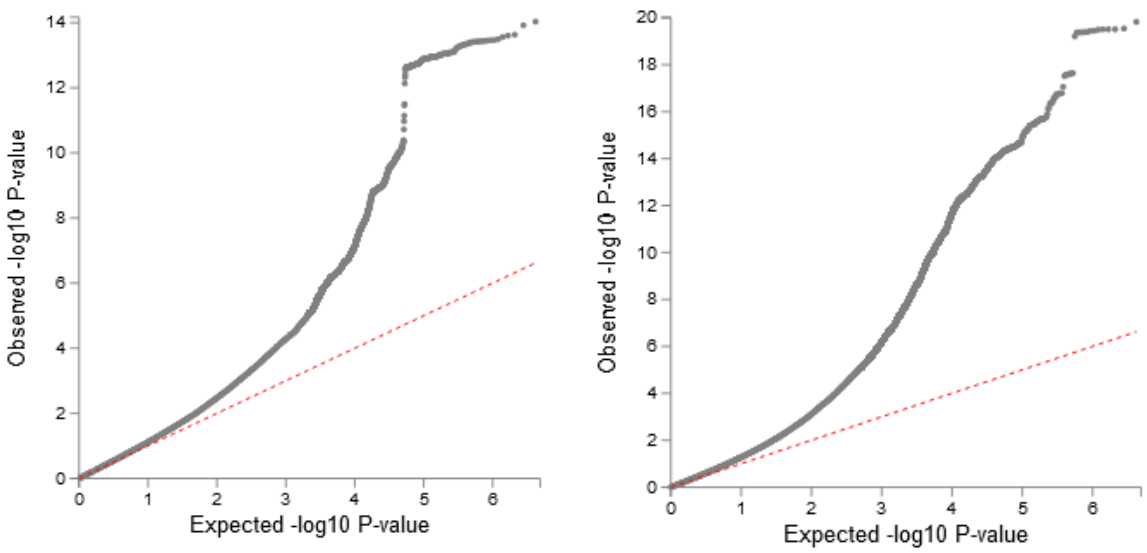



**Figure 3.12: Q-Q plots of *Resilience* GWAS (a) and *EduYears* GWAS (b)**
The x-axis shows the expected distribution of p-values from the GWAS across all SNPs, and the y-axis shows the observed p-values. The genomic inflation factor (lambda) as calculated by LDSC is 1.25 for Resilience and 1.50 for EduYears.

**Table 3.5: Examining the results of effect sizes and P values for 13 SNPs in the contributing GWAS analysis using the independent significant SNPs in *Resilience.***

| | Resilience [a]Neff = 111,316 | | | EY n=163,999 | | RT n=164,000 | | RT n=331,487 | | EY n=330,000 | | EduYears [a]Neff =165,000 | | EY+Res n=157,000 | | EY/NonRes n=176,000 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **SNP** | **Chr** | **[b]Beta** | **P value** | **[b]OR** | **P value** | **[b]OR** | **P value** | **OR** | **P** | **OR** | **P** | **[b]Beta** | **P value** | **[b]OR** | **P value** | **[b]OR** | **P value** |
| **rs12474507** | 2 | -0.06 | 4.19E-08 | 1.01 | 4.34E-01 | 0.98 | 7.89E-04 | 0.97 | 8.66E-09 | 1.01 | 2.01E-01 | 0.02 | 5.82E-03 | 0.96 | 4.79E-07 | 1.02 | 2.88E-03 |
| **rs2352974** | 3 | 0.08 | 9.27E-15 | 0.93 | 3.00E-21 | 1.02 | 6.38E-03 | 1.02 | 1.90E-03 | 0.93 | 1.04E-45 | 0.05 | 6.96E-11 | 1.10 | 1.09E-38 | 1.05 | 1.29E-12 |
| **rs2189234** | 4 | 0.07 | 3.14E-12 | 0.98 | 8.53E-04 | 1.02 | 1.72E-03 | 1.03 | 7.63E-09 | 0.97 | 5.66E-08 | 0.00 | 7.67E-01 | 1.06 | 2.22E-16 | 1.00 | 7.49E-01 |
| **rs6857847** | 4 | -0.08 | 4.99E-09 | 1.01 | 4.13E-01 | 0.96 | 3.46E-05 | 0.96 | 8.06E-10 | 1.01 | 4.12E-01 | 0.03 | 9.16E-04 | 0.95 | 2.25E-07 | 1.03 | 3.37E-04 |
| **rs56335290** | 5 | -0.07 | 2.59E-08 | 1.00 | 9.59E-01 | 0.97 | 2.33E-03 | 0.97 | 3.80E-08 | 1.01 | 1.51E-01 | 0.02 | 7.94E-03 | 0.95 | 2.06E-07 | 1.02 | 4.12E-03 |
| **rs6870103** | 5 | 0.07 | 4.03E-11 | 0.98 | 4.86E-03 | 1.02 | 1.29E-03 | 1.03 | 7.30E-10 | 0.98 | 1.18E-05 | -0.01 | 2.99E-01 | 1.06 | 2.00E-13 | 0.99 | 2.62E-01 |
| **rs7747481** | 6 | 0.06 | 2.29E-08 | 0.95 | 3.21E-13 | 1.01 | 9.45E-02 | 1.01 | 2.75E-02 | 0.95 | 4.86E-25 | 0.04 | 6.72E-07 | 1.07 | 3.25E-21 | 1.04 | 7.10E-08 |
| **rs1029388** | 12 | 0.07 | 6.50E-09 | 1.01 | 1.91E-01 | 1.04 | 6.22E-06 | 1.04 | 8.84E-11 | 1.01 | 3.55E-01 | -0.04 | 2.00E-06 | 1.04 | 1.43E-05 | 0.96 | 2.57E-07 |
| **rs2417261** | 12 | 0.09 | 1.35E-08 | 0.95 | 2.29E-06 | 1.03 | 1.71E-02 | 1.04 | 2.11E-06 | 0.96 | 4.59E-07 | 0.00 | 7.43E-01 | 1.09 | 3.24E-12 | 1.00 | 7.24E-01 |
| **rs6580699** | 12 | -0.07 | 1.29E-10 | 1.03 | 2.81E-04 | 0.97 | 7.71E-05 | 0.97 | 5.55E-09 | 1.03 | 6.19E-07 | 0.00 | 5.43E-01 | 0.95 | 1.72E-13 | 1.00 | 5.12E-01 |
| **rs9569811** | 13 | -0.09 | 1.14E-08 | 1.01 | 3.05E-01 | 0.97 | 2.62E-03 | 0.96 | 2.13E-07 | 1.02 | 1.24E-02 | 0.02 | 8.42E-02 | 0.94 | 4.25E-09 | 1.02 | 6.22E-02 |
| **rs62074125** | 17 | -0.08 | 8.31E-11 | 1.00 | 6.46E-01 | 0.95 | 7.54E-09 | 0.96 | 8.45E-13 | 1.00 | 7.77E-01 | 0.04 | 1.90E-06 | 0.96 | 2.10E-07 | 1.04 | 2.42E-07 |
| **rs4810896** | 20 | 0.07 | 1.94E-10 | 0.97 | 3.19E-04 | 1.03 | 3.67E-04 | 1.03 | 7.81E-07 | 0.97 | 4.30E-07 | 0.00 | 9.46E-01 | 1.06 | 3.22E-14 | 1.00 | 9.42E-01 |

Notes: The P and effect values of the genomic loci for Resilience and EduYears are compared to the values generated for the inputs to GBS (EY+Res and EY/NonRes) and to GWAS of the two variables used to create the phenotype (RT and EY). RT is examined using both a dichotomised sample similar in size to *Resilience* and using the full dataset (n=331,487). [a]Neff = effective N which is the sample size adjusted for GWAS-By-Subtraction, [b]Beta/ OR = effect size of the A1 (minor) allele.

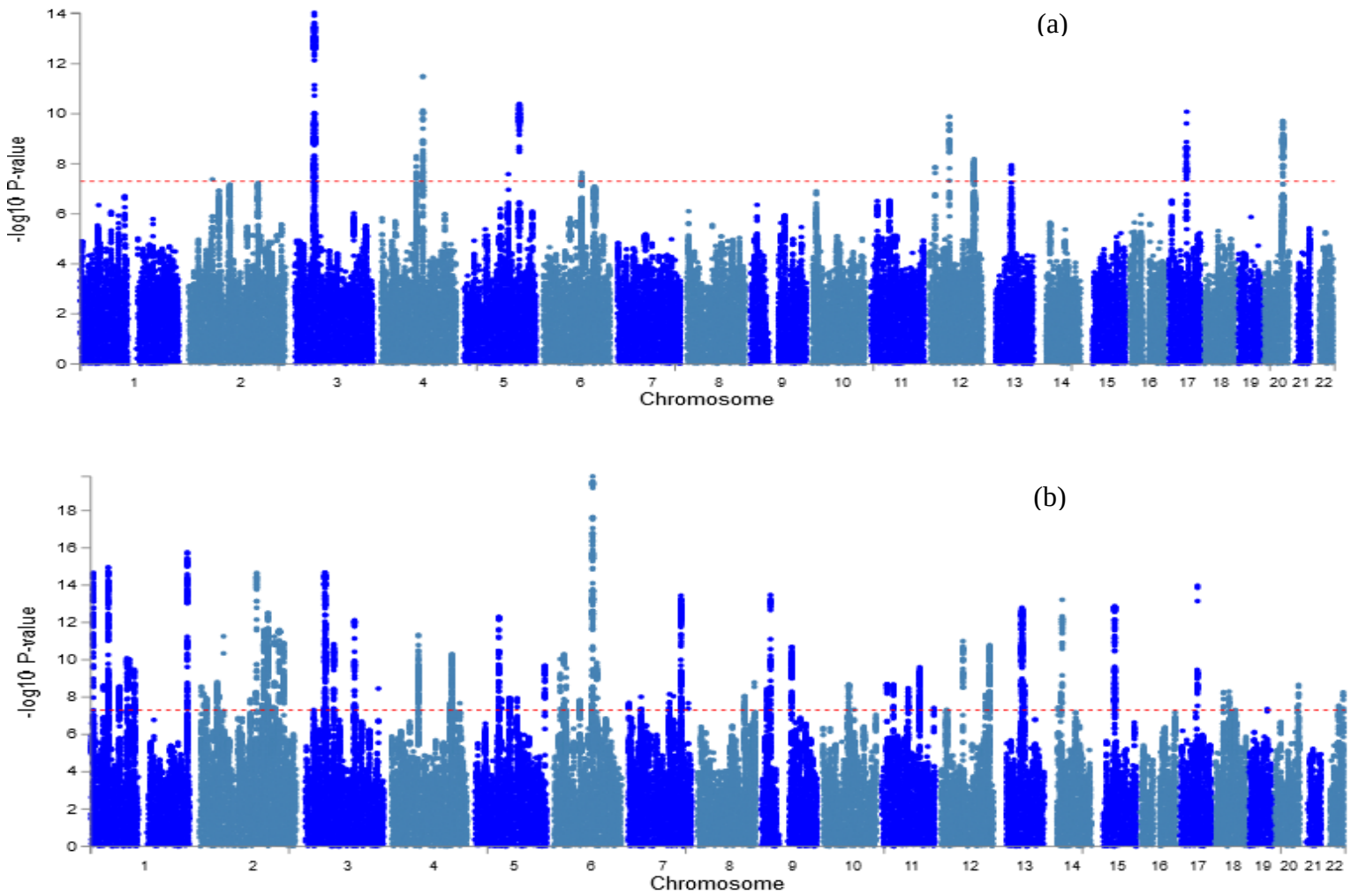



**Figure 3.13: Manhattan plots of *Resilience* GWAS (a) and *EduYears* GWAS (b).**
The y-axis shows the -log10 transformed P-values of each SNP from the GWAS. The x-axis shows the base pair position along the chromosomes. The dotted read line shows the Bonferroni corrected P-value (P<5.0E-8).

### 3.2.9 Correlation analysis

Initially, both EY+Res and EY/NonRes had a strong negative correlation with EY ($r_g$ = -0.88 and $r_g$ = -0.89 respectively (Figure 3.14). The strength of these correlations likely reflects the major contribution of EY to these phenotypes and they are negative because for EY+Res and EY/NonRes, the direction of effect is in the opposite direction to EY, as the cases are low EY whereas for the EY GWAS, the cases are high EY. EY+Res and EY/NonRes had a moderate positive correlation with each other ($r_g$ = 0.54). After GBS there was no genetic correlation between *Resilience* and *EduYears* ($r_g$ = 0.01, $P$ = 0.803) suggesting that the subtraction had successfully separated out the genetic associations for both phenotypes.

Although the EY component of *Resilience* was addressed by the GBS method, the RT component was not and the genetic correlation between *Resilience* and RT was strong ($r_g$ = 0.80; Figure 3.14). This finding was examined further following functional analysis of associated loci and is documented in chapter 5.

| | *Resilience* | *EduYears* | EY+Res | EY/NonRes | EY | RT |
|---|---|---|---|---|---|---|
| *Resilience* | | | | | | |
| *EduYears* | 0.01* | | | | | |
| EY+Res | 0.84 *** | 0.55*** | | | | |
| EY/NonRes | 0.01* | 1.0*** | 0.54*** | | | |
| EY | -0.47*** | -0.9*** | -0.88*** | -0.89*** | | |
| RT | 0.8*** | -0.56*** | 0.36*** | -0.56*** | 0.14** | |

*** P<1x 10 $^{-10}$, **P< .0002, * Not significant.

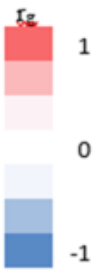



**Figure 3.14: Heat map of genetic correlations**
Genetic correlations between the two GBS GWAS of Resilience and EduYears, the two inputs to GBS (EY+Res and EY/NonRes) and GWAS of the two variables used to create these phenotypes (EY and RT).

## 3.3 Conclusion

In this chapter I outlined the limitations of the available cognitive data in the UKB to explore cognitive resilience. I showed how a resilience variable could be generated using a proxy measure for past cogntive performance but recognised that this variable was strongly influenced by EY which is highly heritability with a polygenic nature that can mask the genetics of resilience.

To overcome this I employed Genomics Structural Equation Modelling (GenomicSEM) (Grotzinger et al., 2019) to perform a GWAS-by-subtraction (Demange et al., 2021) using two GWAS, one which captured genetic variants associated with EY and resilience and a second which captured genetic variants associated with EY but not resilience. Subtracting one from the other generated two new GWAS, one capturing EY and the other capturing the genetics of a processing speed-based cognitive resilience phenotype. Replication of this approach was shown using independent discovery and replication samples. The full GWAS results were now available for functional analysis, which is explored in chapter 4.

# 4 Functional analysis of Resilience

## 4.1 Introduction

In chapter 3, I described how I researched methods to extract the genetic contribution of SNPs associated with *Resilience* from a GWAS that also contained a strong genetic contribution for EY. We showed that this method could be replicated in an independent sample and combining our discovery and replication sample we generated a GWAS of *Resilience* on 330,097 individuals in the UKB.

In this chapter I perform functional analysis using an array of bioinformatic tools described in chapter 2 to connect the genetic associations in *Resilience* to genes, biological processes, tissues and cell types to gain insight into the biological nature of cognitive resilience.

## 4.2 Description of genetic loci

Function analysis was performed on *Resilience* in FUMA v 1.3.6 (Watanabe et al., 2017). (Note: see Table 2.7: Parameters used in FUMA for parameters used and results publicly available in FUMA ID:171).

A total of 1,329 significant SNPs were tagged from the *Resilience* GWAS and were associated with 26 independent lead SNPs ($P < 5 \times 10^{-8}$). By including SNPs in the reference panel that are in LD with the independent SNPs, a total of 1,922 candidate SNPs were identified. Functional annotation of the candidate SNPs showed that 82% were intergenic/intronic (Figure 4.1).

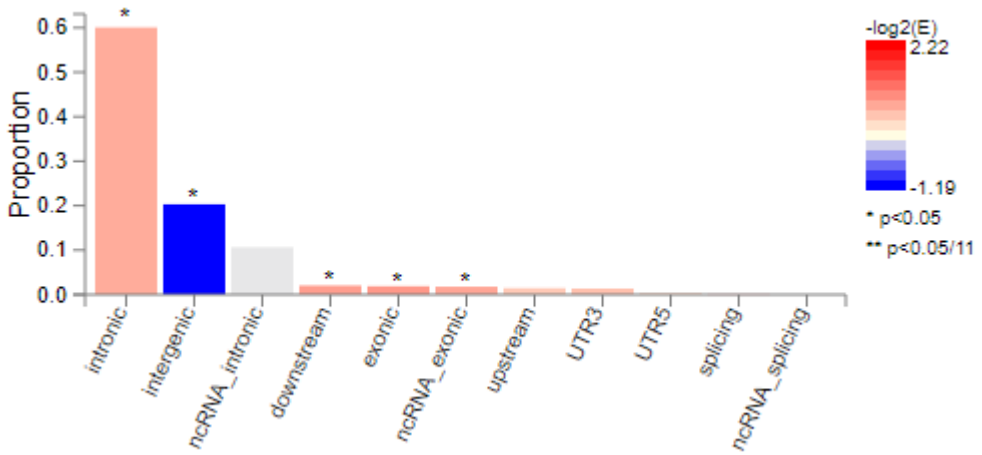



**Figure 4.1: Functional consequences of SNPs on genes**
This histogram displays the proportion of SNPs (all SNPs in LD of Ind. sig. SNPs) which have corresponding functional annotation assigned in FUMA. Bars are coloured by log2(enrichment) relative to all SNPs in the reference panel.

A total of 84 SNPs had a Combined Annotation Dependent Depletion (CADD) score greater than the threshold of 12.37 which indicates that the variants are potentially pathogenic (Kircher et al., 2014) . A full list of these SNPs can be found in Table 4 of the Supplementary material attached to my published research (Fitzgerald et al., 2021).

Lead SNPs were grouped into 13 independent genetic loci that are on 9 different chromosomes (see  Figure 4.2 and Table 4.1). Regional plots for all 13 loci are shown in Figure 4.3 to Figure 4.15.

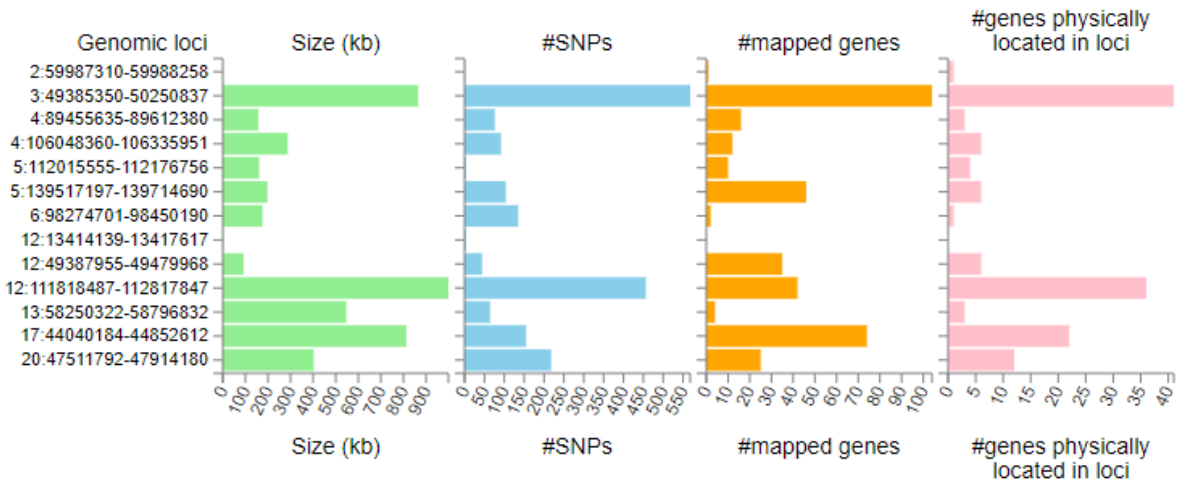



**Figure 4.2: Summary per genomic risk locus**
These histogram display summary results for the genetic risk loci by size, number of SNPS, mapped genes and number of SNPs physically located in the loci.

**Table 4.1: Thirteen genomic risk loci and their independent significant SNPs**

| **Genomic loci** | | | | | | **Independent significant SNPs** | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Locus** | SNP | locus start | locus end | [a]nSNPs | [b]nGWAS SNPs | SNP | Position | P value | [a]nSNPs | [b]nGWAS SNPs |
| **2** | rs12474507 | 59987310 | 59988258 | 2 | 1 | rs12474507 | 59988258 | 4.19E-08 | 2 | 1 |
| **3** | rs2352974 | 49385350 | 50250837 | 568 | 402 | rs2352974 | 49890613 | 9.27E-15 | 117 | 80 |
| | | | | | | rs35999162 | 49597230 | 4.75E-09 | 142 | 92 |
| | | | | | | rs1317140 | 49878652 | 1.84E-10 | 56 | 41 |
| | | | | | | rs2883059 | 49902160 | 2.39E-10 | 205 | 160 |
| | | | | | | rs7428430 | 50174184 | 2.49E-14 | 205 | 146 |
| | | | | | | rs2526389 | 50192826 | 9.35E-11 | 30 | 28 |
| | | | | | | rs1046953 | 50197097 | 1.07E-10 | 5 | 4 |
| | | | | | | rs9858059 | 50227871 | 2.88E-08 | 82 | 67 |
| **4A** | rs6857847 | 89455635 | 89612380 | 76 | 59 | rs6857847 | 89514572 | 4.99E-09 | 62 | 52 |
| | | | | | | rs57672162 | 89459723 | 3.71E-08 | 37 | 28 |
| **4B** | rs2189234 | 106048360 | 106335951 | 92 | 65 | rs2189234 | 106075498 | 3.14E-12 | 44 | 29 |
| | | | | | | rs2726485 | 106263450 | 8.27E-09 | 80 | 56 |
| **5A** | rs56335290 | 112015555 | 112176756 | 4 | 3 | rs56335290 | 112036634 | 2.59E-08 | 4 | 3 |
| **5B** | rs6870103 | 139517197 | 139714690 | 104 | 71 | rs6870103 | 139692515 | 4.03E-11 | 104 | 71 |
| **6** | rs7747481 | 98274701 | 98450190 | 135 | 103 | rs7747481 | 98315696 | 2.29E-08 | 135 | 103 |
| **12A** | rs2417261 | 13414139 | 13417617 | 4 | 3 | rs2417261 | 13414139 | 1.35E-08 | 4 | 3 |
| **12B** | rs6580699 | 49387955 | 49479968 | 44 | 25 | rs6580699 | 49478812 | 1.29E-10 | 44 | 25 |
| **12C** | rs1029388 | 111818487 | 112817847 | 456 | 356 | rs1029388 | 111926901 | 6.50E-09 | 456 | 356 |
| **13** | rs9569811 | 58250322 | 58796832 | 64 | 21 | rs9569811 | 58646190 | 1.14E-08 | 64 | 21 |
| **17** | rs62074125 | 44040184 | 44852612 | 155 | 61 | rs62074125 | 44852612 | 8.31E-11 | 10 | 2 |
| | | | | | | rs10775404 | 44167366 | 1.35E-09 | 66 | 30 |
| | | | | | | rs7225002 | 44189067 | 2.44E-10 | 19 | 1 |
| | | | | | | 17:44224272_G_A | 44224272 | 4.52E-09 | 63 | 28 |
| **20** | rs4810896 | 47511792 | 47914180 | 218 | 159 | rs4810896 | 47535298 | 1.94E-10 | 198 | 145 |
| | | | | | | rs2426132 | 47723127 | 1.54E-09 | 21 | 15 |

Notes: The 13 independent significant SNPs (in the right panel) have a P value > 0.5 x 10-8, with an LD limit of r2 =0.6, Genomic risk loci (left panel) have a further r2 limit of 0.1 and a maximum distance of 250 kb. Loci number reflects the chromosome number. Where there is more than one locus per chromosome the locus is given an A, B, C

in order of position. [a]nSNPs = number of candidate SNPs in LD with the index or lead SNP, [b]nGWASSNPs = number of SNPs in the GWAS in LD with the index or lead SNP.

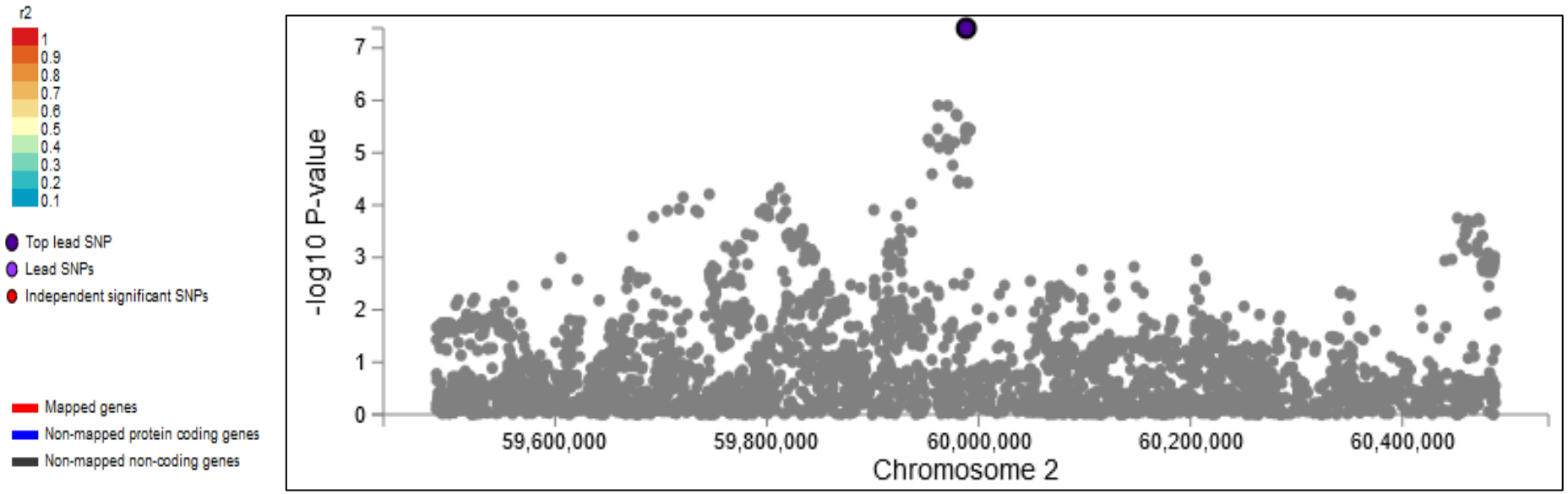



**Figure 4.3: Locus 2 - Regional plot - 2:59987310-59988258**
Regional plots were generated in FUMA (https://fuma.ctglab.nl). The x-axis is chromosomal position, the y-axis is significance of association with *Resilience* represented as -log10 (P). The top lead or Index SNP is **rs1247450**. There are no other SNPs in LD with this SNP. There are no mapped genes in this locus.

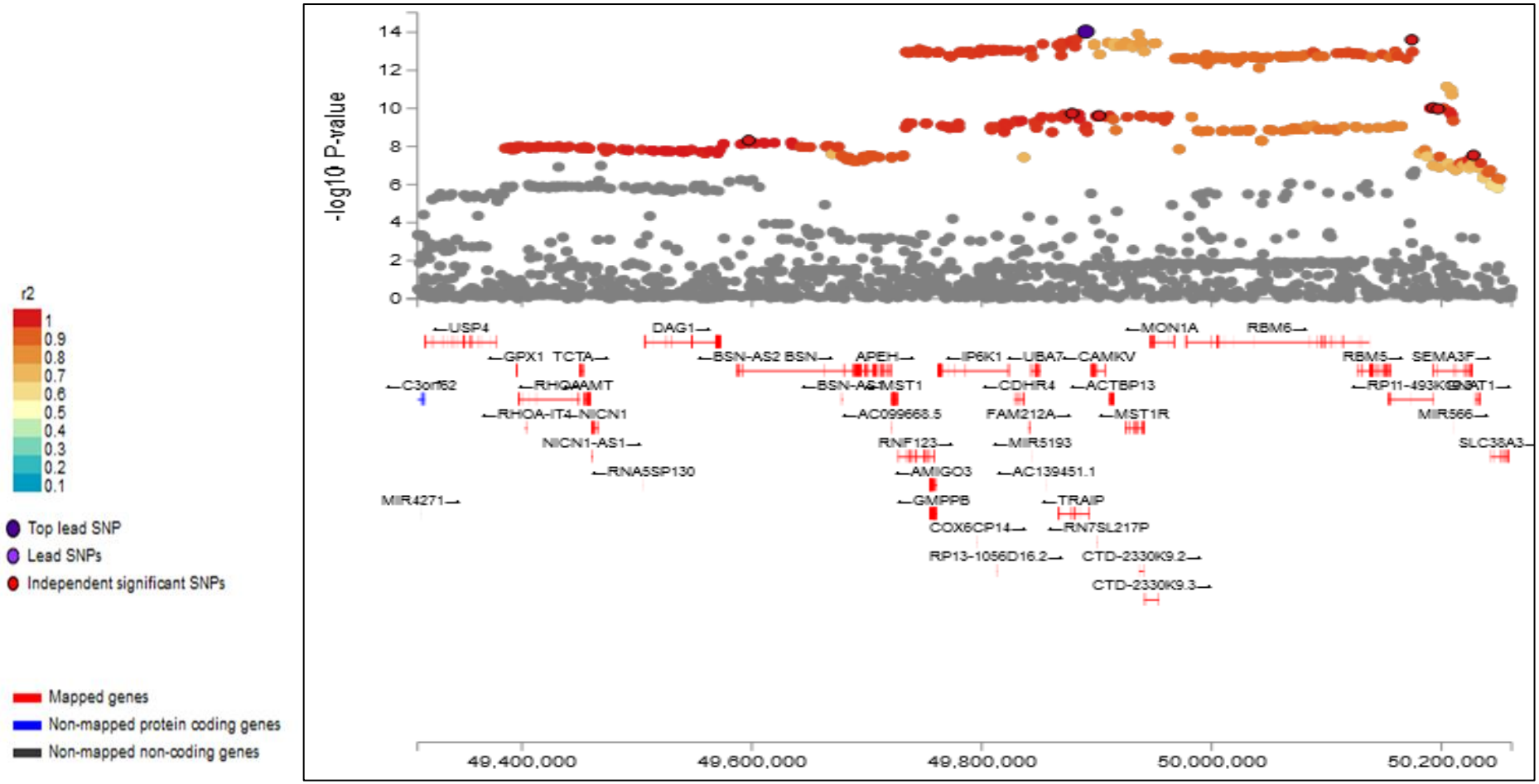



**Figure 4.4: Locus 3 Regional plot - 3:49385350-50250837.**
Regional plots were generated in FUMA (https://fuma.ctglab.nl). The x-axis is chromosomal position, the y-axis is significance of association with *Resilience* represented as -log10 (P). The top lead or Index SNP is rs2352974. There are 402 GWAS SNPs in LD with this SNP including 7 other independent significant SNPs (Table 4-4).

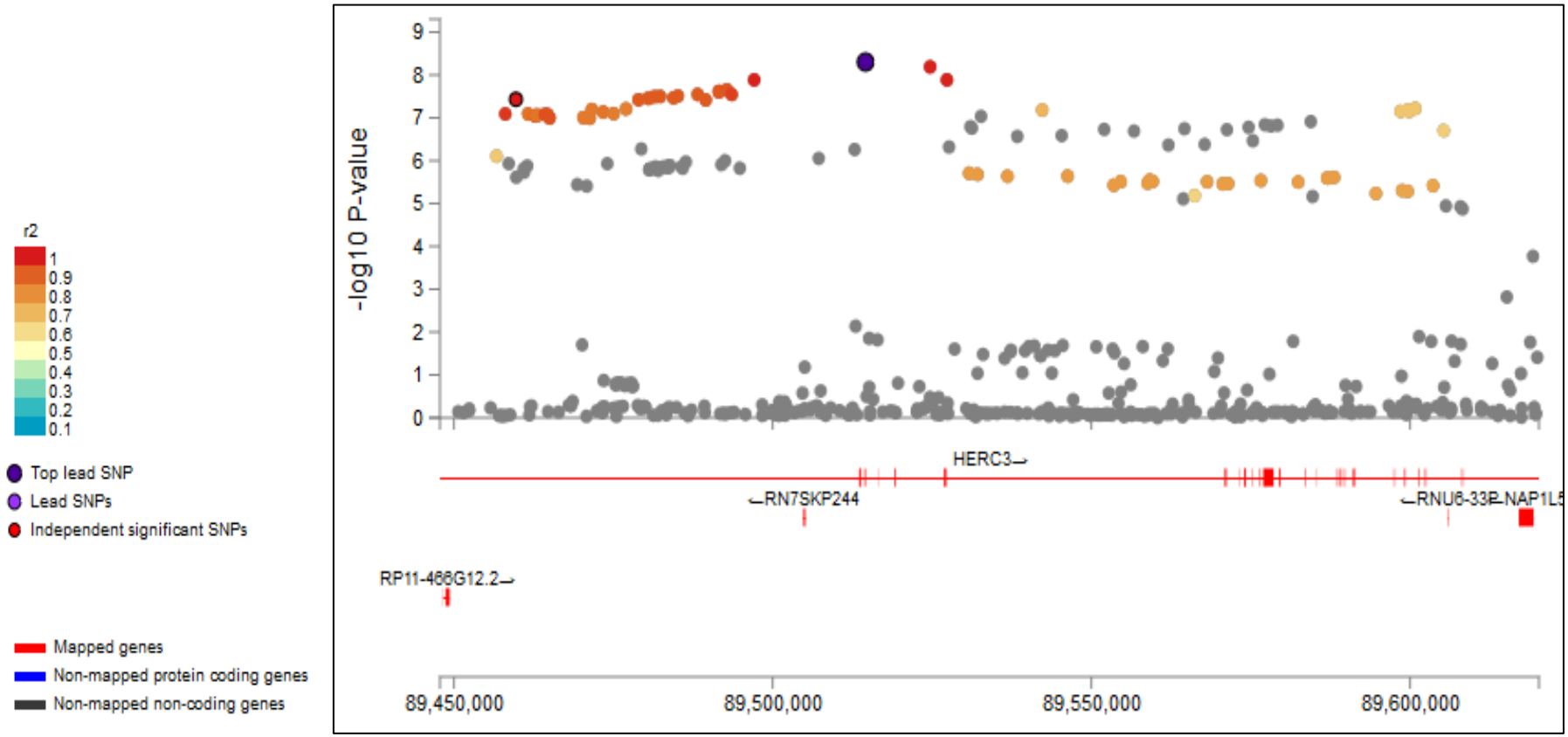



**Figure 4.5: Locus 4A Regional plot - 4:89455635-89612380.**
Regional plots were generated in FUMA (https://fuma.ctglab.nl). The x-axis is chromosomal position, the y-axis is significance of association with *Resilience* represented as -log10 (P). The top lead or Index SNP is rs6857874. There are 59 GWAS SNPs in LD with this SNP including 1 independent significant SNP rs57672162 at 4:89459723.

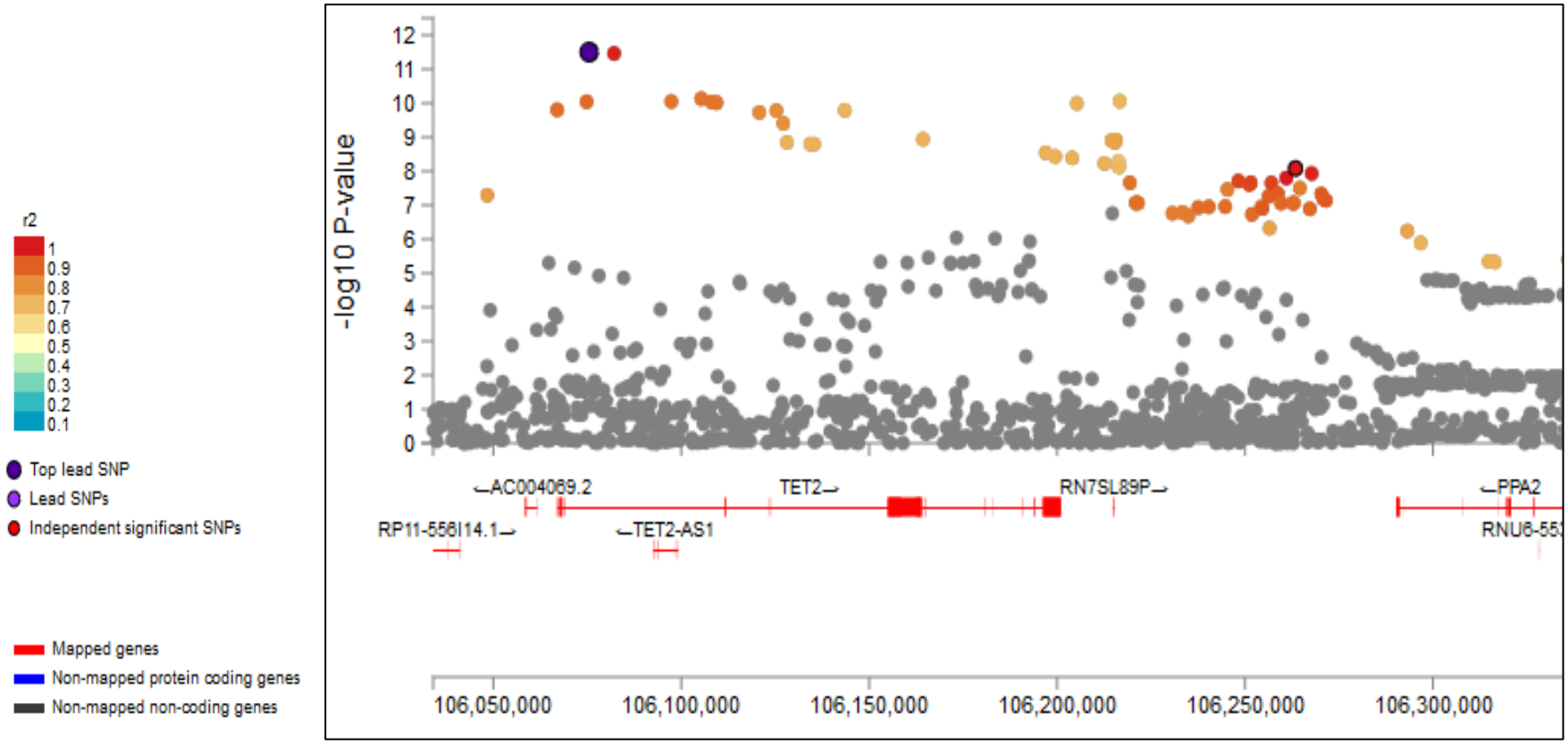



**Figure 4.6: Locus 4B Regional plot - 4:106048360-106335951.**
Regional plots were generated in FUMA (https://fuma.ctglab.nl). The x-axis is chromosomal position, the y-axis is significance of association with *Resilience* represented as -log10 (P). The top lead or Index SNP is rs2189234. There are 65 GWAS SNPs in LD with this SNP including 1 independent significant SNP rs2726485 at 4:106263450.

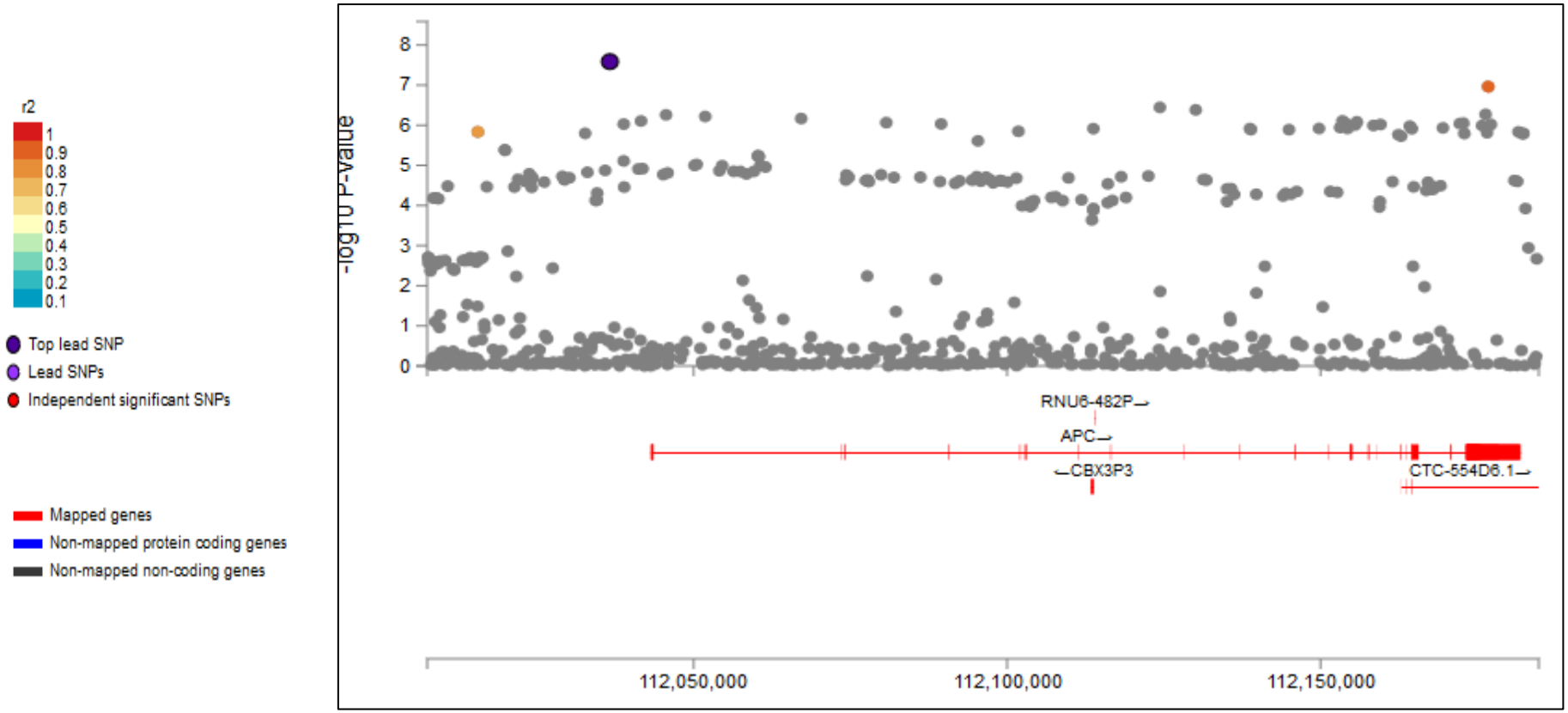



**Figure 4.7: Locus 5A Regional plot 5: 112015555-112176756.**
Regional plots were generated in FUMA (https://fuma.ctglab.nl). The x-axis is chromosomal position, the y-axis is significance of association with *Resilience* represented as -log10 (P). The top lead or Index SNP is rs56335290. There are 3 GWAS SNPs in LD with this SNP.

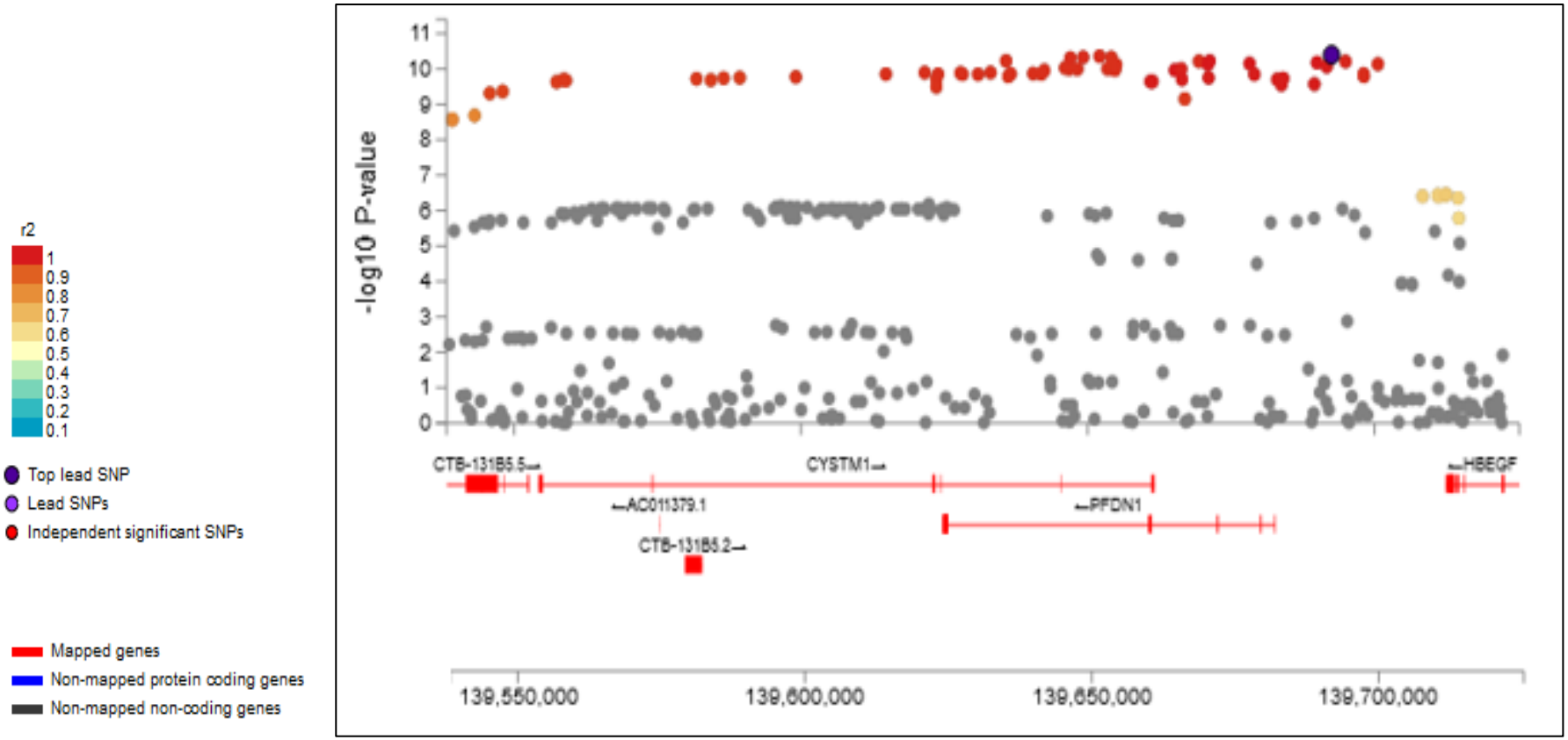



**Figure 4.8: Locus 5B Regional plot 5: 139517197-139714690.**
Regional plots were generated in FUMA (https://fuma.ctglab.nl). The x-axis is chromosomal position, the y-axis is significance of association with *Resilience* represented as -log10 (P). The top lead or Index SNP is rs6870103. There are 71 GWAS SNPs in LD with this SNP.

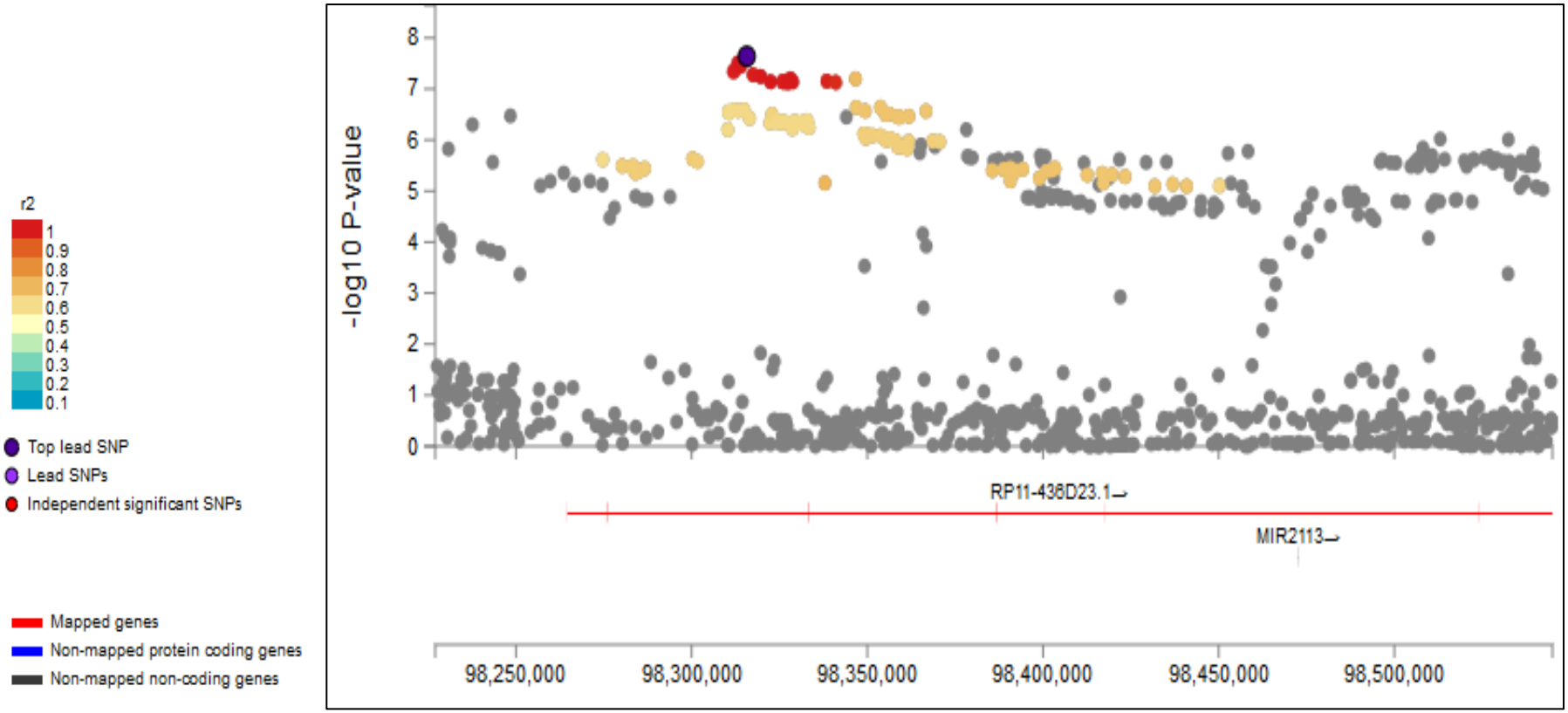



**Figure 4.9: Locus Regional plot – 6: 98274701-98450190.**
Regional plots were generated in FUMA (https://fuma.ctglab.nl). The x-axis is chromosomal position, the y-axis is significance of association with *Resilience* represented as -log10 (P). The top lead or Index SNP is rs7747481. There are 103 GWAS SNPs in LD with this SNP.

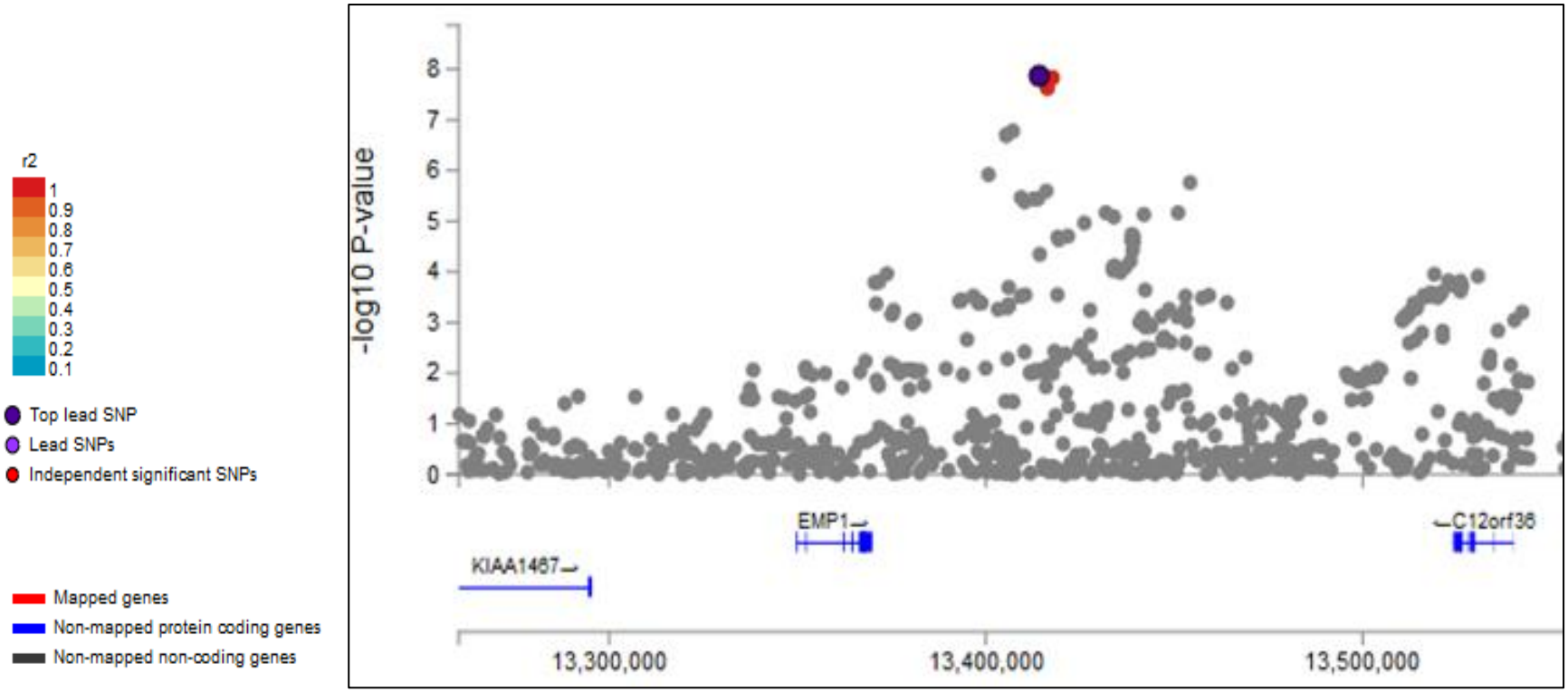



**Figure 4.10: Locus 12A Regional plot – 12: 13414139-13417617.**
Regional plots were generated in FUMA (https://fuma.ctglab.nl). The x-axis is chromosomal position, the y-axis is significance of association with *Resilience* represented as -log10 (P). The top lead or Index SNP is rs2417261. There are 3 GWAS SNPs in LD with this SNP.

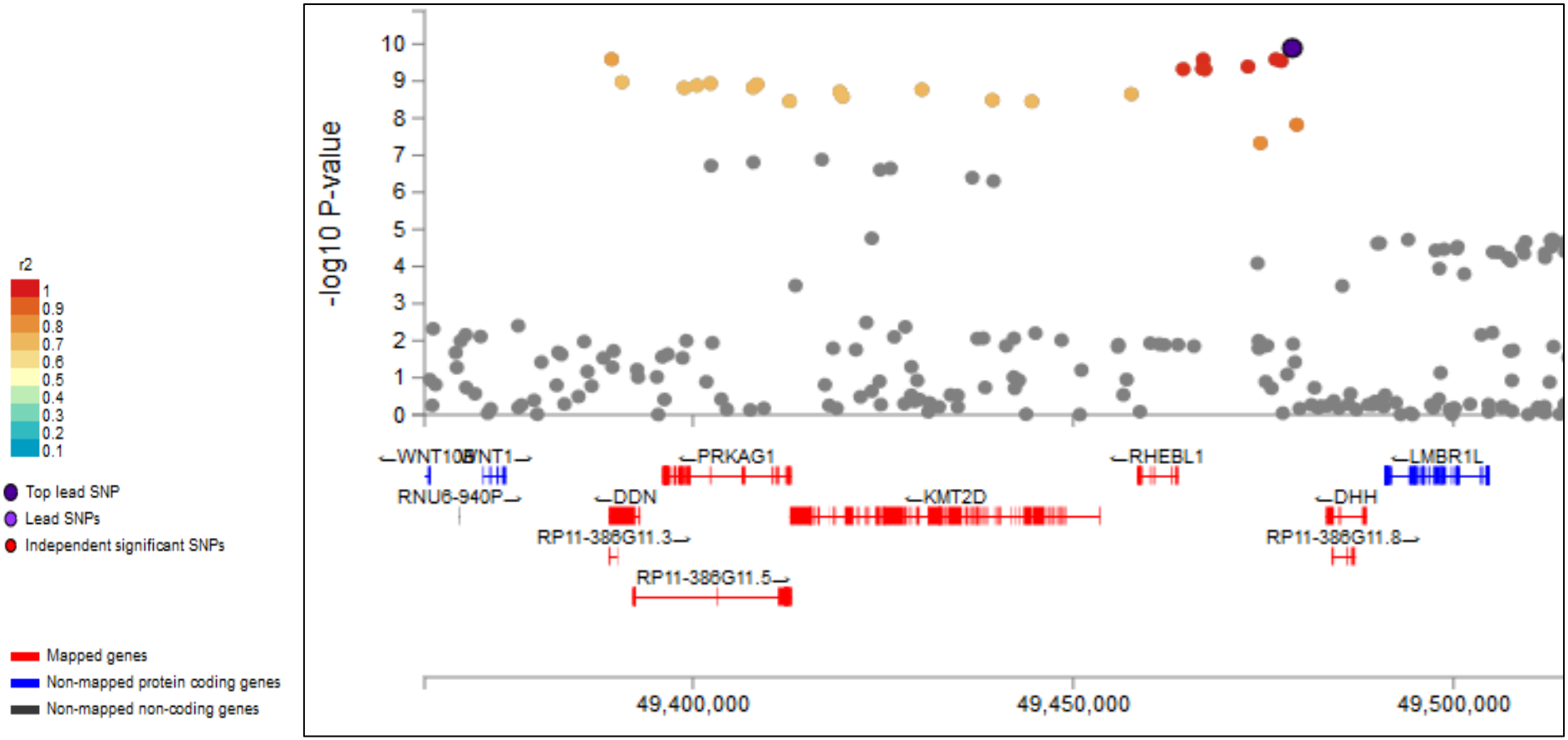



**Figure 4.11: Locus 12B. Regional plot- 12: 49387955-49479968.**
Regional plots were generated in FUMA (https://fuma.ctglab.nl). The x-axis is chromosomal position, the y-axis is significance of association with *Resilience* represented as -log10 (P). The top lead or Index SNP is rs6580699. There are 25 GWAS SNPs in LD with this SNP.

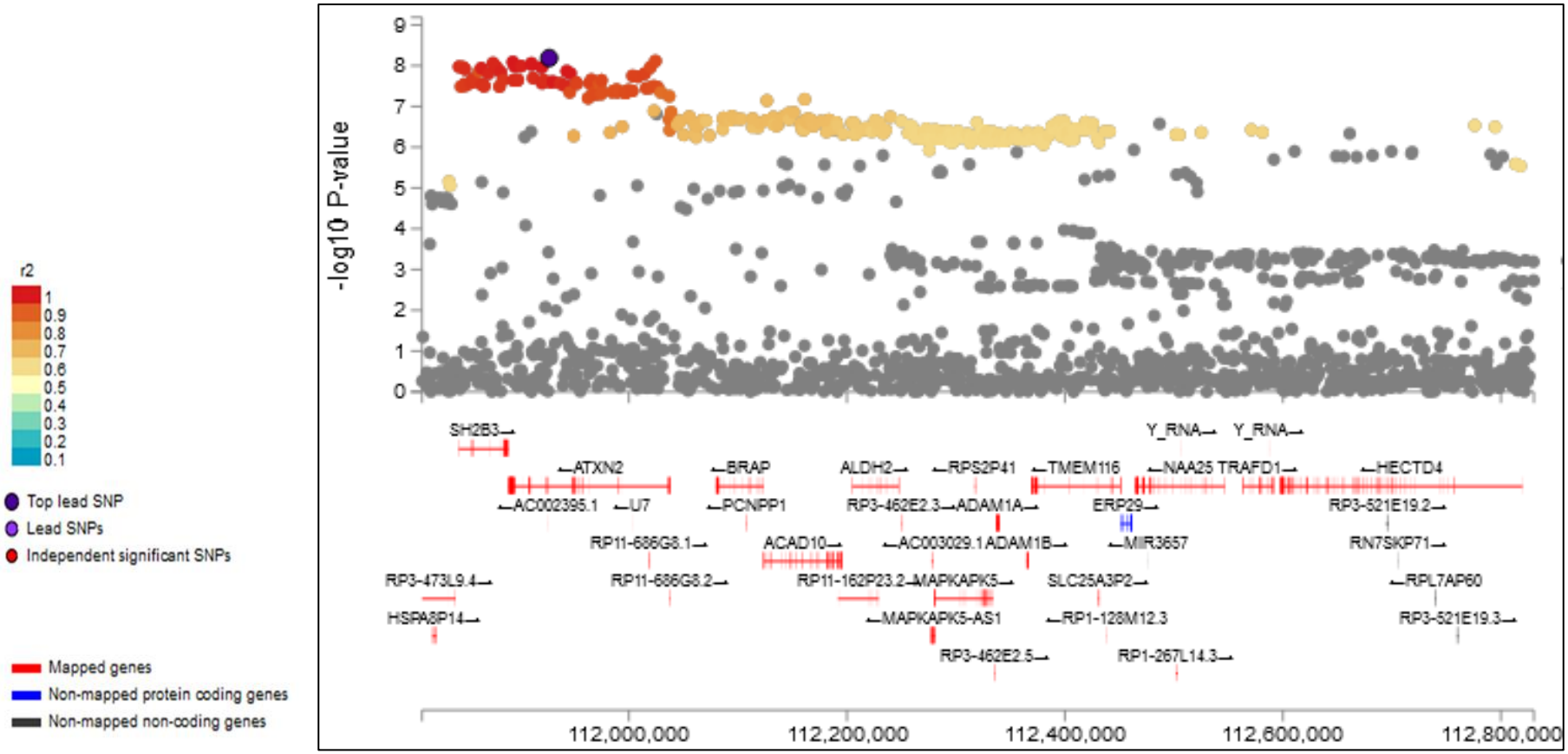



**Figure 4.12: Locus 12C Regional plot – 12: 111818487-112817847.**
Regional plots were generated in FUMA (https://fuma.ctglab.nl). The x-axis is chromosomal position, the y-axis is significance of association with *Resilience* represented as -log10 (P). The top lead or Index SNP is rs1029388. There are 356 GWAS SNPs in LD with this SNP.

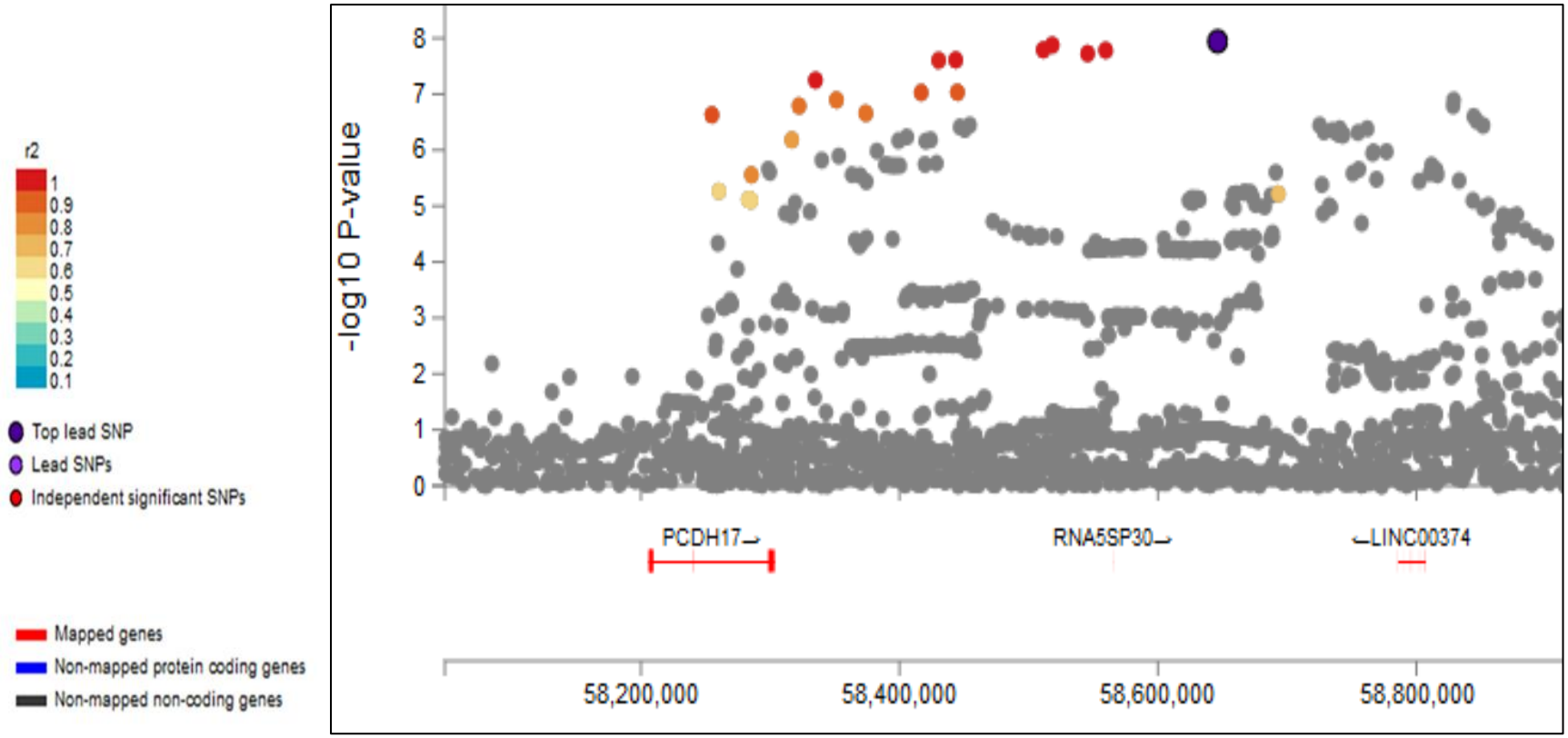



**Figure 4.13: Locus 13 Regional plot – 13:58250322 – 58796832.**
Regional plots were generated in FUMA (https://fuma.ctglab.nl). The x-axis is chromosomal position, the y-axis is significance of association with Resilience represented as -log10 (P). The top lead or Index SNP is rs9569811. There are 21 GWAS SNPs in LD with this SNP.

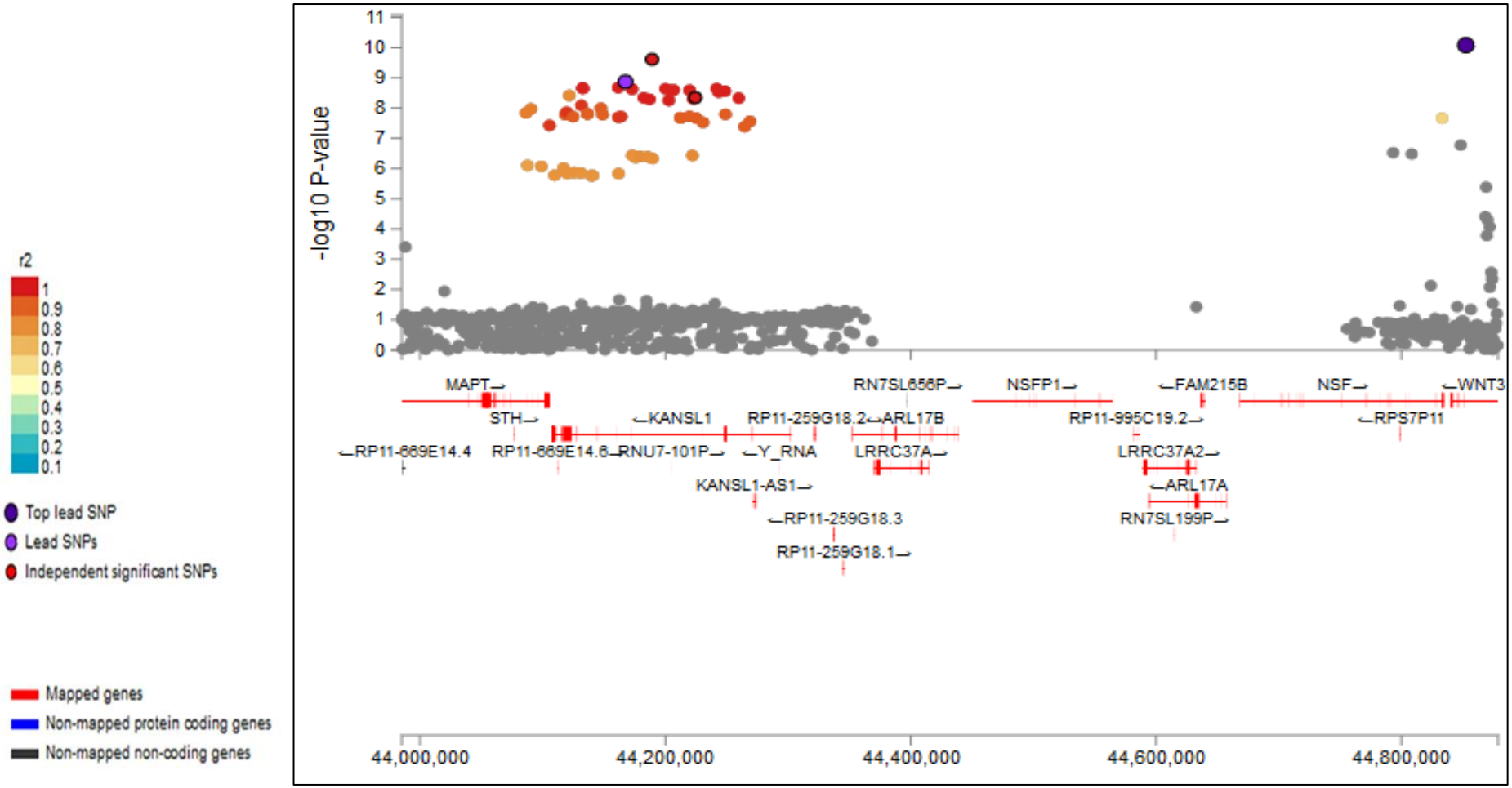



**Figure 4.14: Locus 17 Regional plot- 17: 44040184-44852612.**
Regional plots were generated in FUMA (https://fuma.ctglab.nl). The x-axis is chromosomal position, the y-axis is significance of association with *Resilience* represented as -log10 (P). The top lead or Index SNP is rs62074125. There are 61 GWAS SNPs in LD with this SNP including 3 other independent lead SNPs (Table 4-4).

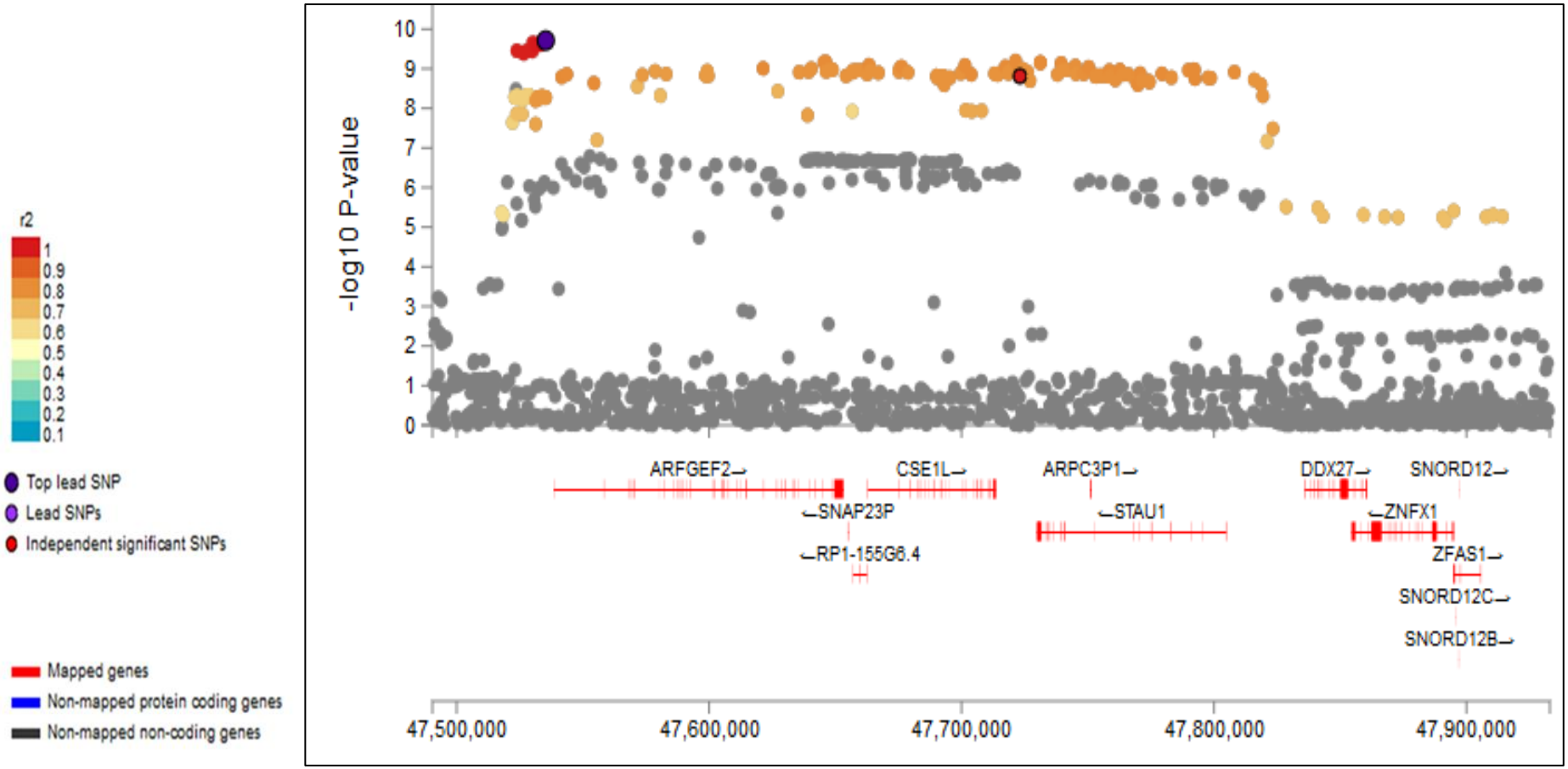



**Figure 4.15: Locus 20 Regional plot – 20: 47511792-47914180.**
Regional plots were generated in FUMA (https://fuma.ctglab.nl). The x-axis is chromosomal position, the y-axis is significance of association with *Resilience* represented as -log10 (P). The top lead or Index SNP is rs4810896. There are 159 GWAS SNPs in LD with this SNP including 1 other independent lead SNP rs2426132 20:47723127.

### 4.2.1 Conditional Analysis

Conditional analysis was performed at each locus where there was more than one independent significant SNP within 1000 kb of the index SNP. These analyses showed that the significance of all independent lead SNPs at each locus was reduced when the GWAS was conditioned for the index SNP, confirming the linkage of the index SNP to each lead SNP (Table 4.2).

### 4.2.2 Fine Mapping

FINEMAP (Benner et al., 2016) was used to provide further information on significant SNPs in LD with the index SNP on each locus using the GWAS SNPs generated by FUMA . The $\log_{10}$ Bayes factor ($B_{10}$) quantifies causal evidence for a particular SNP and a posterior probability value yielding a $B_{10}$ greater than 2 indicates considerable evidence of causality (Benner et al., 2016). One SNP, rs62074125, on chromosome 17, exceeded this value ($B_{10}$ = 2.64). This SNP is an intron within the *WNT3* gene, which is associated with cognitive function (Davies et al., 2018). Further examination in various genomic browsers does not identify a specific function for this SNP, however eQTL mapping shows expression in various brain regions associated with *WNT3* and *KANSL1*. It has no reported clinical significance. The next highest result was on chromosome 4 where rs2189234 had a value slightly below 2 ($B_{10}$ = 1.62). This SNP is an intronic variant in the *TET2* gene, which is discussed below. FINEMAP analysis showed that the index SNP had the highest Bayes Factor for all loci with four exceptions. A summary of the FINEMAP analysis is shown in Table 4.3. Detail of the analysis by SNP are in Supplementary Table 7 (Fitzgerald et al., 2021).

**Table 4.2: Conditional analysis of each locus conditioning for the index SNP in EY+Res**

| | | | | | **Resilience** | | | **EY + Res** | | | **Conditioned on index SNP** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Locus** | rsID | position | start | end | [a]Beta | SE | P value | [a]OR | SE | P value | [a]OR | SE | P value |
| **3** | **rs2352974** | **49890613** | **49385350** | **50250837** | **0.080** | **0.010** | **9.27E-15** | **1.101** | **0.007** | **1.09E-38** | NA | NA | NA |
| | rs35999162 | 49597230 | | | -0.065 | 0.011 | 4.75E-09 | 0.914 | 0.008 | 2.37E-29 | 0.959 | 0.011 | 7.63E-05 |
| | rs1317140 | 49878652 | | | 0.069 | 0.011 | 1.84E-10 | 1.077 | 0.008 | 5.50E-21 | 1.005 | 0.011 | 6.56E-01 |
| | rs2883059 | 49902160 | | | 0.065 | 0.010 | 2.39E-10 | 1.090 | 0.007 | 4.35E-31 | 1.040 | 0.010 | 1.56E-04 |
| | rs7428430 | 50174184 | | | -0.078 | 0.010 | 2.49E-14 | 0.911 | 0.007 | 2.23E-36 | 0.946 | 0.011 | 3.94E-07 |
| | rs2526389 | 50192826 | | | 0.067 | 0.010 | 9.35E-11 | 1.082 | 0.007 | 5.81E-26 | 1.035 | 0.009 | 2.45E-04 |
| | rs1046953 | 50197097 | | | -0.067 | 0.010 | 1.07E-10 | 0.924 | 0.008 | 3.40E-26 | 0.970 | 0.010 | 2.07E-03 |
| | rs9858059 | 50227871 | | | 0.057 | 0.010 | 2.88E-08 | 1.075 | 0.007 | 2.99E-22 | 1.027 | 0.009 | 3.51E-03 |
| **4A** | **rs6857847** | **89514572** | **89455635** | **89612380** | **-0.076** | **0.013** | **4.99E-09** | **0.953** | **0.009** | **2.25E-07** | NA | NA | NA |
| | rs57672162 | 89459723 | | | -0.074 | 0.014 | 3.71E-08 | 0.952 | 0.010 | 4.76E-07 | 0.974 | 0.018 | 1.54E-01 |
| **4B** | **rs2189234** | **106075498** | **106048360** | **106335951** | **0.073** | **0.010** | **3.14E-12** | **1.064** | **0.008** | **2.22E-16** | NA | NA | NA |
| | rs2726485 | 106263450 | | | 0.059 | 0.010 | 8.27E-09 | 1.058 | 0.007 | 3.98E-14 | 1.023 | 0.012 | 4.60E-02 |
| **17** | **rs62074125** | **44852612** | **44040184** | **44852612** | **-0.077** | **0.012** | **8.31E-11** | **0.957** | **0.008** | **2.10E-07** | NA | NA | NA |
| | rs10775404 | 44167366 | | | 0.080 | 0.013 | 1.35E-09 | 1.038 | 0.009 | 1.01E-04 | 1.026 | 0.010 | 8.08E-03 |
| | rs7225002 | 44189067 | | | 0.065 | 0.010 | 2.44E-10 | 1.050 | 0.007 | 7.83E-11 | 1.040 | 0.008 | 3.67E-06 |
| | 17:44224272_G_A | 44224272 | | | -0.060 | 0.010 | 4.52E-09 | 0.961 | 0.007 | 7.77E-08 | 0.973 | 0.009 | 3.05E-03 |
| **20** | **rs4810896** | **47535298** | **47511792** | **47914180** | **0.067** | **0.011** | **1.94E-10** | **1.060** | **0.008** | **3.22E-14** | NA | NA | NA |
| | rs2426132 | 47723127 | | | -0.062 | 0.010 | 1.54E-09 | 0.948 | 0.007 | 4.42E-13 | 0.972 | 0.010 | 4.95E-03 |

Notes: The effect of conditioning of the index SNPs (highlighted in bold) on the independent significant SNPs was examined in EY+Res and not in Resilience due to the effects of the GBS manipulation. The effect of adding the index SNP as a covariate to the GWAS analysis diminished the effect of all the independent significant SNPs. aBeta/ OR = effect size of the A1 (minor) allele.

**Table 4.3: Summary of Fine Mapping**

| **Index SNPs** | | | | | | **Independent Significant SNPs** | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **SNP** | **Chr** | **position** | **P value** | **start** | **end** | **SNP** | **position** | **P Value** | **maf** | **[a]beta** | **se** | **[b]prob** | **[c]log10bf** |
| **rs12474507** | 2 | 59988258 | 4.19E-08 | 59987310 | 59988258 | rs12474507 | | 4.19E-08 | 0.401 | -0.057 | 0.010 | | |
| | 3 | | | | | rs35999162 | | 4.75E-09 | 0.307 | -0.065 | 0.011 | 0.001 | -0.449 |
| | 3 | | | | | rs1317140 | | 1.84E-10 | 0.320 | 0.069 | 0.011 | 0.002 | -0.330 |
| **rs2352974** | 3 | 49890613 | 9.27E-15 | 49385350 | 50250837 | **rs2352974** | | **9.27E-15** | **0.492** | **0.080** | **0.010** | **0.072** | **1.318** |
| | 3 | | | | | rs2883059 | | 2.39E-10 | 0.428 | 0.065 | 0.010 | 0.000 | -1.033 |
| | 3 | | | | | rs7428430 | | 2.49E-14 | 0.487 | -0.078 | 0.010 | 0.033 | 0.965 |
| | 3 | | | | | rs2526389 | | 9.35E-11 | 0.425 | 0.067 | 0.010 | 0.002 | -0.186 |
| | 3 | | | | | rs1046953 | | 1.07E-10 | 0.415 | -0.067 | 0.010 | 0.001 | -0.644 |
| | 3 | | | | | rs9858059 | | 2.88E-08 | 0.448 | 0.057 | 0.010 | 0.000 | -1.102 |
| | 4A | | | | | rs57672162 | | 3.71E-08 | 0.171 | -0.074 | 0.014 | 0.008 | -0.492 |
| **rs6857847** | 4A | 89514572 | 4.99E-09 | 89455635 | 89612380 | **rs6857847** | | **4.99E-09** | **0.193** | **-0.076** | **0.013** | **0.165** | **0.879** |
| **rs2189234** | 4B | 106075498 | 3.14E-12 | 106048360 | 106335951 | **rs2189234** | | **3.14E-12** | **0.383** | **0.073** | **0.010** | **0.496** | **1.621** |
| | 4B | | | | | rs2726485 | | 8.27E-09 | 0.423 | 0.059 | 0.010 | 0.002 | -1.158 |
| **rs56335290** | 5A | 112036634 | 2.59E-08 | 112015555 | 112176756 | **rs56335290** | | **2.59E-08** | **0.220** | **-0.068** | **0.012** | **0.838** | **0.758** |
| **rs6870103** | 5B | 139692515 | 4.03E-11 | 139517197 | 139714690 | **rs6870103** | **139692515** | **4.03E-11** | **0.450** | **0.068** | **0.010** | **0.059** | **0.464** |
| | 5B | | | | | *rs13186198* | *139652057* | *4.40E-11* | *0.44* | *0.068* | *0.01* | *0.059* | *0.465* |
| **rs7747481** | 6 | 98315696 | 2.29E-08 | 98274701 | 98450190 | **rs7747481** | | **2.29E-08** | **0.405** | **0.058** | **0.010** | **0.121** | **0.968** |
| **rs2417261** | 12A | 13414139 | 1.35E-08 | 13414139 | 13417617 | **rs2417261** | **13414139** | **1.35E-08** | **0.105** | **0.094** | **0.017** | **0.106** | **-0.879** |
| | 12A | | | | | *rs10772644* | *13417617* | *1.47E-08* | *0.11* | *0.093* | *0.016* | *0.526* | *0.092* |
| **rs6580699** | 12B | 49478812 | 1.29E-10 | 49387955 | 49479968 | **rs6580699** | | **1.29E-10** | **0.434** | **-0.066** | **0.010** | **0.160** | **0.478** |
| **rs1029388** | 12C | 111926901 | 6.50E-09 | 111818487 | 112817847 | **rs1029388** | 111926901 | **6.50E-09** | **0.211** | **0.073** | **0.013** | **0.016** | **0.576** |
| | 12C | | | | | *rs11065898* | *111862575* | *1.39E-08* | *0.21* | *0.07* | *0.012* | *0.054* | *1.132* |
| | 12C | | | | | *rs7953810* | *112029291* | *4.51E-08* | *0.21* | *0.068* | *0.012* | *0.021* | *0.710* |
| **rs9569811** | 13 | 58646190 | 1.14E-08 | 58250322 | 58796832 | **rs9569811** | | **1.14E-08** | **0.130** | **-0.086** | **0.015** | **0.160** | **0.397** |
| | 17 | | | | | rs10775404 | | 1.35E-09 | 0.185 | 0.080 | 0.013 | 0.087 | 0.578 |

| **Index SNPs** | | | | | | **Independent Significant SNPs** | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **SNP** | **Chr** | **position** | **P value** | **start** | **end** | **SNP** | **position** | **P Value** | **maf** | **[a]beta** | **se** | **[b]prob** | **[c]log10bf** |
| | 17 | | | | | rs7225002 | | 2.44E-10 | 0.416 | 0.065 | 0.010 | 0.019 | -0.114 |
| | 17 | | | | | 17:44224272_G_A | | 4.52E-09 | 0.452 | -0.060 | 0.010 | 0.007 | -0.556 |
| **rs62074125** | 17 | 44852612 | 8.31E-11 | 44040184 | 44852612 | **rs62074125** | | **8.31E-11** | **0.253** | **-0.077** | **0.012** | **0.916** | **2.637** |
| **rs4810896** | 20 | 47535298 | 1.94E-10 | 47511792 | 47914180 | **rs4810896** | **47535298** | **1.94E-10** | **0.366** | **0.067** | **0.011** | **0.004** | **-0.416** |
| | 20 | | | | | rs2426132 | 47723127 | 1.54E-09 | 0.459 | -0.062 | 0.010 | 0.011 | 0.075 |
| | 20 | | | | | *rs1043361* | *47731158* | *7.09E-10* | *0.42* | *0.064* | *0.01* | *0.021* | *0.358* |

Notes: Index SNPs are compared with independent significant SNPs for causality. SNPs other than independent SNPs having a higher bayes factor than the index SNPs are added in italics and their comparative positions are shown. [a]beta = effect of A1 (minor allele), [b]prob = the posterior probabilities that configurations are the causal configuration, clog10bf= the log10 Bayes factor. The Bayes factor quantifies the evidence for a causal configuration over the null configuration (no SNPs are causal).

### 4.2.3 Gene mapping

Three approaches were used in FUMA to map the associated variants to genes: (a) Positional mapping SNPs to 141 genes based on their genomic location within a 10 kilobase window of known gene boundaries. (b) Expression quantitative trait (eQTL) analysis mapped 207 cis-eQTL SNPs to genes whose expression they were associated with. (c) Chromatin interaction analysis using the 3D DNA to DNA interactions mapped SNPs to 243 genes. Circos plots for all loci are attached in Figure 4.16 to Figure 4.26.

The circos plot from chromosome 3 shows that 102 genes were mapped to this region, representing 42% of the total genes mapped (I further explore the influence of this locus on the functional analysis in chapter 4). In addition, the circos plot from chromosome 17 shows two distinct clusters of SNPs. Genes in this region (MAPT, WNT3, CRHR1, KANSL1, and NSF) have been previously associated with general cognitive function but also with other cognitive indicators (Davies et al., 2018). Details of this gene mapping analysis is in Supplementary Table 9 (Fitzgerald et al., 2021).

(a) (b)

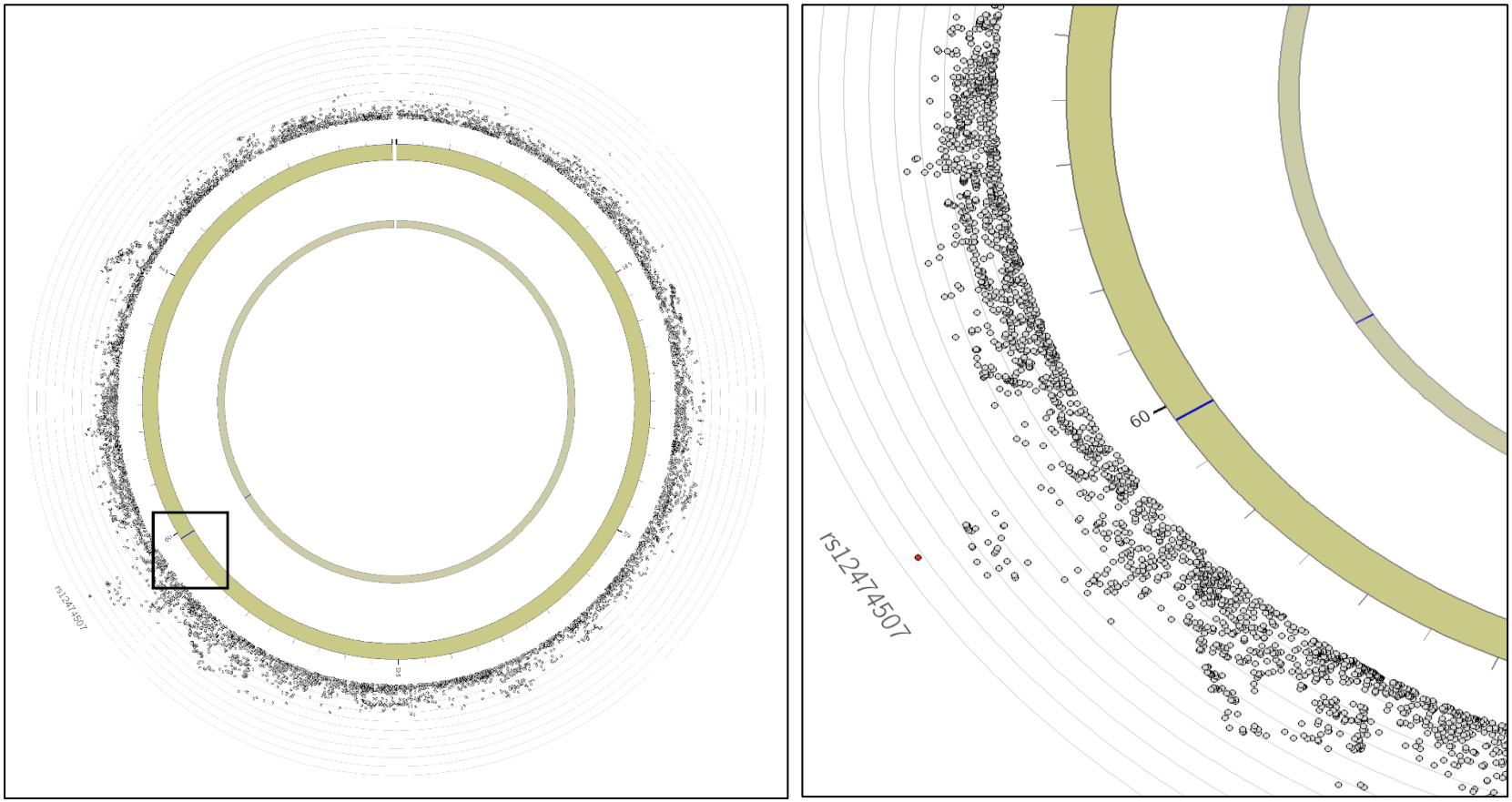



**Figure 4.16: Chromosome 2 circos plot showing mapped genes by eQTL and chromatin interactions**
Genomic loci are highlighted in blue. Orange represents genes that are mapped by chromatin interaction and green represents eQTL mapping. If genes are mapped by both, they are highlighted in red. The dark blue portion of the inner circles represents the loci. There are no mapped genes by eQTL and chromatin interactions on Chromosome 2. Figure (a) is the full chromosome. Figure (b) is an enlargement of the region in the box in figure 2a.

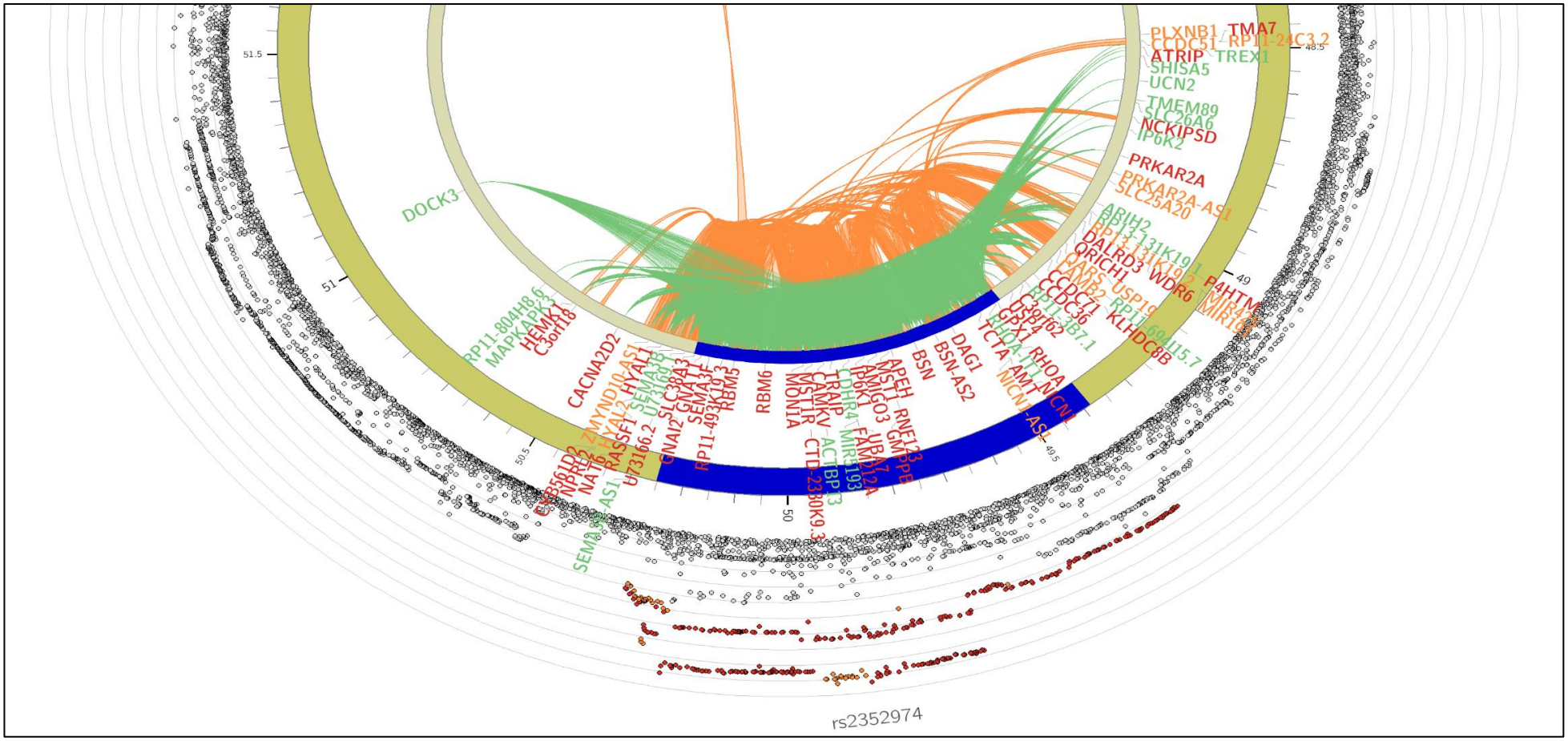



**Figure 4.17: Chromosome 3 partial circos plot showing mapped genes by eQTL and chromatin interactions**
Genomic loci are highlighted in blue. Orange represents genes that are mapped by chromatin interaction and green represents eQTL mapping. If genes are mapped by both, they are highlighted in red. The dark blue portion of the inner circles represents the loci. All interactions are showed except for one chromatin interaction between RBM5 3:50126341-50156454 to PBRM1 3:52579368-52719933.

(a) (b)

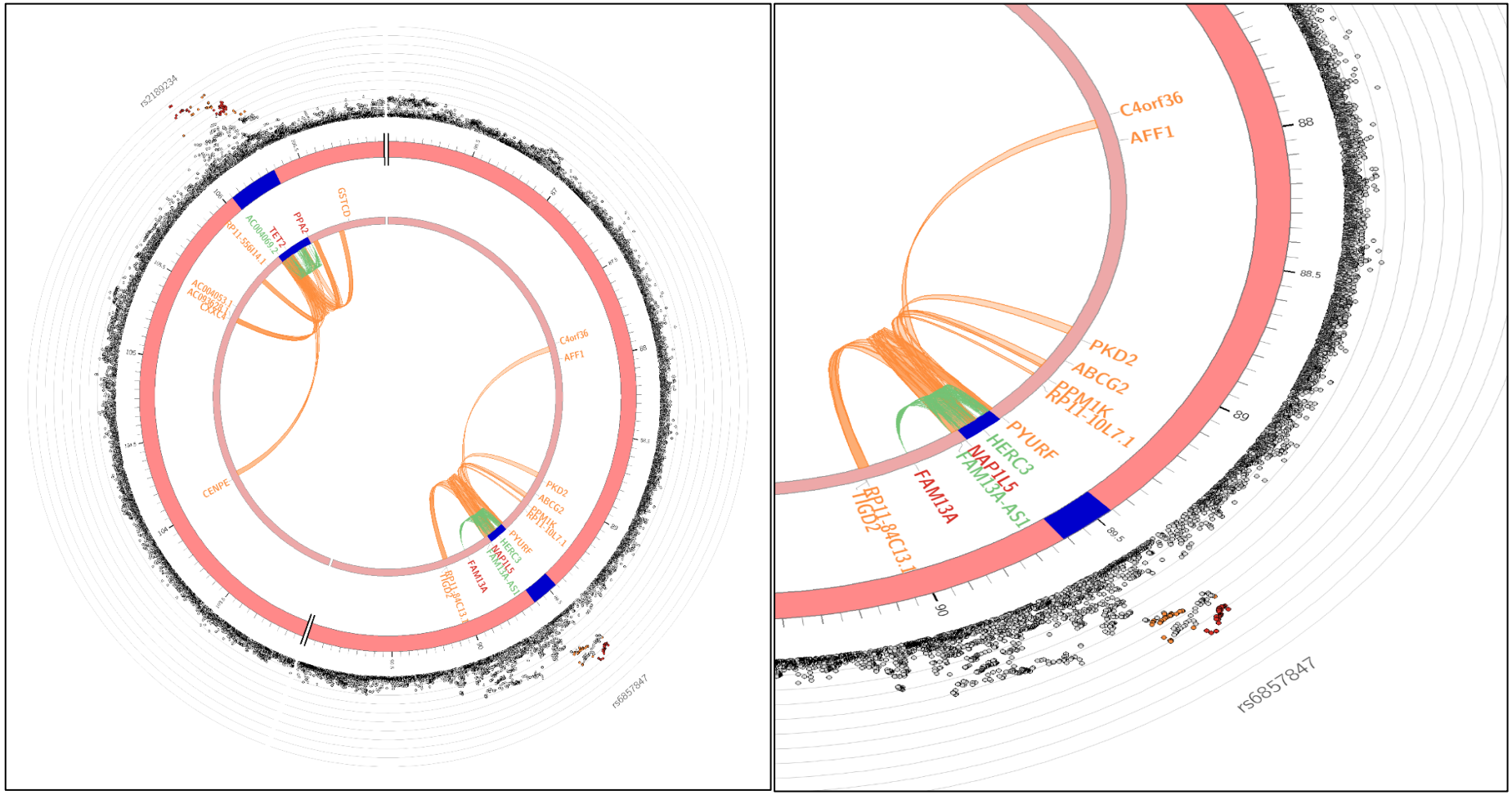



**Figure 4.18: Chromosome 4 circos plot showing mapped genes by eQTL and chromatin interactions**
Genomic loci are highlighted in blue. Orange represents genes that are mapped by chromatin interaction and green represents eQTL mapping. If genes are mapped by both, they are highlighted in red. The dark blue portion of the inner circles represents the loci. There are two loci associated with *Resilience* on chromosome 4 - locus 4A and 4B. Figure (a) shows the full circos plot for chromosome 4 and figure (b) is and enlargement of the locus 4A.

(a) (b)

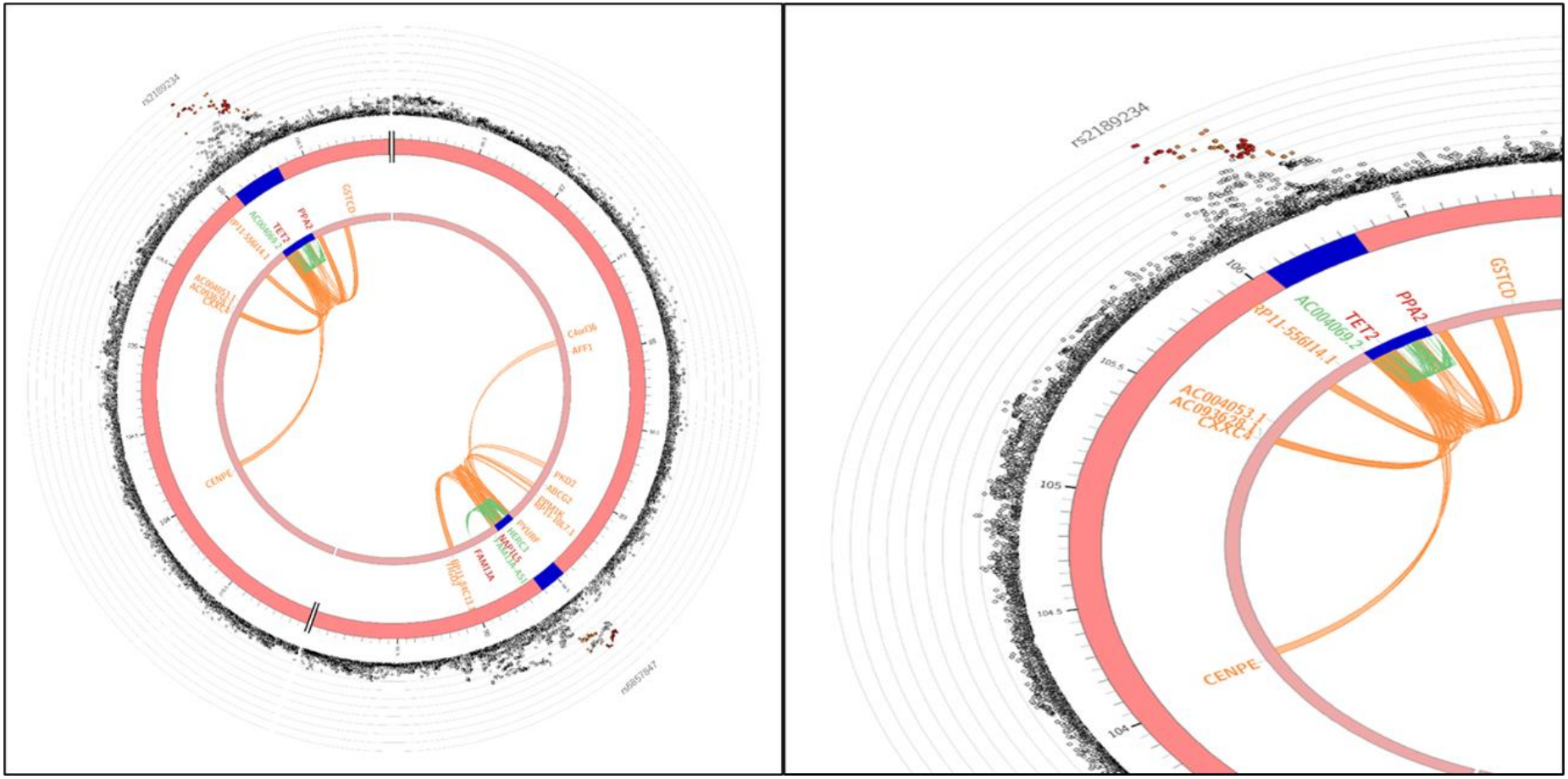



**Figure 4.19: Chromosome 4 circos plot showing mapped genes by eQTL and chromatin interactions**
Genomic loci are highlighted in blue. Orange represents genes that are mapped by chromatin interaction and green represents eQTL mapping. If genes are mapped by both, they are highlighted in red. The dark blue portion of the inner circles represents the loci. There are two loci associated with Resilience on chromosome 4 - locus 4A and 4B. F Figure (a) shows the full circos plot for chromosome 4 and figure (b) is and enlargement of the locus 4B.

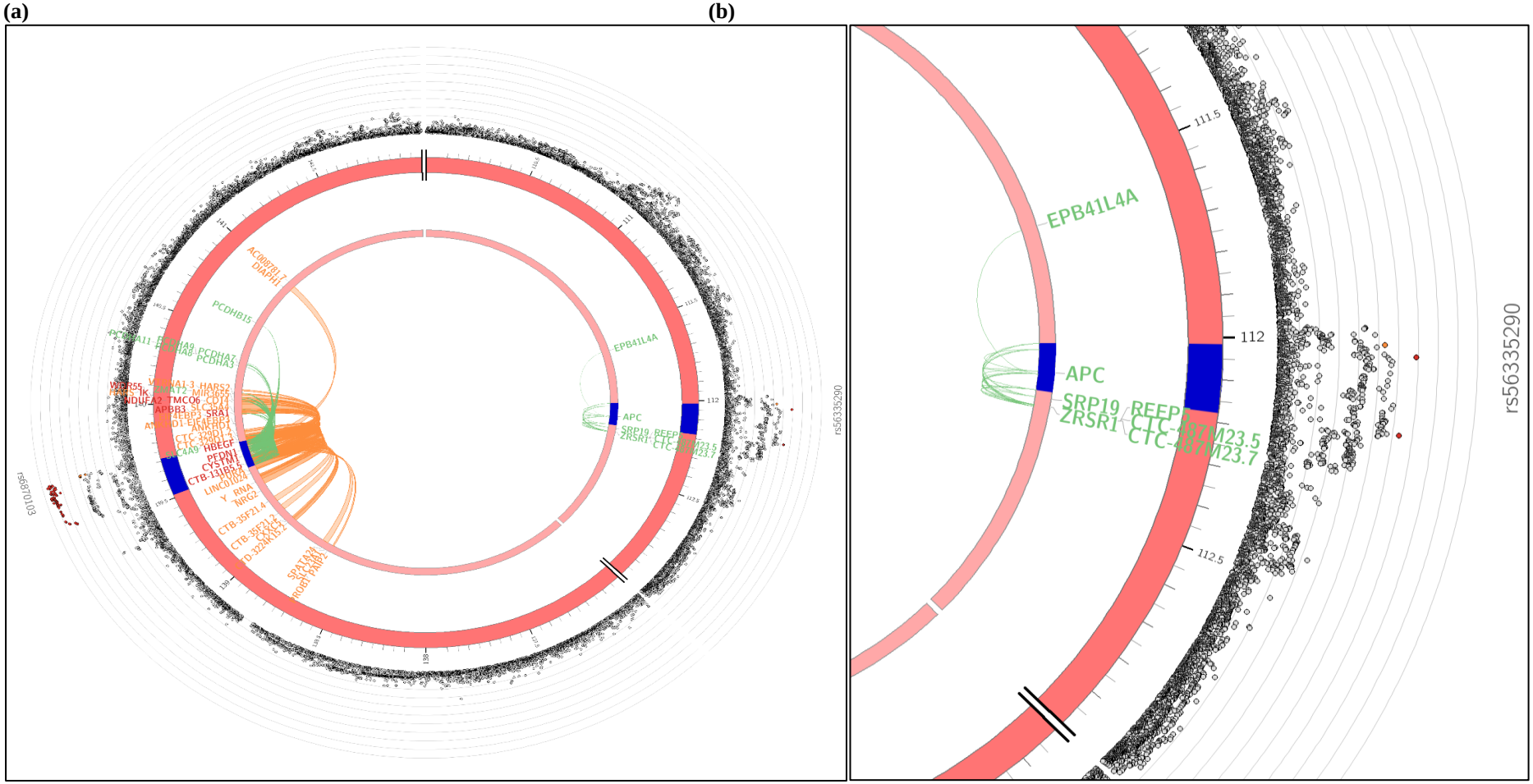



**Figure 4.19: Chromosome 5 circos plot showing mapped genes by eQTL and chromatin interactions**
Genomic loci are highlighted in blue. Orange represents genes that are mapped by chromatin interaction and green represents eQTL mapping. If genes are mapped by both, they are highlighted in red. The dark blue portion of the inner circles represents the loci. There are two loci associated with *Resilience* on chromosome 5 - locus 5A and 5B. This figure includes full circos plot of chromosome 5 (a) a partial circos plot of locus 5A (b).

(a) (b)

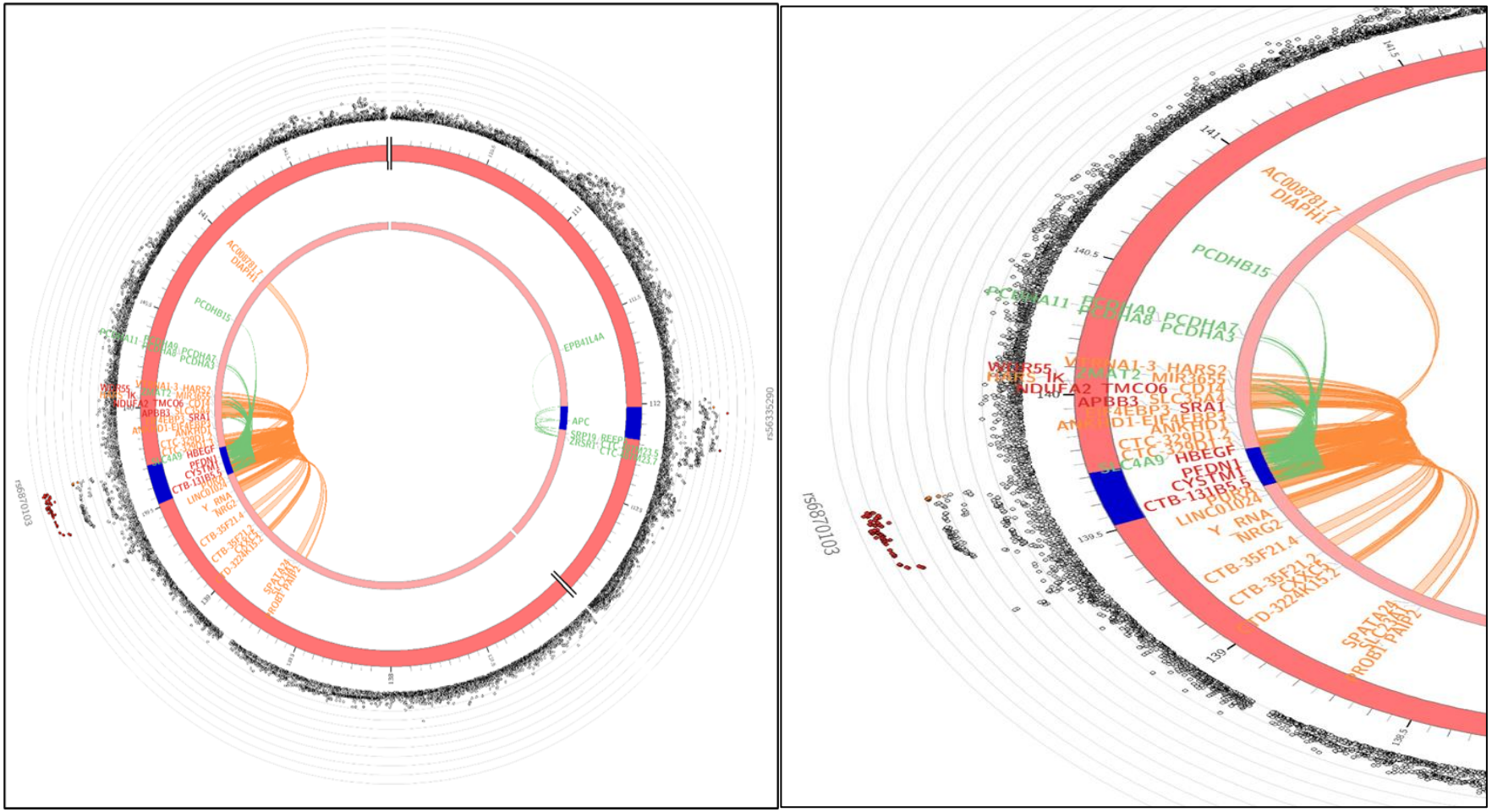



**Figure 4.20: Chromosome 5 circos plot showing mapped genes by eQTL and chromatin interactions**
Genomic loci are highlighted in blue. Orange represents genes that are mapped by chromatin interaction and green represents eQTL mapping. If genes are mapped by both, they are highlighted in red. The dark blue portion of the inner circles represents the loci. There are two loci associated with Resilience on chromosome 5 - locus 5A and 5B. This figure includes full circos plot of chromosome 5 (a) a partial circos plot of locus 5B (b).

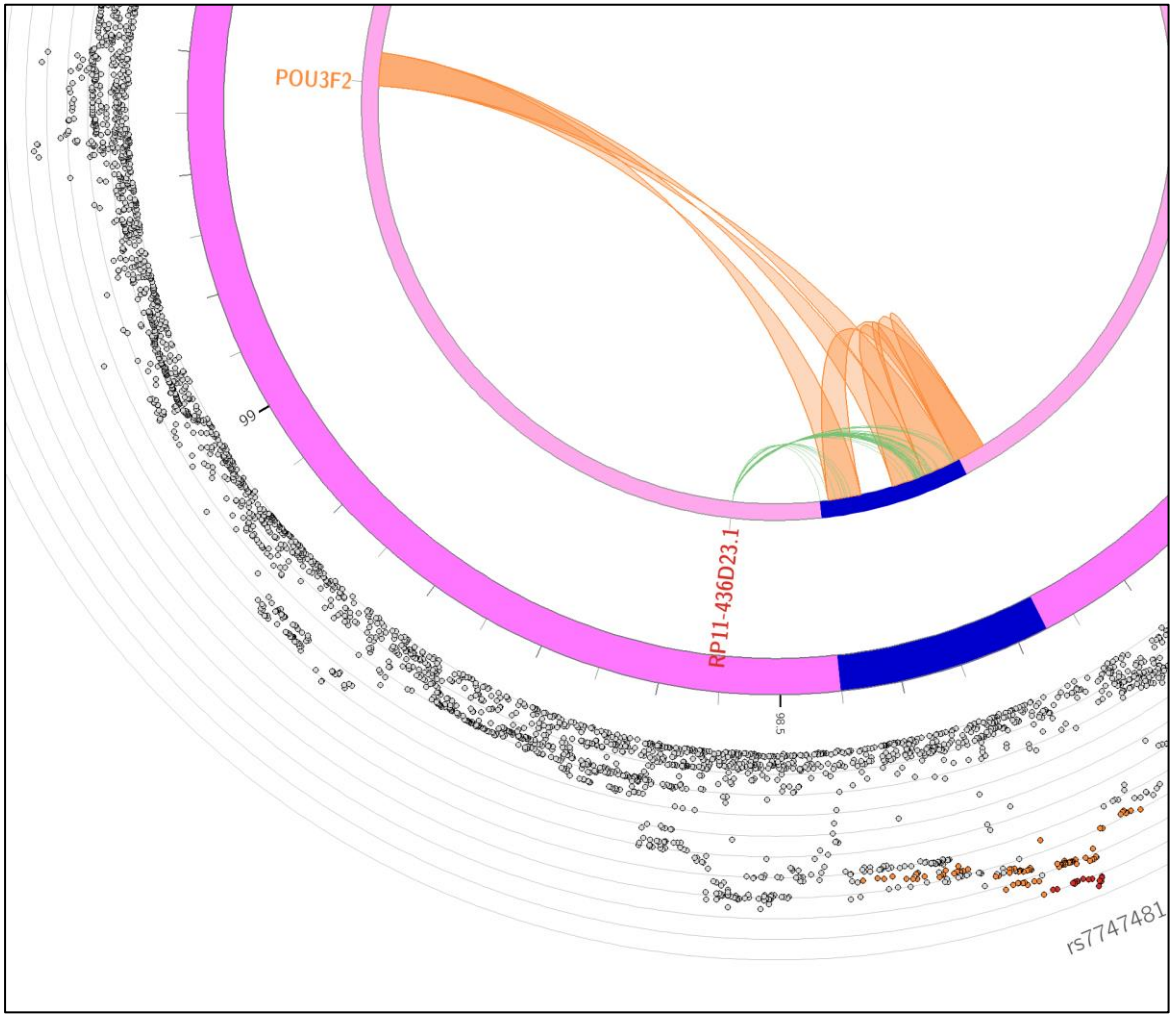



**Figure 4.21: Partial circos plot showing locus 6 on chromosome 6 showing mapped genes by eQTL and chromatin interactions**
Genomic loci are highlighted in blue. Orange represents genes that are mapped by chromatin interaction and green represents eQTL mapping. If genes are mapped by both, they are highlighted in red. The dark blue portion of the inner circles represents the loci.

(a) (b)

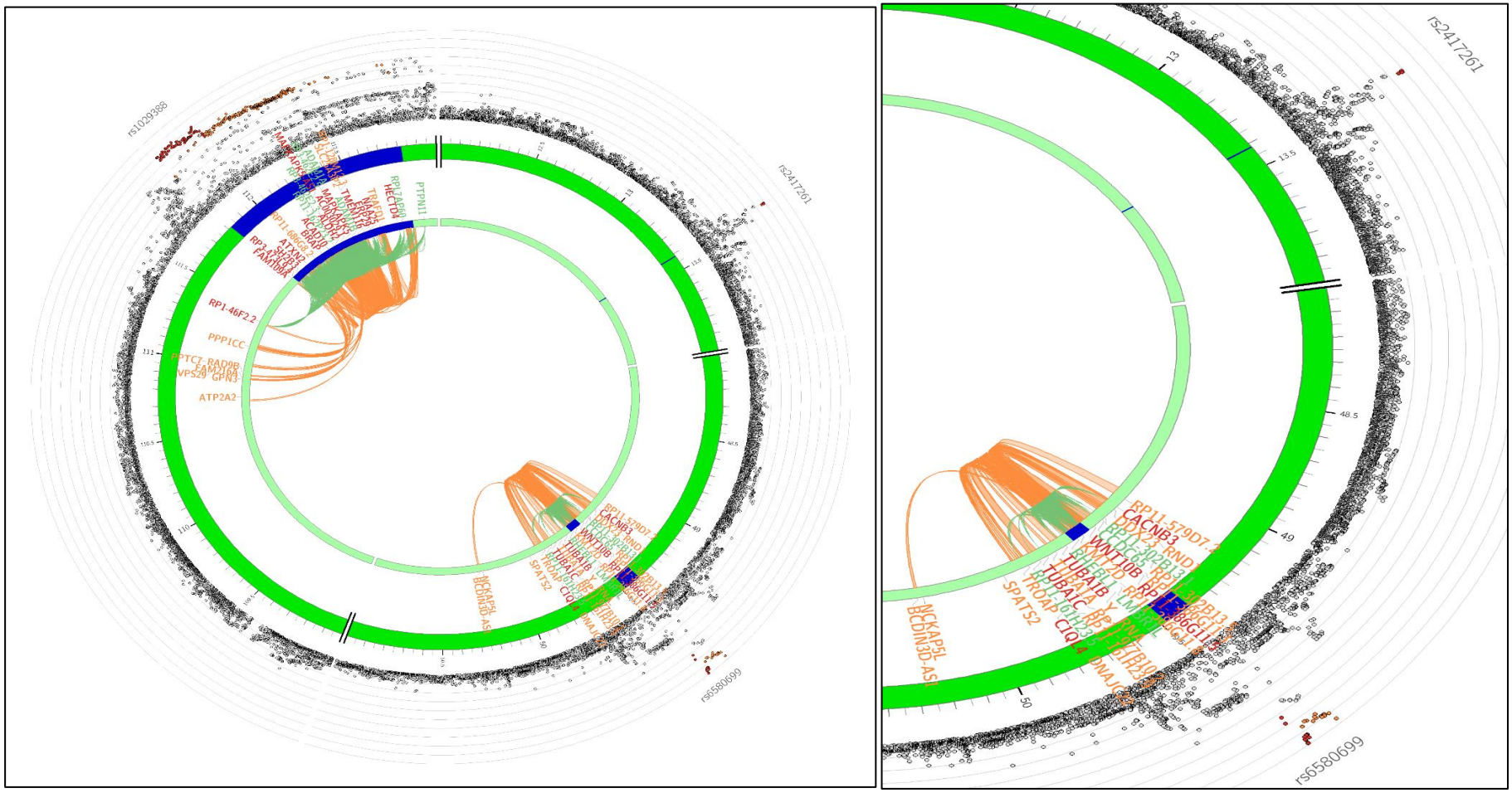



**Figure 4.22: Chromosome 12 circos plot showing mapped genes by eQTL and chromatin interactions**
Genomic loci are highlighted in blue. Orange represents genes that are mapped by chromatin interaction and green represents eQTL mapping. If genes are mapped by both, they are highlighted in red. The dark blue portion of the inner circles represents the loci. There are three loci associated with *Resilience* on chromosome 12 - locus 12A,12B and 12C. This figure includes a full plot of all three loci (a) and a partial circos plot of 12A (rs2417261) and 12B (rs6580699) (b).

(a) (b)

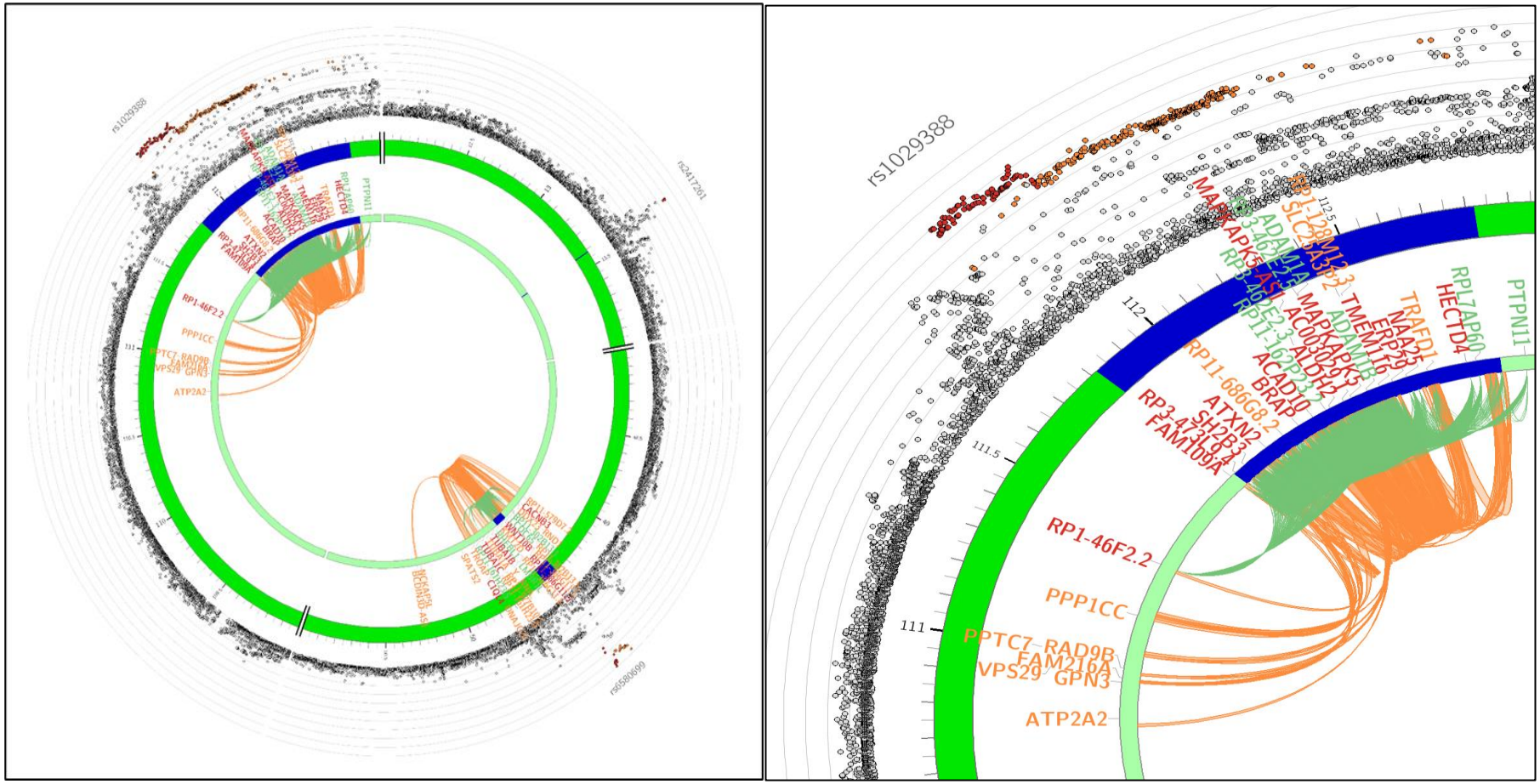



**Figure 4.23: Chromosome 12 circos plot showing mapped genes by eQTL and chromatin interactions.**
Genomic loci are highlighted in blue. Orange represents genes that are mapped by chromatin interaction and green represents eQTL mapping. If genes are mapped by both, they are highlighted in red. The dark blue portion of the inner circles represents the loci. There are three loci associated with Resilience on chromosome 12 - locus 12A,12B and 12C. This figure includes a full plot of all three loci (a) and a partial circos plot of 12C (b).

(a) (b)

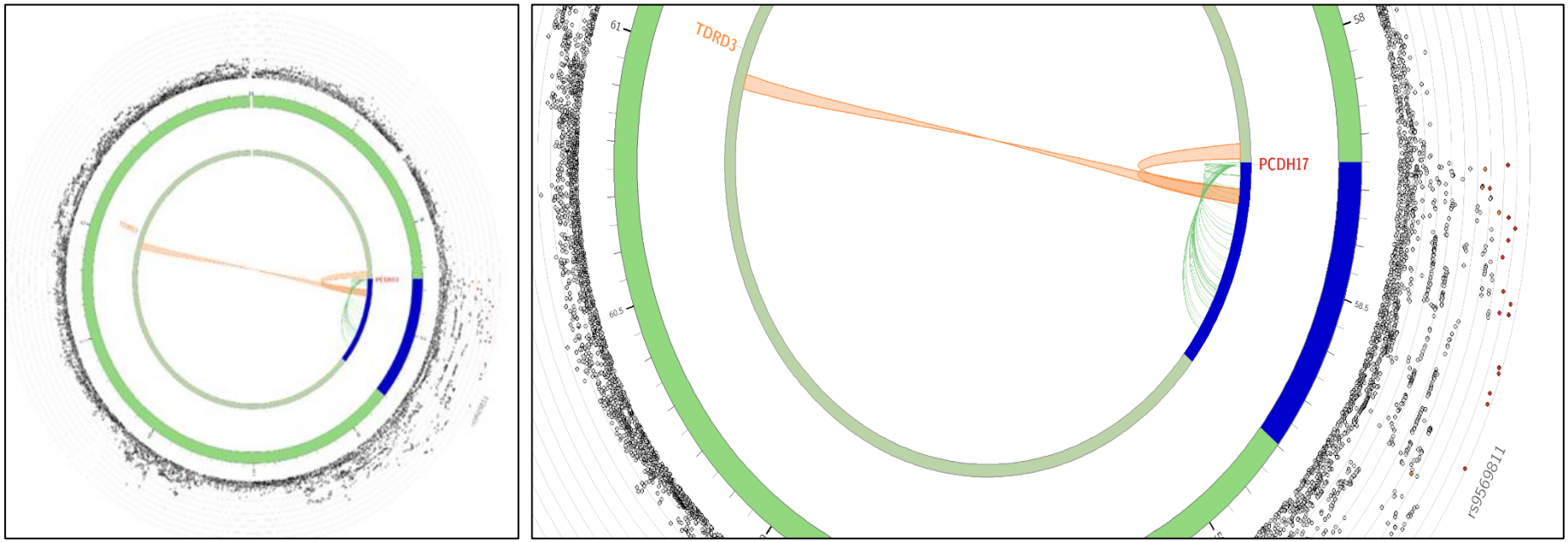



**Figure 4.24: Chromosome 13 circos plot showing mapped genes by eQTL and chromatin interactions**
Genomic loci are highlighted in blue. Orange represents genes that are mapped by chromatin interaction and green represents eQTL mapping. If genes are mapped by both, they are highlighted in red. The dark blue portion of the inner circles represents the loci (a). This figure includes a partial circos plot of 13 showing more detail (b).

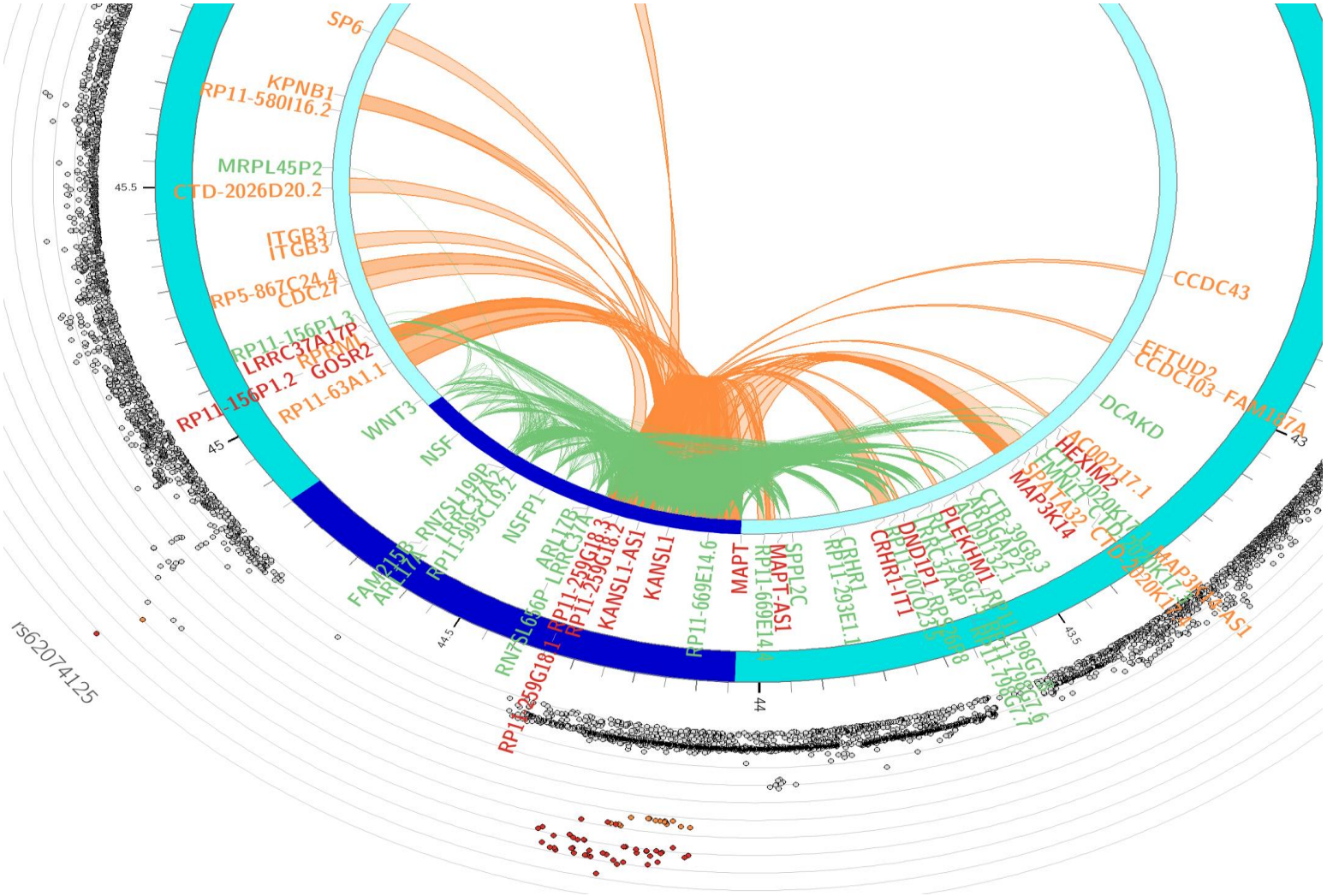



**Figure 4.25: Chromosome 17 partial circos plot showing mapped genes by eQTL and chromatin interactions**
Genomic loci are highlighted in blue. Orange represents genes that are mapped by chromatin interaction and green represents eQTL mapping. If genes are mapped by both, they are highlighted in red. The dark blue portion of the inner circles represents the loci. There is one interaction that is not included and that is a chromatin interaction between rs10775404 17:44224272 and HOXB 17:44605888 – 46683776.

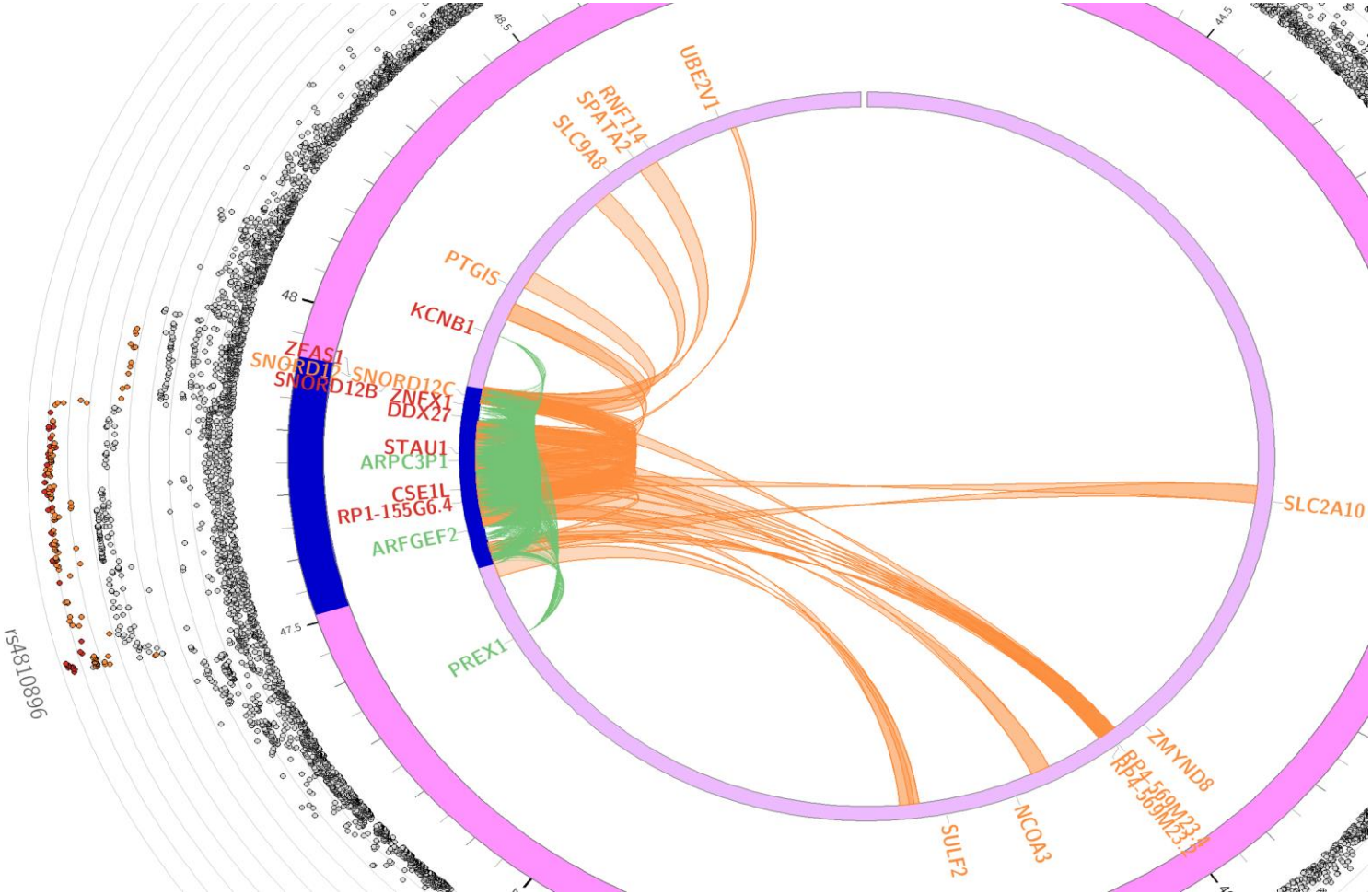



**Figure 4.26: Chromosome 20 circos plot showing mapped genes by eQTL and chromatin interactions**
Genomic loci are highlighted in blue. Orange represents genes that are mapped by chromatin interaction and green represents eQTL mapping. If genes are mapped by both, they are highlighted in red. The dark blue portion of the inner circles represents the loci.

### 4.2.4 Genome-wide gene-based association analysis (GWGAS)

In addition to the three approaches above we also performed a genome-wide gene-based association analysis (GWGAS) using the MAGMA function within FUMA (Watanabe et al., 2017), which looks at the aggregate association results of all SNPs in a gene in contrast to the previous analyses that examined the association signals at the level of individual SNPs. A GWGAS was performed using the *Resilience* GWAS on 18,879 protein-coding genes containing at least one SNP from the GWAS. Based on the number of genes tested, a Bonferroni-corrected threshold of $P < 2.65 \times 10^{-6}$ was used (see Q-Q plot of this association – Figure 4.27).

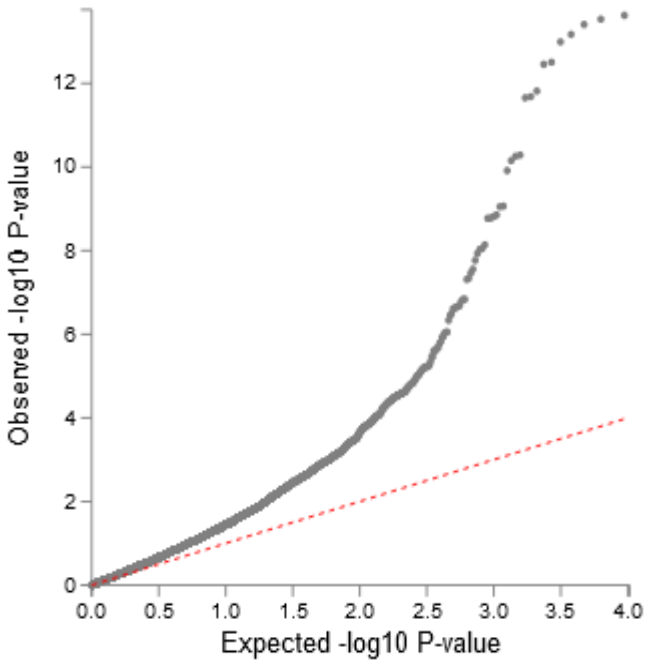



**Figure 4.27: Q-Q plot of the gene-based test computed by MAGMA**
The x-axis illustrates the expected distribution of p-values from the association test across all SNPs, and the y-axis shows the observed p-values. The genomic inflation factor (lambda) is 1.30.

A total of 52 protein coding genes were identified as significantly associated (Table 4.4), 40 of which were identified by the previously described strategies. A full list of results for all 18,897 protein coding genes is in Supplementary Table 10 (Fitzgerald et al., 2021).

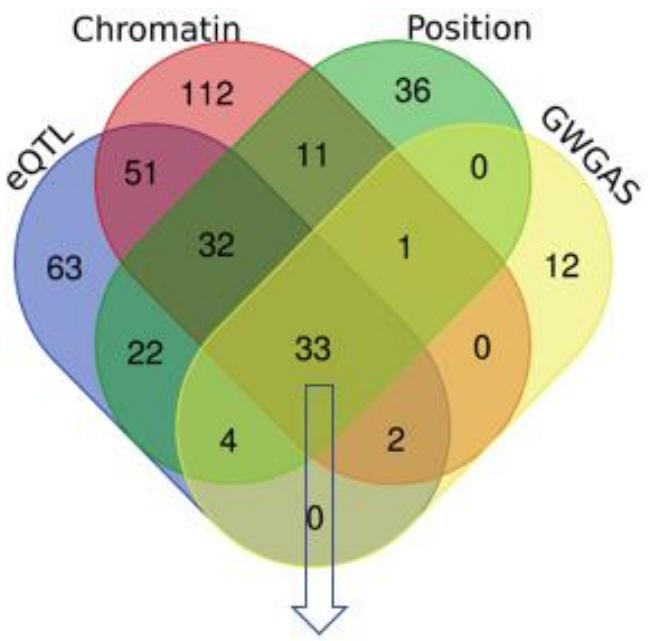
**Table 4.4: Genes significantly associated with *Resilience***

| GENE | Chr | Start | End | nSNPs | [b]Z stat | P value | Symbol |
|---|---|---|---|---|---|---|---|
| ENSG00000164068 | 3 | 49726932 | 49758962 | 64 | 7.5392 | 2.36E-14 | RNF123 |
| ENSG00000183763 | 3 | 49866034 | 49894007 | 45 | 7.5114 | 2.93E-14 | TRAIP |
| ENSG00000176095 | 3 | 49761727 | 49823975 | 135 | 7.4732 | 3.91E-14 | IP6K1 |
| ENSG00000004534 | 3 | 49977440 | 50137478 | 321 | 7.4014 | 6.74E-14 | RBM6 |
| ENSG00000164078 | 3 | 49924435 | 49941299 | 30 | 7.3465 | 1.02E-13 | MST1R |
| ENSG00000164076 | 3 | 49895421 | 49907655 | 20 | 7.1956 | 3.11E-13 | CAMKV |
| ENSG00000003756 | 3 | 50126341 | 50156454 | 35 | 7.1795 | 3.50E-13 | RBM5 |
| ENSG00000187492 | 3 | 49828165 | 49837268 | 26 | 6.9757 | 1.52E-12 | CDHR4 |
| ENSG00000228008 | 3 | 49941278 | 49954370 | 16 | 6.9327 | 2.06E-12 | CTD-2330K9.3 |
| ENSG00000164077 | 3 | 49946302 | 49967606 | 32 | 6.9226 | 2.22E-12 | MON1A |
| ENSG00000181418 | 12 | 49388932 | 49393092 | 6 | 6.462 | 5.17E-11 | DDN |
| ENSG00000113068 | 5 | 139624624 | 139682706 | 107 | 6.4508 | 5.56E-11 | PFDN1 |
| ENSG00000182179 | 3 | 49842640 | 49851379 | 16 | 6.4152 | 7.03E-11 | UBA7 |
| ENSG00000001617 | 3 | 50192478 | 50226508 | 65 | 6.3309 | 1.22E-10 | SEMA3F |
| ENSG00000120306 | 5 | 139554227 | 139661637 | 236 | 6.0215 | 8.64E-10 | CYSTM1 |
| ENSG00000124207 | 20 | 47662849 | 47713489 | 141 | 6.0163 | 8.92E-10 | CSE1L |
| ENSG00000145022 | 3 | 49449639 | 49453908 | 8 | 5.9428 | 1.40E-09 | TCTA |
| ENSG00000181929 | 12 | 49396057 | 49412980 | 18 | 5.9267 | 1.55E-09 | PRKAG1 |
| ENSG00000167548 | 12 | 49412758 | 49453557 | 48 | 5.9137 | 1.67E-09 | KMT2D |
| ENSG00000124214 | 20 | 47729878 | 47804904 | 182 | 5.9119 | 1.69E-09 | STAU1 |
| ENSG00000204842 | 12 | 111890018 | 112037480 | 199 | 5.6674 | 7.25E-09 | ATXN2 |
| ENSG00000173402 | 3 | 49506146 | 49573048 | 118 | 5.633 | 8.85E-09 | DAG1 |
| ENSG00000111252 | 12 | 111843752 | 111889427 | 56 | 5.624 | 9.33E-09 | SH2B3 |
| ENSG00000233276 | 3 | 49394609 | 49396033 | 2 | 5.5844 | 1.17E-08 | GPX1 |
| ENSG00000145020 | 3 | 49454211 | 49460186 | 8 | 5.5184 | 1.71E-08 | AMT |
| ENSG00000164061 | 3 | 49591922 | 49708978 | 201 | 5.432 | 2.79E-08 | BSN |
| ENSG00000164483 | 6 | 130465460 | 130686570 | 1090 | 5.3953 | 3.42E-08 | SAMD3 |
| ENSG00000173531 | 3 | 49721380 | 49726934 | 7 | 5.3428 | 4.58E-08 | MST1 |
| ENSG00000168769 | 4 | 106067032 | 106200973 | 361 | 5.3303 | 4.90E-08 | TET2 |
| ENSG00000111271 | 12 | 112123857 | 112194903 | 96 | 5.1285 | 1.46E-07 | ACAD10 |
| ENSG00000145029 | 3 | 49460379 | 49466759 | 6 | 5.1283 | 1.46E-07 | NICN1 |
| ENSG00000257767 | 12 | 112191694 | 112229222 | 46 | 5.1068 | 1.64E-07 | RP11-162P23.2 |
| ENSG00000089234 | 12 | 112079950 | 112123790 | 62 | 5.0607 | 2.09E-07 | BRAP |
| ENSG00000111275 | 12 | 112204691 | 112247782 | 57 | 5.0567 | 2.13E-07 | ALDH2 |
| ENSG00000114349 | 3 | 50229045 | 50233949 | 3 | 5.0471 | 2.24E-07 | GNAT1 |
| ENSG00000138641 | 4 | 89442199 | 89629693 | 486 | 5.0336 | 2.41E-07 | HERC3 |
| ENSG00000138640 | 4 | 89647106 | 90032549 | 1224 | 5.0298 | 2.46E-07 | FAM13A |
| ENSG00000067560 | 3 | 49396578 | 49450431 | 107 | 4.9771 | 3.23E-07 | RHOA |
| ENSG00000124198 | 20 | 47538427 | 47653230 | 318 | 4.9599 | 3.53E-07 | ARFGEF2 |
| ENSG00000089022 | 12 | 112279782 | 112334343 | 135 | 4.9066 | 4.63E-07 | MAPKAPK5 |
| ENSG00000134569 | 11 | 46878419 | 46940193 | 123 | 4.7789 | 8.81E-07 | LRP4 |
| ENSG00000165912 | 11 | 47199076 | 47207994 | 19 | 4.7726 | 9.09E-07 | PACSIN3 |
| ENSG00000198270 | 12 | 112369086 | 112450970 | 177 | 4.7667 | 9.36E-07 | TMEM116 |

| GENE | Chr | Start | End | nSNPs | [b]Z stat | P value | Symbol |
|---|---|---|---|---|---|---|---|
| ENSG00000175216 | 11 | 46764598 | 46867847 | 178 | 4.7276 | 1.14E-06 | CKAP5 |
| ENSG00000129158 | 11 | 17809595 | 18034709 | 410 | 4.7157 | 1.20E-06 | SERGEF |
| ENSG00000132535 | 17 | 7093209 | 7123021 | 56 | 4.6671 | 1.53E-06 | DLG4 |
| ENSG00000149179 | 11 | 46958240 | 47185936 | 358 | 4.6491 | 1.67E-06 | C11orf49 |
| ENSG00000100985 | 20 | 44637547 | 44645200 | 28 | 4.6029 | 2.08E-06 | MMP9 |
| ENSG00000196141 | 2 | 201170604 | 201346986 | 466 | 4.5994 | 2.12E-06 | SPATS2L |
| ENSG00000165118 | 9 | 86553226 | 86571901 | 45 | 4.5751 | 2.38E-06 | C9orf64 |
| ENSG00000138430 | 2 | 174937175 | 175113426 | 570 | 4.5657 | 2.49E-06 | OLA1 |
| ENSG00000072778 | 17 | 7120444 | 7128592 | 14 | 4.5648 | 2.50E-06 | ACADVL |

Note: Resilience GWAS SNPs were assigned to 18879 protein coding genes. Significant associations (P = 0.05/18879 = 2.648e-6). [a]nPARAM = the number of relevant parameters used in the model, [b]Z stat = the Z-value for the gene based on its p-value

In total, 33 genes were identified by all four mapping strategies (Figure 4.28 and Table 4.5).




ACAD10, ALDH2, AMT, ATXN2, BRAP, BSN, CAMKV, CSE1L, CTD-2330K9.3, MAPKAPK5, MON1A,CYSTM1, DAG1, DDN, GNAT1, GPX, IP6K1, MST1 RBM5, RHOA, RNF123, MST1R, NICN1, PFDN1, PRKAG1, SEMA3F, SH2B3, STAU1, TCTA, TET2, TMEM116, TRAIP, UBA7

**Figure 4.28: Venn diagram of overlapping mapped genes**
Showing 33 genes were mapped by all four strategies. These genes are listed underneath.

**Table 4.5: Mapped genes by all 4 strategies- position, eQTL, chromatin interaction mapping and GWGAS.**

| symbol | Locus | GWGAS P | [a]pos Map SNPs | [b]pos Map Max CADD | [c]eqtl Map SNPs | [d]min GWAS P |
|---|---|---|---|---|---|---|
| **SEMA3F** | 3 | 1.22E-10 | 43 | 19.09 | 435 | 9.27E-15 |
| **RBM5** | 3 | 3.50E-13 | 33 | 18.47 | 262 | 9.27E-15 |
| **RHOA** | 3 | 6.74E-14 | 38 | 19.22 | 369 | 9.27E-15 |
| **MAPKAPK5** | 12C | 3.23E-07 | 119 | 15.93 | 402 | 6.50E-09 |
| **BRAP** | 12C | 2.50E-06 | 21 | 11.71 | 395 | 6.50E-09 |
| **SH2B3** | 12C | 4.63E-07 | 34 | 21.4 | 209 | 6.50E-09 |
| **ACAD10** | 12C | 2.09E-07 | 45 | 11.71 | 322 | 6.50E-09 |
| **ALDH2** | 12C | 2.08E-06 | 26 | 14.56 | 400 | 6.50E-09 |
| **PFDN1** | 5B | 9.33E-09 | 56 | 14.14 | 89 | 4.03E-11 |
| **GNAT1** | 3 | 1.46E-07 | 20 | 18.21 | 107 | 1.20E-14 |
| **CYSTM1** | 5B | 2.13E-07 | 68 | 16.53 | 88 | 4.03E-11 |
| **CSE1L** | 20 | 5.56E-11 | 54 | 15.91 | 192 | 1.94E-10 |
| **STAU1** | 20 | 2.24E-07 | 75 | 11.52 | 192 | 1.94E-10 |
| **AMT** | 3 | 8.64E-10 | 10 | 19.06 | 460 | 9.27E-15 |
| **TCTA** | 3 | 3.53E-07 | 11 | 19.06 | 399 | 9.27E-15 |
| **NICN1** | 3 | 8.92E-10 | 12 | 19.06 | 456 | 9.27E-15 |
| **BSN** | 3 | 1.69E-09 | 46 | 12.52 | 304 | 1.20E-14 |
| **RNF123** | 3 | 1.20E-06 | 36 | 21 | 465 | 9.27E-15 |
| **CAMKV** | 3 | 1.53E-06 | 24 | 18.51 | 410 | 9.27E-15 |
| **MON1A** | 3 | 8.81E-07 | 27 | 13.15 | 340 | 9.27E-15 |
| **MST1R** | 3 | 2.49E-06 | 21 | 18.51 | 465 | 9.27E-15 |
| **TET2** | 4B | 2.46E-07 | 26 | 13.4 | 61 | 3.14E-12 |
| **DAG1** | 3 | 2.41E-07 | 35 | 15.6 | 258 | 9.27E-15 |
| **MST1** | 3 | 1.71E-08 | 12 | 21 | 373 | 9.27E-15 |
| **IP6K1** | 3 | 1.40E-09 | 86 | 17.96 | 385 | 9.27E-15 |
| **DDN** | 12B | 1.46E-07 | 8 | 10.57 | 34 | 1.29E-10 |
| **PRKAG1** | 12B | 1.67E-06 | 18 | 10.57 | 35 | 1.29E-10 |
| **UBA7** | 3 | 2.79E-08 | 18 | 13.02 | 467 | 9.27E-15 |
| **TRAIP** | 3 | 2.36E-14 | 37 | 17.61 | 115 | 9.27E-15 |
| **TMEM116** | 12C | 3.11E-13 | 72 | 17.1 | 402 | 6.50E-09 |
| **ATXN2** | 12C | 2.22E-12 | 99 | 18.59 | 287 | 6.50E-09 |
| **CTD-2330K9.3** | 3 | 1.02E-13 | 20 | 13.15 | 419 | 9.27E-15 |
| **GPX1** | 3 | 3.42E-08 | 11 | 19.22 | 399 | 9.27E-15 |

Note: [a]pos Map = positional mapping, [b]CADD = Combined Annotation-Dependent depletion score, [c]eqtl Map = eqtl mapping, [d]min Gwas P = minimum GWAS P value of SNP(s)associated with the gene.

Many of these 33 genes have been connected with cognitive performance, neurodegenerative disorders or ageing and represent potential therapeutic targets: *STAU1* (chr 20) and *SEMA3F* (chr 3) are predicted to control cognitive decline in ageing through formation of neural circuits and synaptic transmission (Tasaki et al., 2018). *BSN* (chr 3) codes for bassoon presynaptic cytomatrix protein which is implicated in the regulation of neurotransmitters at inhibitory and excitatory synapses (Annamneedi et al., 2018). *IP6K1* (chr 3) codes for inositol pyrophosphate biosynthesis, and mouse studies have shown its involvement in short term memory by altering presynaptic vesicle release and short-term facilitation of glutamatergic synapses in the hippocampus (Kim et al., 2020). *MST1* (chr 3) has been shown to play a role in protecting cells from oxidative stress which leads to ageing and eventual cell death (Wang et al., 2019). *TET2* (chr 4) codes for ten eleven translocation methyl cytosine dioxygenase 2 which catalyses the production of 5-hydroxymethylcytosine and is associated with increased neurogenesis in the hippocampus and cognition in animal studies (Gontier et al., 2018). *ATXN2* (chr 20) is involved in regulating mRNA and is linked to decline in cognitive function in older adults (Gardiner et al., 2019), general cognitive function (Davies et al., 2018) and neurodegenerative disorders (Lastres-Becker, Nonis, Nowock, & Auburger, 2019). The *ATXN2/BRAP* locus has a strong association with parental lifespan (Timmers et al., 2019). Another mapped gene close to *ATXN2* and *BRAP* is *SH2B3*, which encodes lymphocyte adaptor protein LNK, and plays a role in human ageing though the mechanism involved is not fully understood (Melzer, Pilling, & Ferrucci, 2020). The gene *ALDH2* (chr 12) codes for aldehyde dehydrogenase and there is a link between this enzyme and life span as well as cardiovascular ageing (Wu & Ren, 2019).

Among the associated SNPs at the 33 prioritized genes are two UTR3 variants on chromosome 3 (rs2681781 (CADD=17.77) and rs4625 (CADD=15.6)) that map to *RBM5* and *DAG1* respectively. Animal studies have shown that *RBM5* is a likely regulator of Rab4a, which in involved in many neurobiological functions including the transport of transmembrane proteins required for neurotransmission (Jackson, Kotermanski, & Kochanek, 2017). *DAG1* has been associated with increased cognitive performance and is associated with GABAergic signalling in the hippocampus (Panzanelli, Früh, & Fritschy, 2017). In addition, one other variant of note is rs1130146 that maps to *DDX27* (chr 20), a gene that was mapped by all strategies except for GWGAS and is associated with longevity (McLaren et al., 2016). This missense SNP has a CADD score of 31 and is predicted by SIFT to be deleterious and by PolyPhen to be possibly damaging.

### 4.2.5 Tissue, cell type and pathway enrichment analysis

(See method section 2.1.10.1)

#### 4.2.5.1 Gene – tissue expression

Using gene expression data for 53 tissues obtained from GTEx (Consortium, 2015), I found all brain regions to be significantly enriched for our associated genes with the strongest enrichments for the frontal cortex, BA9 ($P$ = 2.26 x $10^{-11}$), the cortex ($P$ = 8.48 x $10^{-11}$) and the cerebellar hemisphere ($P$ = 1.18 x $10^{-10}$); (Figure 4.29, Figure 4.30). Table 4.6 shows the results for the 53 tissue types. There was no significant enrichment in other non-brain tissues of the body .

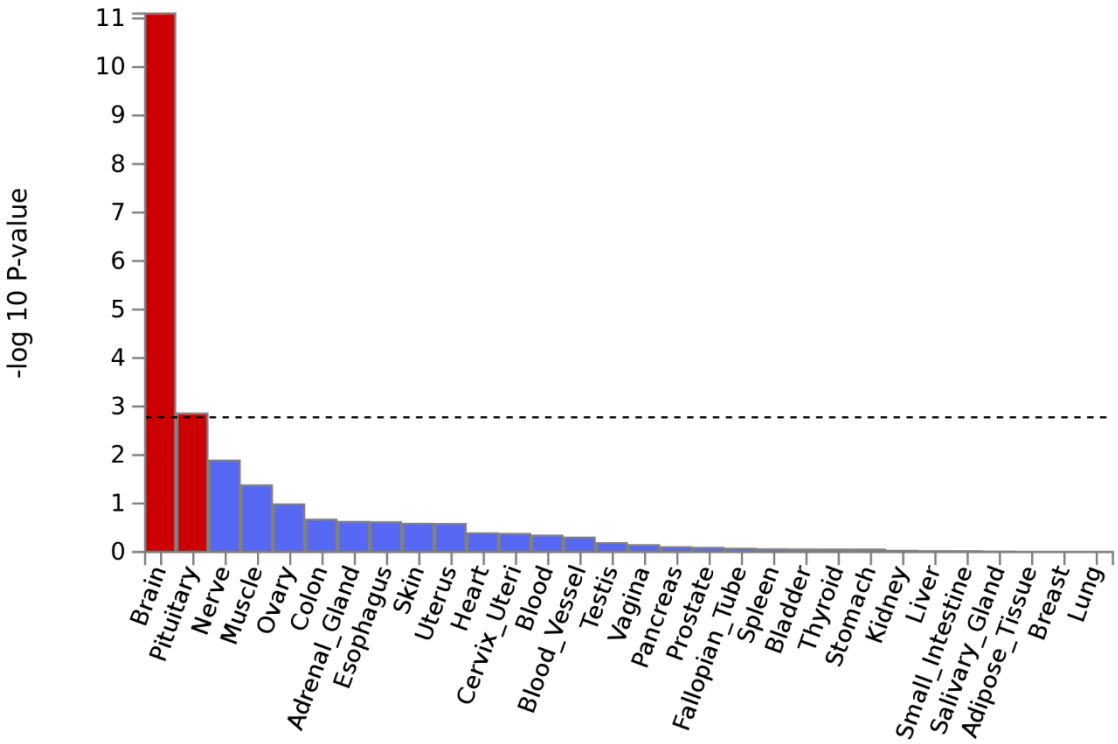



**Figure 4.29: MAGMA expression analysis across all tissue types.**
The y-axis shows the -log 10 transformed P-values of the GWGAS and the x-axis shows the tissue location. Significant associations are shown in red.

**Table 4.6: Expression of mapped genes for the 53 tissue samples analysed.**

| Tissue | Beta | SE | P value |
|---|---|---|---|
| **Brain_Frontal_Cortex_BA9** | **0.050** | **0.008** | **2.26E-11** |
| **Brain_Cortex** | **0.050** | **0.008** | **8.48E-11** |
| **Brain_Cerebellar_Hemisphere** | **0.042** | **0.007** | **1.81E-10** |
| **Brain_Cerebellum** | **0.042** | **0.007** | **4.07E-10** |
| **Brain_Anterior_cingulate_cortex_BA24** | **0.048** | **0.008** | **4.11E-10** |
| **Brain_Nucleus_accumbens_basal_ganglia** | **0.046** | **0.008** | **1.73E-08** |
| **Brain_Hypothalamus** | **0.047** | **0.009** | **5.38E-08** |
| **Brain_Amygdala** | **0.044** | **0.009** | **1.01E-07** |
| **Brain_Caudate_basal_ganglia** | **0.044** | **0.009** | **1.67E-07** |
| **Brain_Hippocampus** | **0.043** | **0.009** | **3.04E-07** |
| **Brain_Putamen_basal_ganglia** | **0.043** | **0.009** | **3.88E-07** |
| Brain_Substantia_nigra | 0.036 | 0.009 | 6.71E-05 |
| Brain_Spinal_cord_cervical_c-1 | 0.023 | 0.009 | 7.98E-03 |
| Pituitary | 0.012 | 0.010 | 1.05E-01 |
| Cells_EBV-transformed_lymphocytes | 0.005 | 0.005 | 1.84E-01 |
| Muscle_Skeletal | 0.002 | 0.008 | 3.83E-01 |
| Cells_Cultured_fibroblasts | -0.001 | 0.007 | 5.49E-01 |
| Nerve_Tibial | -0.002 | 0.011 | 5.70E-01 |
| Colon_Sigmoid | -0.010 | 0.014 | 7.61E-01 |
| Ovary | -0.011 | 0.011 | 8.46E-01 |
| Testis | -0.007 | 0.006 | 8.82E-01 |
| Whole_Blood | -0.008 | 0.006 | 8.99E-01 |
| Heart_Left_Ventricle | -0.014 | 0.010 | 9.21E-01 |
| Esophagus_Muscularis | -0.019 | 0.014 | 9.21E-01 |
| Skin_Not_Sun_Exposed_Suprapubic | -0.012 | 0.008 | 9.24E-01 |
| Esophagus_Gastroesophageal_Junction | -0.021 | 0.014 | 9.29E-01 |
| Adrenal_Gland | -0.017 | 0.011 | 9.35E-01 |
| Skin_Sun_Exposed_Lower_leg | -0.013 | 0.008 | 9.44E-01 |
| Artery_Tibial | -0.021 | 0.011 | 9.70E-01 |
| Heart_Atrial_Appendage | -0.020 | 0.010 | 9.72E-01 |
| Uterus | -0.023 | 0.012 | 9.74E-01 |
| Esophagus_Mucosa | -0.018 | 0.008 | 9.86E-01 |
| Pancreas | -0.021 | 0.009 | 9.89E-01 |
| Cervix_Endocervix | -0.029 | 0.013 | 9.90E-01 |
| Spleen | -0.020 | 0.008 | 9.95E-01 |
| Cervix_Ectocervix | -0.035 | 0.013 | 9.96E-01 |
| Liver | -0.019 | 0.007 | 9.97E-01 |
| Artery_Aorta | -0.032 | 0.011 | 9.98E-01 |
| Vagina | -0.034 | 0.012 | 9.98E-01 |
| Colon_Transverse | -0.040 | 0.013 | 9.99E-01 |

| Tissue | Beta | SE | P value |
|---|---|---|---|
| Thyroid | -0.034 | 0.011 | 9.99E-01 |
| Kidney_Cortex | -0.031 | 0.010 | 9.99E-01 |
| Prostate | -0.044 | 0.013 | 1.00E+00 |
| Small_Intestine_Terminal_Ileum | -0.036 | 0.010 | 1.00E+00 |
| Fallopian_Tube | -0.045 | 0.013 | 1.00E+00 |
| Kidney_Medulla | -0.035 | 0.010 | 1.00E+00 |
| Stomach | -0.047 | 0.013 | 1.00E+00 |
| Artery_Coronary | -0.048 | 0.013 | 1.00E+00 |
| Minor_Salivary_Gland | -0.039 | 0.010 | 1.00E+00 |
| Bladder | -0.056 | 0.014 | 1.00E+00 |
| Adipose_Subcutaneous | -0.051 | 0.012 | 1.00E+00 |
| Adipose_Visceral_Omentum | -0.060 | 0.012 | 1.00E+00 |
| Breast_Mammary_Tissue | -0.068 | 0.014 | 1.00E+00 |
| Lung | -0.047 | 0.010 | 1.00E+00 |

Note: Total genes = 17234. Padj = 2.9E-6, significant associations are shown in bold.

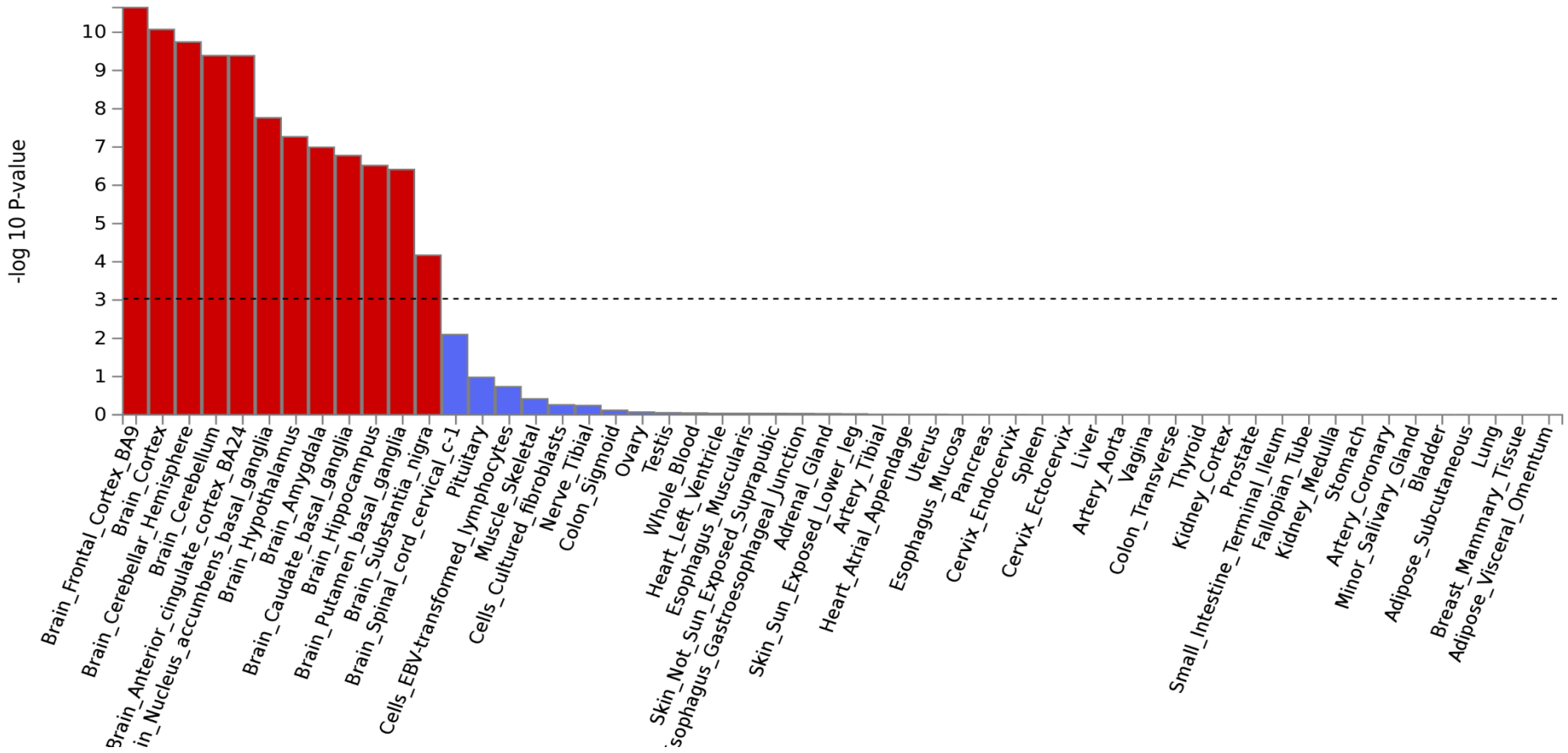



**Figure 4.30: Gene-tissue expression analysis based on GTEx RNA-seq data.**
The y-axis shows the -log 10 transformed P-values of the GWGAS and the x-axis shows the tissue type. Significant associations are shown in red.

#### 4.2.5.2 Gene-cell type expression

Expression analysis at the cellular level was performed using datasets from the Human Prefrontal cortex by age (Zhong et al., 2018), the Human Cortex (Darmanis et al., 2015) and Linnarsson Mouse Brain Atlas (Zeisel et al., 2018). We analysed significant cell types across datasets, independent cell type associations based on within-dataset conditional analyses and pair-wise cross-datasets conditional analyses. Figure 4.31 and Table 4.7 show significant enrichments. For full analysis see Supplementary Tables 13 and 14 (Fitzgerald et al., 2021). These analyses identified four neuronal cell types to be enriched for our associated genes. For human data, these were neurons in the cortex ($P$ = 2.16 x $10^{-6}$), and GW26 GABAergic neurons in the prefrontal cortex ($P$ = 6.59 x $10^{-8}$). For mouse data, these were excitatory glutamatergic neurons in cortical pyramidal layer 5 of the cerebral cortex (TEGLU10; $P$ = 6.98 x $10^{-6}$) and excitatory glutamatergic/nitric oxide neurons in the tegmental reticular nucleus of the pons in the hindbrain (HBGLU8; $P$ = 6.74 x $10^{-7}$). The enrichment in GABAergic neurons is interesting because there is growing evidence to suggest that impairment of the GABAergic system caused by ageing results in an imbalance in the inhibitory/excitatory process involved in the neuronal response to cellular challenges and environmental changes. This results in increased vulnerability to synaptopathy and cognitive decline (Rozycka & Liguz-Lecznar, 2017) .

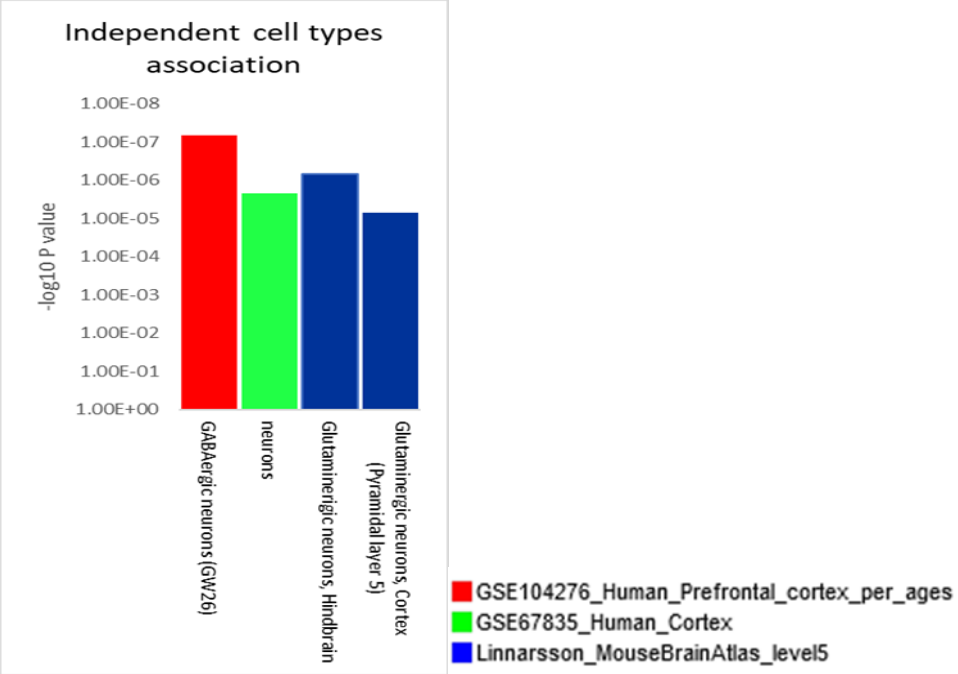



**Figure 4.31: Independent cell type associations based on within-dataset conditional analyses.**
The y-axis shows the -log 10 transformed P-values of the GWGAS and the x-axis shows the cell type.

**Table 4.7: Significant cell type associations across datasets.**

| Dataset | Cell_type | Symbol | Neurotransmitter | SE | P value |
|---|---|---|---|---|---|
| **GSE104276_Human_Prefrontal_cortex_per_ages** | **GW26_GABAergic_neurons** | | | **0.012** | **6.59E-08** |
| **Linnarsson_MouseBrainAtlas_level5** | **Excitatory neurons, hindbrain** | **HBGLU8** | **VGLUT1VGLUT2** | **0.019** | **6.74E-07** |
| **GSE67835_Human_Cortex** | **neurons** | | | **0.009** | **2.16E-06** |
| **Linnarsson_MouseBrainAtlas_level5** | **Excitatory neurons, cerebral cortex** | **TEGLU10** | **VGLUT1** | **0.035** | **6.98E-06** |
| Linnarsson_MouseBrainAtlas_level5 | Excitatory neurons, hindbrain | HBGLU7 | VGLUT1VGLUT2 | 0.029 | 6.78E-07 |
| Linnarsson_MouseBrainAtlas_level5 | Excitatory neurons, hindbrain | HBGLU6 | VGLUT1VGLUT2 | 0.029 | 1.54E-06 |
| Linnarsson_MouseBrainAtlas_level5 | Inhibitory neurons, hindbrain | HBINH5 | GABAGly | 0.036 | 1.80E-06 |
| Linnarsson_MouseBrainAtlas_level5 | Excitatory neurons, hindbrain | HBGLU9 | VGLUT1VGLUT2 | 0.034 | 2.89E-06 |
| Linnarsson_MouseBrainAtlas_level5 | Inhibitory neurons, hindbrain | HBINH9 | GABAGly | 0.057 | 3.33E-06 |
| Linnarsson_MouseBrainAtlas_level5 | Cholinergic neurons, hindbrain | HBCHO2 | VGLUT1VGLUT2ACh | 0.026 | 6.30E-06 |
| Linnarsson_MouseBrainAtlas_level5 | Excitatory neurons, hindbrain | HBGLU1 | VGLUT2VGLUT3 | 0.029 | 8.41E-06 |
| Linnarsson_MouseBrainAtlas_level5 | Excitatory neurons, cerebral cortex | TEGLU4 | VGLUT1VGLUT2 | 0.039 | 1.55E-05 |
| Linnarsson_MouseBrainAtlas_level5 | Cholinergic neurons, hindbrain | HBCHO1 | VGLUT1VGLUT2ACh | 0.025 | 2.46E-05 |
| Linnarsson_MouseBrainAtlas_level5 | Excitatory neurons, hippocampus CA1 | TEGLU21 | VGLUT1 | 0.035 | 3.52E-05 |
| GSE104276_Human_Prefrontal_cortex_per_ages | GW26 oligodendroxyte progenitor cell | | | 0.014 | 3.64E-05 |
| Linnarsson_MouseBrainAtlas_level5 | Inhibitory neurons, hindbrain | HBINH8 | GABAGly | 0.031 | 4.74E-05 |
| Linnarsson_MouseBrainAtlas_level5 | Inhibitory neurons, hindbrain | HBINH2 | GABAGlyACh | 0.026 | 8.92E-05 |
| GSE67835_Human_Cortex | hybrid | | | 0.012 | 1.23E-04 |
| Linnarsson_MouseBrainAtlas_level5 | Inhibitory neurons, hindbrain | HBINH7 | GABAGly | 0.036 | 1.27E-04 |

Notes: Associations of Resilience genes with individual cell types in three scRNA-seq datasets - human pre-frontal cortex, human cortex and the Linarrsson brain atlas. Datasets that survived conditional analysis within datasets are in bold**.** This table shows only significant associations.

#### 4.2.5.3 Pathway enrichment analysis

Gene-set analysis performed on curated gene sets and Gene Ontology (GO) (Ashburner et al., 2000) terms using the full distribution of SNP P-values from the *Resilience* GWAS identified two GO terms to be significantly enriched after adjustment for multiple testing. These were the biological processes“ neuron differentiation” ($P = 9.7 \times 10^{-07}$) and the cellular component “synaptic part” ($P = 2.14 \times 10^{-06}$). Bi-directional conditional analysis using MAGMA 1.08 (de Leeuw, Mooij, Heskes, & Posthuma, 2015) showed that these two annotations were independent of each other. (See Table 4.8 for top 10 enrichment GO terms). One term that was nominally significant for enrichment with a very large beta value was “Wnt signalosome”. There are 12 genes in this gene set, and I examined their P value in the GWGAS discussed in section 4.2.4. A total of 6 of the 12 genes had significant P values and three of these remained significant after correcting for multiple testing ($P_{bon=}$.0042) (See Table 4.9). Two of these genes are mapped in *Resilience* loci by eQTL analysis. The first is *APC* on locus 5A (Figure 4.19). The other is *WNT3* on locus 17 (Figure 4.25). Deficient Wnt signalling is associated with loss of cognitive ability (Palomer, Buechler, & Salinas, 2019). This is discussed further in chapter 6.

**Table 4.8: MAGMA Gene-set Analysis**

| VARIABLE | N Genes | Beta | SE | P value | Conditional P value |
|---|---|---|---|---|---|
| **GO_bp:go_neuron_differentiation** | **1277** | **0.129** | **0.027** | **9.73E-07** | **7.97E-05** |
| **GO_cc:go_synapse_part** | **890** | **0.146** | **0.032** | **2.14E-06** | **3.58E-05** |
| GO_bp:go_synaptic_vesicle_localization | 156 | 0.328 | 0.073 | 3.25E-06 | |
| GO_cc:go_synapse | 1116 | 0.128 | 0.029 | 4.10E-06 | |
| GO_bp:go_neuron_development | 1040 | 0.133 | 0.030 | 4.27E-06 | |
| GO_cc:go_wnt_signalosome | 12 | 1.223 | 0.276 | 4.75E-06 | |
| GO_bp:go_neurogenesis | 1515 | 0.110 | 0.025 | 4.96E-06 | |
| GO_bp:go_synaptic_signaling | 681 | 0.158 | 0.036 | 6.50E-06 | |
| GO_bp:go_signal_release | 442 | 0.198 | 0.046 | 6.81E-06 | |
| GO_bp:go_developmental_cell_growth | 201 | 0.284 | 0.066 | 7.80E-06 | |

Note: Analysis is performed for curated gene sets and GO terms from MsigDB. Bidirectional conditional analysis in magma shows that both sets are independent.

**Table 4.9: GWGAS P values for genes in the Wnt signalosome gene set**

| Gene | P value |
|---|---|
| *WNT2* | 6.16E-01 |
| *ADGRA2* | 2.48E-01 |
| *CTNNB1* | 1.45E-03** |
| *LRP5* | 1.34E-02* |
| *WNT3* | 1.53E-03** |
| *FZD1* | 9.06E-02 |
| *RECK* | 3.18E-02* |
| *LRP6* | 1.27E-01 |
| *DVL1* | 8.51E-01 |
| *APC* | 1.23E-05** |
| *GSK3B* | 2.07E-01 |
| *LRRK2* | 9.87E-03* |

*Nominally significant, **Significant ($P_{bon=}$ 0.0042)

### 4.2.6 Genetic correlations with other traits

I compared our *Resilience* GWAS with recent published GWAS of cognitive phenotypes, psychiatric and neurological disorders, and global brain imaging phenotypes using LDSR analysis (see online methods 2.1.11). A moderate negative correlation of *Resilience* with intelligence (Savage et al., 2018) ($r_g$= -0.26, $P$ = 1.29 x $10^{-17}$) and educational attainment (Lee et al., 2018) ($r_g$= -0.45, $P$ = 1.64 x $10^{-56}$) is as expected given that the resilience phenotype was derived from individuals within the UKB that had lower than average education years. Of the 13-independent genome-wide significant SNPs for *Resilience*, 6 are associated with intelligence at genome-wide significant levels ($P$ < 5 x $10^{-8}$) but the remaining 7 SNPs are not associated with intelligence ($P$ > 0.01). This indicates that some of genetic basis of *Resilience* does not overlap fully with the genetics of intelligence.

When genetic correlation analyses between *Resilience* and psychiatric phenotypes were corrected for multiple testing ($P_{bon}$ < 2.4 x $10^{-3}$), *Resilience* had a small positive correlation with unipolar depression (Nagel et al., 2018) ($r_g$ = 0.17, $P$ = 5.0 x $10^{-10}$), a small negative correlation with schizophrenia (Pardiñas et al., 2018) ($r_g$ = -0.18, $P$ = 1.24 x $10^{-12}$) and bipolar disorder (Stahl et al., 2019) ($r_g$ = -0.17, $P$ = 1.84 x $10^{-7}$), and a nominally significant negative correlation with neuroticism (Luciano et al., 2018) ($r_g$= -0.07, $P$ = 2.02 x $10^{-2}$). Examination of neurological disorders showed *Resilience* had a small nominally significant correlation with amyotrophic lateral sclerosis (ALS) (van Rheenen et al., 2016) ($r_g$ = -0.21, $P$= 1.44 x $10^{-2}$), stroke (Malik et al., 2018) ($r_g$= 0.08, P = 1.89 x $10^{-2}$), and Parkinson's disease (Nalls et

al., 2019) ($r_g$ = -0.08, $P$ = 4.58 x $10^{-02}$), but Alzheimer's disease (AD) (Jansen et al., 2019) was not significant ($r_g$ = 0.04, $P$=0.358), (Table 4.10 and Figure 4.32). Links to public studies are outlined in Table 2.6).

**Table 4.10: LD score regression analysis (LDSC) of cognitive, psychiatric, and neurological traits**

| Symbol | Trait | N | Z score | $^a r_g$ | SE | P value |
|---|---|---|---|---|---|---|
| | *Cognitive* | | | | | |
| **INT** | Intelligence | 269867 | -8.544 | -0.258 | 0.030 | 1.29E-17 |
| **EA** | Educational attainment | 766345 | -15.840 | -0.452 | 0.029 | 1.64E-56 |
| | *Psychiatric and Neurological* | | | | | |
| **ALS** | Amyotrophic lateral sclerosis | 36052 | 2.446 | 0.213 | 0.087 | **1.44E-02** |
| **AD** | Alzheimer's Disease | 455258 | 0.920 | 0.047 | 0.051 | 3.58E-01 |
| **MDD** | Unipolar Depression | 358000 | 6.219 | 0.173 | 0.028 | **4.99E-10** |
| **SCZ** | Schizophrenia | 105318 | -7.101 | -0.181 | 0.025 | **1.24E-12** |
| **BIP** | Bipolar disorder | 35802 | -5.215 | -0.179 | 0.034 | **1.84E-07** |
| **PK** | Parkinson's Disease | 482730 | -1.997 | -0.079 | 0.039 | **4.58E-02** |
| **STK** | Stroke | 520000 | 1.870 | 0.084 | 0.045 | **1.89E-02** |
| **NEU** | Neuroticism | 329000 | 1.511 | 0.044 | 0.029 | 1.31E-01 |

Note: $^a r_g$ =the genetic correction of the trait under examination and Resilience. P bon= 0.6E - 2

The GWAS of 11 brain phenotypes from the UKB (Smith et al., 2020) were examined by LDSC for genetic correlation with the *Resilience* (Table 4.11 and Figure 4.32). The volume of global white and grey matter and cerebral white matter in the left and right hemisphere were examined based on the relationship between brain volumes and cognition (Nave, Jung, Karlsson Linnér, Kable, & Koellinger, 2019). Volume of cerebrospinal fluid was included based of its documented association with brain atrophy (Orellana et al., 2016) and the hippocampus, amygdala and nucleus accumbens were examined as moderators of cognitive function (Lisman et al., 2017) (Floresco, 2015). After adjusting for multiple testing ($P_{bon=}$ 2.4 x $10^{-3}$), the only significant correlations found were for white matter volumes where a small positive correlation was found between *Resilience* and global white matter volume ($r_g$ = 0.14, $P$ = 1.19 x $10^{-3}$), and the volume of cerebral white matter in the left ($r_g$ = 0.148, $P$ = 1.74 x $10^{-03}$) and right hemisphere ($r_g$ = 0.160, $P$ = 7.34 x $10^{-04}$).

**Table 4.11: LDSC of brain imaging phenotypes**

| Volume | [c] ref | [d]$r_g$ | SE | P value |
|---|---|---|---|---|
| **ventricular cerebrospinal fluid [a]** | 3 | -0.11 | 0.05 | 2.27E-02 |
| **grey matter [a]** | 5 | 0.03 | 0.05 | 5.16E-01 |
| **white matter [a]** | **7** | **0.15** | **0.05** | **1.19E-03** |
| **left hippocampus** | 19 | 0.02 | 0.05 | 6.63E-01 |
| **right hippocampus** | 20 | 0.11 | 0.06 | 7.83E-02 |
| **left amygdala** | 21 | 0.09 | 0.06 | 1.14E-01 |
| **right amygdala** | 22 | 0.03 | 0.07 | 6.51E-01 |
| **left accumbens** | 23 | 0.04 | 0.06 | 5.24E-01 |
| **right accumbens** | 24 | 0.06 | 0.06 | 3.58E-01 |
| **cerebral white matter in the left hemisphere [b]** | **190** | **0.15** | **0.05** | **1.74E-03** |
| **cerebral white matter in the right hemisphere[b]** | **207** | **0.16** | **0.05** | **7.34E-04** |

Note: a = normalised for head size, b = generated by subcortical volumetric segmentation. P bon= 3.0E-3. Significant results highlighted in bold. [c]ref = Reference of the imaging variable in the Oxford Brain Imaging Genetics Server - BIG40 atlas. [d]$r_g$ =the genetic correction of the trait under examination and Resilience.

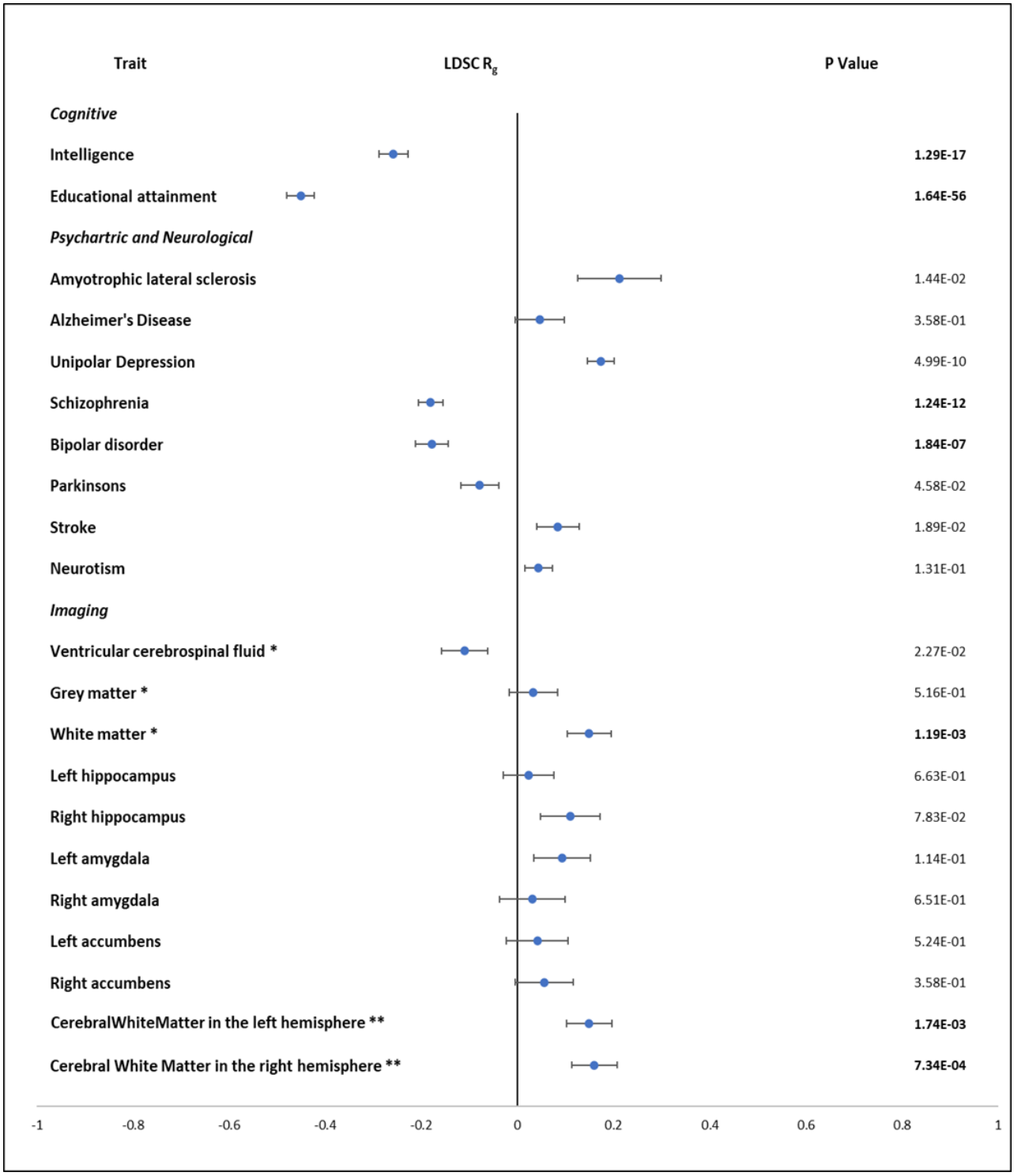



**Figure 4.32: Genetic correlation with cognitive traits, psychiatric and brain disorders, and brain imaging**
Significant P values corrected for multiple testing are in bold. *Normalised for head size, **generated by subcortical volumetric segmentation.

The correlations of cognitive and psychiatric and neurological disorders are largely supported by gene enrichment analysis of the genes associated with *Resilience* here and previous

GWAS of cognitive and psychiatric phenotypes. An analysis of published research from the GWAS catalog (Buniello et al., 2019) showed that the significant SNPs found in this study were previously cited 294 times. A total of 47% of these citations were from studies of cognitive phenotypes (educational attainment, cognitive ability, maths ability and RT) and 5% were from studies of psychiatric disorders (Table 4.12).

**Table 4.12: *Resilience* SNPs overlapping in published research in GWAS catalog.**

| independent sig SNP | Locus | N studies | Trait |
|---|---|---|---|
| **rs12474507** | 2 | 1 | Reaction time |
| **rs35999162** | 3 | 6 | Cognition/intelligence |
| **rs2352974** | 3 | 4 | Cognition/intelligence |
| **rs1317140** | 3 | 3 | Cognition/intelligence |
| **rs2883059** | 3 | 3 | Cognition/intelligence |
| **rs7428430** | 3 | 4 | Cognition/intelligence |
| **rs1046953** | 3 | 1 | Cognition/intelligence |
| **rs2526389** | 3 | 3 | Cognition/intelligence |
| **rs9858059** | 3 | 2 | Cognition/intelligence |
| **rs35999162** | 3 | 3 | Math ability |
| **rs2526389** | 3 | 3 | Math ability |
| **rs57672162** | 4A | 1 | Reaction time |
| **rs6857847** | 4A | 1 | Reaction time |
| **rs2189234** | 4B | 3 | Cognition/intelligence |
| **rs2726485** | 4B | 4 | Cognition/intelligence |
| **rs2189234** | 4B | 1 | Extremely high intelligence |
| **rs2726485** | 4B | 1 | Math ability |
| **rs2189234** | 4B | 1 | Reaction time |
| **rs56335290** | 5A | 1 | Reaction time |
| **rs6870103** | 5B | 4 | Cognition/intelligence |
| **rs6870103** | 5B | 1 | Educational attainment |
| **rs6870103** | 5B | 1 | Math ability |
| **rs6870103** | 5B | 1 | Reaction time |
| **rs7747481** | 6 | 4 | Cognition/intelligence |
| **rs7747481** | 6 | 1 | Educational attainment |
| **rs2417261** | 12A | 3 | Educational attainment |
| **rs6580699** | 12B | 5 | Cognition/intelligence |
| **rs6580699** | 12B | 2 | Educational attainment |
| **rs6580699** | 12B | 2 | Math ability |
| **rs6580699** | 12B | 1 | Reaction time |
| **rs1029388** | 12C | 2 | Reaction time |
| **rs10775404** | 17 | 1 | Reaction time |
| **17:44224272_G_A** | 17 | 1 | Reaction time |
| **rs7225002** | 17 | 1 | Reaction time |
| **rs62074125** | 17 | 1 | Reaction time |
| **rs4810896** | 20 | 3 | Cognition/intelligence |
| **rs2426132** | 20 | 5 | Cognition/intelligence |
| **rs4810896** | 20 | 2 | Educational attainment |
| **rs2426132** | 20 | 1 | Educational attainment |
| **rs4810896** | 20 | 1 | Reaction time |

In addition, when this exercise was repeated for overlapping mapped genes, I found that there was considerable overlap with these phenotypes amongst others. The most significant overlap was where 40 mapped genes in the *Resilience* analysis overlapped with the 99 reported genes for short sleep duration ($P$= 2.03 x $10^{-57}$). In a recent Mendelian randomisation study on sleep duration it was suggested that sleep duration may represent a potential causal pathway for differences in cognitive ability (Henry et al., 2019) and increased sleep in adults over 60 is associated with poorer cognitive function (Low, Wu, & Spira, 2019). There was also a significant overlap with genes associated with extremely high intelligence (Coleman et al., 2019) where 32 *Resilience* mapped genes overlapped with the 81 associated genes reported in that study ($P$ = 1.17 x $10^{-45}$). Many of the overlapping genes for sleep duration and extremely high intelligence were on chromosome 3 (Table 4.13 and Supplementary Table 19 (Fitzgerald et al., 2021)).

**Table 4.13: Genetic overlap of gene enrichment**

| GeneSet | n Genes | n overlap | adjusted P value |
|---|---|---|---|
| **Sleep duration (short sleep)** | 99 | 40 | 3.68E-54 |
| **Extremely high intelligence** | 81 | 32 | 1.06E-42 |
| **Ulcerative colitis** | 366 | 43 | 1.85E-33 |
| **Regular attendance at a religious group** | 78 | 26 | 2.80E-32 |
| **Inflammatory bowel disease** | 640 | 49 | 8.12E-30 |
| **Crohn's disease** | 600 | 47 | 5.06E-29 |
| **Body mass index** | 1358 | 63 | 9.39E-27 |
| **Intelligence** | 313 | 35 | 1.70E-26 |
| **General cognitive ability** | 242 | 31 | 3.51E-25 |
| **Regular attendance at sports club** | 42 | 18 | 1.13E-24 |
| **Reaction time** | 49 | 18 | 3.18E-23 |
| **Mood instability** | 61 | 14 | 1.31E-14 |
| **Blood protein levels** | 1776 | 53 | 6.70E-14 |
| **Cognitive function** | 85 | 14 | 1.59E-12 |
| **Handedness** | 10 | 7 | 3.62E-11 |
| **Alcohol use disorder** | 39 | 10 | 7.85E-11 |
| **Handedness** | 12 | 7 | 2.08E-10 |
| **Response to alcohol consumption** | 13 | 7 | 4.22E-10 |

Note: This table shows the top 18 of 86 associations. For the full list see Supplementary Table 19 (Fitzgerald et al., 2021).

### 4.2.7 Mendelian randomisation

To investigate whether genetic correlations reflected causal effects, I examined the potential credible causality of the relationship between *Resilience* and phenotypes where independent samples were available using Generalised Summary statistics-based Mendelian Randomisation (Zhu et al., 2018) (GSMR) (Table 4.14 and Section 2.1.12). I observed a significant bidirectional causal effect of *Resilience* on schizophrenia (*bxy* = -0.25, $P = 7.02 \times 10^{-9}$) and schizophrenia on *Resilience* (*bxy* = -0.07, $P = 3.80 \times 10^{-7}$) indicating an inter-relationship between the two phenotypes. By contrast, bipolar disorder and ALS did not have significant credible causality relationships with *Resilience*.

GSMR analysis was also performed using white matter volume variables and *Resilience.* To maintain independence between GWAS datasets, I used the *discovery*.*Resilience* GWAS that did not include UKB participants with imaging data. The low level of independent significant SNPs in the discovery GWAS did not allow for analysis of the causal effect of *Resilience* on white matter. A nominally significant causal association of white matter volume with *Resilience* was detected ($b_{xy} = 0.13$, $P = 0.049$) along with causal associations of left and right cerebral hemisphere white matter volume with *Resilience*. The association with the right hemisphere survived multiple test correction (left: $b_{xy} = 0.15$, $P = 0.005$; right: $b_{xy} = 0.17$, $P = 0.002$). There is no evidence of substantial pleiotropy in the GSMR analysis.

**Table 4.14: Generalised Summary data based Mendelian Randomisation.**

| Exposure | Outcome | [a]$b_{xy}$ | se | p | [b]NSNP | N SNPS filtered by [c]HEIDI | Pleiotropic SNPs |
|---|---|---|---|---|---|---|---|
| ***Resilience*** | White matter | nan | nan | nan | nan | 44 | 0 |
| **White matter** | *Resilience* | 0.13 | 0.07 | 4.90E-02 | 9 | | |
| ***Resilience*** | Cerebral white matter (left) | nan | nan | nan | nan | 1 | 0 |
| **Cerebral white matter (left)** | *Resilience* | 0.15 | 0.05 | **5.12E-03** | 12 | | |
| ***Resilience*** | Cerebral white matter (right) | nan | nan | nan | nan | 1 | 0 |
| **Cerebral white matter (right)** | *Resilience* | 0.17 | 0.05 | **1.96E-03** | 12 | | |
| ***Resilience*** | Amyotrophic lateral sclerosis | -0.01 | 0.08 | 9.43E-01 | 13 | 0 | 0 |
| **Amyotrophic lateral sclerosis** | *Resilience* | nan | nan | nan | nan | | |
| ***Resilience*** | Schizophrenia | -0.25 | 0.04 | **7.02E-09** | 13 | 2 | 5 |
| **Schizophrenia** | *Resilience* | -0.07 | 0.01 | **3.80E-07** | 152 | | |
| ***Resilience*** | Bipolar disorder | -0.05 | 0.06 | 4.12E-01 | 12 | 0 | 1 |
| **Bipolar disorder** | *Resilience* | 0.00 | 0.04 | 9.92E-01 | 13 | | |

Note: Bidirectional GSMR results with various phenotypes and Resilience. [a]bxy = the estimate effect coefficient, [b]NSNP = Number of independent significant SNPs after clumping, [c]HEIDI = is an outlier method to remove horizontal pleiotropic SNPs, [d]nan= not applicable as NSNP is less that the threshold value, P bon= 4.2E-3. Significant results highlighted in bold.

## 4.3 Conclusion

I have successfully identified 13 independent genome-wide significant loci resulting in 366 mapped genes and 33 prioritized genes for *Resilience*. Functional analysis showed significant expression of associated genes in all brain tissues, and particularly in the frontal cortex. Significant enrichment of associated genes was also found at the cellular level in both GABAergic and glutamatergic neurons indicating an excitatory/inhibitory control in the prefrontal cortex, and within biological processes related to neuron differentiation and synaptic activity.

Mapping of GWAS results identified genes that have been previously associated with cognitive decline including *STAU1*, *SEMF3A*, *IP6K1*, *MST1*, the *ATNX2/BRAP* locus, *ALDH2 and DDX27*, where a likely functional missense variant is highly associated. Other associated genes involved with synaptic activity and neurogenesis include *BNS*, *DAG1*, *IP6K1* and *TET2*, pointing to potential targets for improvement of cognitive resilience.

On completion of this analysis, I performed ad hoc analysis to satisfy questions arising from this analysis.

- Am I detecting a true Resilience variation, or a variation based solely on high/low RT?
- Almost 30% of the mapped genes and over 50% of prioritized genes can be attributed to a single large locus on chromosome 3. Could the size of this locus have an undue influence on the functional analysis findings?
- Is it possible, given the lack of current genetic data on cognitive change in ageing to replicate these findings in other datasets?

I explore these questions in Chapter 5.

# 5 Ad hoc testing to explore GWAS of *Resilience*.

## 5.1 Introduction:

In Chapter 3, I described how I created a cognitive resilience variable using the UKB and extracted the SNPs that were associated with *Resilience* and in Chapter 4, I outlined functional analysis processes to help identify the biology behind cognitive resilience. In this chapter I further probe three questions that arose from this analysis to strengthen our findings. These are as follows:

- What role does RT play in *Resilience*?
- Does the large locus on chromosome 3 unduly influence the functional analysis?
- Can findings be replicated in a longitudinal sample without using proxy measures for past cognitive performance?

## 5.2 Analysis

### 5.2.1 Examination of the relationship of *Resilience* with RT

Given the strong positive correlation of *Resilience* with RT ($r_g = 0.80$) a possible concern was that I was just identifying genetic associations with RT that are independent of EY. To examine this further I performed a functional analysis on a GWAS of a dichotomised RT phenotype using all suitable participants in the UKB (n=333,664). This GWAS was perfectly correlated ($r_g = 1$, $P = 7.24 \times 10^{-115}$) with a previously published GWAS where RT was studied as a quantitative phenotype (Davies et al., 2018). However, when I compared results for the 13 index SNPs in the *Resilience* GWAS in a similar sized GWAS of RT as a stand-alone phenotype only one of the 13 SNPs was itself genome-wide significant for RT and just three others were associated at P < 1 x 10-4 (Table 5.1). I compared the *Resilience* GWAS to both dichotomised and quantitative RT phenotypes.

**Table 5.1: Comparison of the thirteen index SNPs associated with *Resilience* in GWAS of RT**

| | **Resilience [a]Neff = 111316** | | | | | **RT n=164000 (dichotomised)** | | **RT n=165224 (continuous)** | | **RT n=333,664 (dichotomised)** | | **Davies low RT n=330,069 (continuous)** | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **SNP** | Chr | A1 | A2 | Beta | P | OR | P | Beta | P | OR | P | A1 | A2 | Beta | P |
| **rs1029388** | 12 | C | T | 0.07 | 6.50E-09 | 1.04 | 6.22E-06 | -0.007 | 2.35E-08 | 1.040 | 8.84E-11 | C | T | -0.012 | 3.01E-11 |
| **rs12474507** | 2 | T | C | -0.06 | 4.19E-08 | 0.98 | 7.89E-04 | 0.005 | 1.43E-05 | 0.971 | 8.66E-09 | T | C | 0.011 | 6.44E-11 |
| **rs2189234** | 4 | T | G | 0.07 | 3.14E-12 | 1.02 | 1.72E-03 | -0.005 | 1.74E-06 | 1.030 | 7.63E-09 | G | T | 0.008 | 2.13E-06 |
| **rs2352974** | 3 | T | C | 0.08 | 9.27E-15 | 1.02 | 6.38E-03 | -0.002 | 2.82E-02 | 1.016 | 1.90E-03 | T | C | -0.006 | 8.48E-04 |
| **rs2417261** | 12 | T | G | 0.09 | 1.35E-08 | 1.03 | 1.71E-02 | -0.006 | 1.56E-03 | 1.039 | 2.11E-06 | G | T | 0.006 | 1.09E-03 |
| **rs4810896** | 20 | A | C | 0.07 | 1.94E-10 | 1.03 | 3.67E-04 | -0.004 | 1.54E-04 | 1.026 | 7.81E-07 | C | A | 0.008 | 4.57E-06 |
| **rs56335290** | 5 | A | C | -0.07 | 2.59E-08 | 0.97 | 2.33E-03 | 0.006 | 1.52E-05 | 0.968 | 3.80E-08 | A | C | 0.012 | 1.97E-12 |
| **rs62074125** | 17 | C | A | -0.08 | 8.31E-11 | 0.95 | 7.54E-09 | 0.007 | 1.12E-08 | 0.960 | 8.45E-13 | C | A | 0.012 | 5.28E-12 |
| **rs6580699** | 12 | G | T | -0.07 | 1.29E-10 | 0.97 | 7.71E-05 | 0.005 | 2.76E-05 | 0.971 | 5.55E-09 | G | T | 0.011 | 5.06E-11 |
| **rs6857847** | 4 | A | G | -0.08 | 4.99E-09 | 0.96 | 3.46E-05 | 0.005 | 2.06E-04 | 0.962 | 8.06E-10 | A | G | 0.008 | 3.46E-06 |
| **rs6870103** | 5 | G | T | 0.07 | 4.03E-11 | 1.02 | 1.29E-03 | -0.004 | 3.76E-04 | 1.031 | 7.30E-10 | T | G | 0.010 | 1.88E-09 |
| **rs7747481** | 6 | T | C | 0.06 | 2.29E-08 | 1.01 | 9.45E-02 | -0.001 | 4.42E-01 | 1.011 | 2.75E-02 | T | C | -0.006 | 1.15E-03 |
| **rs9569811** | 13 | A | C | -0.09 | 1.14E-08 | 0.97 | 2.62E-03 | 0.003 | 3.57E-02 | 0.963 | 2.13E-07 | A | C | 0.007 | 1.87E-05 |

Note: I compared Resilience with RT output at a similar sample size and a full sample size using both dichotomised (case/control) variables and continuous variables. I also compared Resilience with a published GWAS (Davies et al., 2018).

5.2.1.1 **Functional analysis** (see method section 2.1.9. Identical settings were used to in this analysis to functional analysis of Resilience)

To examine this further I performed a functional analysis on a dichotomised RT GWAS of all suitable participants in the UKB (n=333,664) and found that while nine of the 13 loci identified in the *Resilience* GWAS overlapped with RT, five loci did not match (Table 5.2).

**Table 5.2: Genetic loci overlap between RT and *Resilience.***

| Resilience locus | SNP | position | P value in *Resilience* | n SNPs | Matched with RT |
|---|---|---|---|---|---|
| **2** | rs12474507 | 59988258 | 4.19E-08 | 2 | No |
| **3** | rs2352974 | 49890613 | 9.27E-15 | 568 | No |
| **4A** | rs6857847 | 89514572 | 4.99E-09 | 76 | Yes |
| **4B** | rs2189234 | 106075498 | 3.14E-12 | 92 | Yes |
| **5A** | rs56335290 | 112036634 | 2.59E-08 | 4 | Yes |
| **5B** | rs6870103 | 139692515 | 4.03E-11 | 104 | Yes |
| **6** | rs7747481 | 98315696 | 2.29E-08 | 135 | No |
| **12A** | rs2417261 | 13414139 | 1.35E-08 | 4 | Yes |
| **12B** | rs6580699 | 49478812 | 1.29E-10 | 44 | Yes |
| **12C** | rs1029388 | 111926901 | 6.50E-09 | 456 | Yes |
| **13** | rs9569811 | 58646190 | 1.14E-08 | 64 | Yes |
| **17** | rs62074125 | 44852612 | 8.31E-11 | 155 | Yes |
| **20** | rs4810896 | 47535298 | 1.94E-10 | 218 | No |

Note: Column 6 shows the nine index SNPs in *Resilience* that were also significant in RT and the five that were not significant.

There was a total of 534 mapped genes for RT and 366 for *Resilience.* Of these, 301 were unique to RT and 133 unique to *Resilience*. There were 223 shared genes (Figure 5.1).

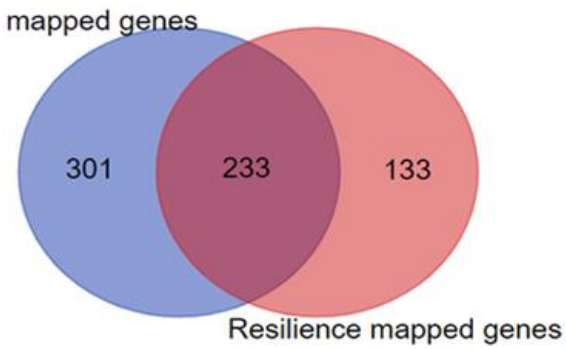



**Figure 5.1: Venn diagram of overlap between mapped genes of *Resilience* and RT**

In addition, examination of prioritized genes using the four strategies in 2.1.9 showed 27 prioritised genes for RT only 11 of whom overlapped with the 33 prioritized genes for *Resilience* (Figure 5.2 (a) and (b) and Table 5.3).

(a) (b)

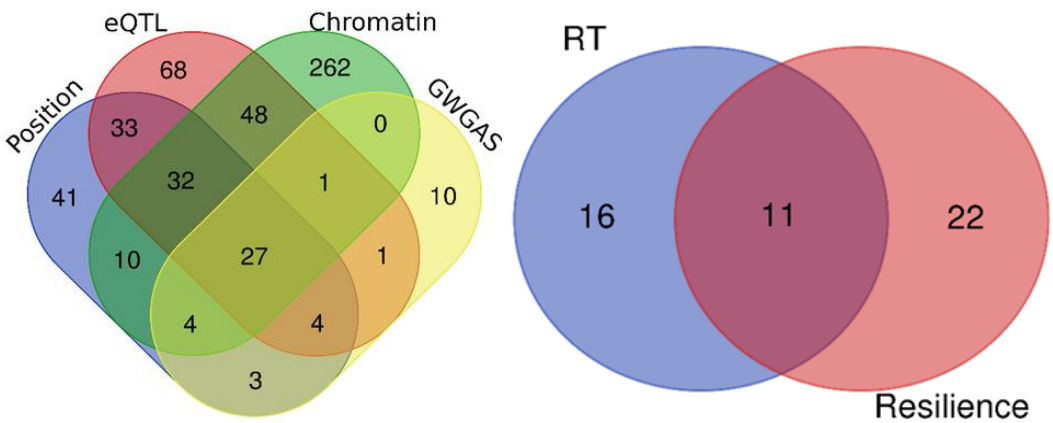



**Figure 5.2: Prioritised genes for RT and their overlap with prioritised genes for *Resilien*ce**
(a) Twenty-seven prioritised genes in RT of which (b) eleven overlap with prioritised genes for *Resilience.*

**Table 5.3: Prioritized genes in RT and *Resilience*.**

| RT | **Resilience** |
|---|---|
| ACAD10 | ACAD10 |
| ALDH2 | ALDH2 |
| ARFGAP2 | AMT |
| ARHGAP1 | ATXN2 |
| ATXN2 | BRAP |
| BRAP | BSN |
| C11orf49 | CAMKV |
| C9orf64 | CSE1L |
| CUX2 | CTD-2330K9.3 |
| CYSTM1 | CYSTM1 |
| DDN | DAG1 |
| DYNC1I2 | DDN |
| F2 | GNAT1 |
| FDXR | GPX1 |
| GRIN2C | IP6K1 |
| LRP4 | MAPKAPK5 |
| MAPKAPK5 | MON1A |
| METAP1D | MST1 |
| PFDN1 | MST1R |
| PRKAG1 | NICN1 |
| SAMD3 | PFDN1 |
| SH2B3 | PRKAG1 |
| SPATS2L | RBM5 |
| TMEM104 | RHOA |
| TMEM116 | RNF123 |
| TMEM200A | SEMA3F |
| ZNF408 | SH2B3 |
| | STAU1 |
| | TCTA |
| | TET2 |
| | TMEM116 |
| | TRAIP |
| | UBA7 |

Note: Shared genes highlighted in red

#### 5.2.1.2 **Pathway enrichment analysis:** (Method section 2.1.10.1)

Pathway enrichment analysis identified shared biological processes between RT and *Resilience* associated with synaptic activity (for example the cellular component "synaptic part" (RT: $P = 2.2 \times 10^{-07}$; *Resilience* $P = 2.14 \times 10^{-06}$) but also showed pathways related to neuronal processes that are only significant in *Resilience* (Table 5.4).

**Table 5.4: Comparison of GO terms between *Resilience* and RT.**

| RT GO term | P | *Resilience* GO term | P |
|---|---|---|---|
| **synapse part** | 2.20E-07 | neuron differentiation | 9.73E-07 |
| **synapse** | 2.08E-06 | **synaptic part** | 2.14E-06 |
| postsynapse | 2.32E-06 | synaptic vesicle localization | 3.25E-06 |
| **wnt signalosome** | 3.78E-06 | **synapse** | 4.10E-06 |
| intrinsic component of post synaptic membrane | 7.13E-06 | neuron development | 4.27E-06 |
| inhibitory extracellular ligand gated ion channel activity | 8.10E-06 | **wnt signalosome** | 4.75E-06 |
| presynapse | 1.38E-05 | neurogenesis | 4.96E-06 |
| neuromuscular process controlling balance | 1.60E-05 | synaptic signalling | 6.50E-06 |
| mesenchymal stem cell differentiation | 1.61E-05 | signal release | 6.81E-06 |
| postsynaptic specialization membrane | 2.20E-05 | development cell growth | 7.80E-06 |

Note: Shared biological processes are in bold

#### 5.2.1.3 **Tissue and cell type enrichment analysis** (Method section 2.1.10.1)

Analysis of tissue and cell type enrichment analysis for RT showed similar findings to those of the *Resilience* GWAS where tissues from global brain regions showed significant enrichment with the strongest enrichments for the frontal cortex, BA9 ($P = 5.05 \times 10^{-12}$), and the cortex ($P = 2.44 \times 10^{-11}$) (Table 5.5). Both GWAS showed enrichment of GW26 GABAergic neurons in the prefrontal cortex (RT: $P = 7.5 \times 10^{-8}$; *Resilience* $P = 6.59 \times 10^{-8}$) and excitatory glutamatergic neurons in cortical pyramidal layer 5 of the cerebral cortex (RT: $P = 2.0 \times 10^{-5}$; *Resilience* $P = 6.98 \times 10^{-6}$). Both phenotypes were significantly enriched for cells in the hindbrain whereas RT shows enrichment of cholinergic neurons in the medulla ($P = 1.1 \times 10^{-5}$), *Resilience* shows enrichment of glutamatergic/nitric oxide neurons in the tegmental reticular nucleus of the pons (Figure 5.3).

**Table 5.5: Comparison of tissue enrichment of *Resilience* and RT**

| | ***Resilience*** | | | **RT** | | |
|---|---|---|---|---|---|---|
| **Tissue** | **Beta** | **SE** | **P value** | **Beta** | **SE** | **P value** |
| **Brain_Frontal_Cortex_BA9** | **0.050** | **0.008** | **2.26E-11** | **0.051** | **0.007** | **5.05E-12** |
| **Brain_Cortex** | **0.050** | **0.008** | **8.48E-11** | **0.051** | **0.008** | **2.44E-11** |
| **Brain_Cerebellar_Hemisphere** | **0.042** | **0.007** | **1.81E-10** | **0.040** | **0.007** | **8.65E-10** |
| **Brain_Cerebellum** | **0.042** | **0.007** | **4.07E-10** | **0.040** | **0.007** | **3.27E-09** |
| **Brain_Anterior_cingulate_cortex_BA24** | **0.048** | **0.008** | **4.11E-10** | **0.047** | **0.008** | **9.64E-10** |
| **Brain_Nucleus_accumbens_basal_ganglia** | **0.046** | **0.008** | **1.73E-08** | **0.042** | **0.008** | **1.69E-07** |
| **Brain_Hypothalamus** | **0.047** | **0.009** | **5.38E-08** | **0.043** | **0.009** | **5.00E-07** |
| **Brain_Amygdala** | **0.044** | **0.009** | **1.01E-07** | **0.040** | **0.009** | **1.14E-06** |
| **Brain_Caudate_basal_ganglia** | **0.044** | **0.009** | **1.67E-07** | **0.041** | **0.009** | **8.27E-07** |
| **Brain_Hippocampus** | **0.043** | **0.009** | **3.04E-07** | **0.043** | **0.009** | **3.41E-07** |
| **Brain_Putamen_basal_ganglia** | **0.043** | **0.009** | **3.88E-07** | **0.041** | **0.009** | **7.84E-07** |
| Brain_Substantia_nigra | 0.036 | 0.009 | 6.71E-05 | 0.036 | 0.009 | 7.04E-05 |
| Brain_Spinal_cord_cervical_c-1 | 0.023 | 0.009 | 7.98E-03 | 0.024 | 0.009 | 5.25E-03 |
| Pituitary | 0.012 | 0.010 | 1.05E-01 | 0.010 | 0.010 | 1.57E-01 |
| Cells_EBV-transformed_lymphocytes | 0.005 | 0.005 | 1.84E-01 | 0.003 | 0.005 | 2.72E-01 |
| Muscle_Skeletal | 0.002 | 0.008 | 3.83E-01 | 0.006 | 0.008 | 2.03E-01 |
| Cells_Cultured_fibroblasts | -0.001 | 0.007 | 5.49E-01 | -0.006 | 0.007 | 7.99E-01 |
| Nerve_Tibial | -0.002 | 0.011 | 5.70E-01 | -0.015 | 0.011 | 9.16E-01 |
| Colon_Sigmoid | -0.010 | 0.014 | 7.61E-01 | -0.015 | 0.014 | 8.67E-01 |

| | ***Resilience*** | | | **RT** | | |
|---|---|---|---|---|---|---|
| **Tissue** | **Beta** | **SE** | **P value** | **Beta** | **SE** | **P value** |
| Ovary | -0.011 | 0.011 | 8.46E-01 | -0.023 | 0.011 | 9.85E-01 |
| Testis | -0.007 | 0.006 | 8.82E-01 | -0.006 | 0.006 | 8.38E-01 |
| Whole_Blood | -0.008 | 0.006 | 8.99E-01 | -0.002 | 0.006 | 6.54E-01 |
| Heart_Left_Ventricle | -0.014 | 0.010 | 9.21E-01 | -0.010 | 0.010 | 8.54E-01 |
| Esophagus_Muscularis | -0.019 | 0.014 | 9.21E-01 | -0.032 | 0.014 | 9.91E-01 |
| Skin_Not_Sun_Exposed_Suprapubic | -0.012 | 0.008 | 9.24E-01 | -0.005 | 0.008 | 7.32E-01 |
| Esophagus_Gastroesophageal_Junction | -0.021 | 0.014 | 9.29E-01 | -0.035 | 0.014 | 9.94E-01 |
| Adrenal_Gland | -0.017 | 0.011 | 9.35E-01 | -0.022 | 0.011 | 9.79E-01 |
| Skin_Sun_Exposed_Lower_leg | -0.013 | 0.008 | 9.44E-01 | -0.006 | 0.008 | 7.53E-01 |
| Artery_Tibial | -0.021 | 0.011 | 9.70E-01 | -0.013 | 0.011 | 8.85E-01 |
| Heart_Atrial_Appendage | -0.020 | 0.010 | 9.72E-01 | -0.022 | 0.010 | 9.83E-01 |
| Uterus | -0.023 | 0.012 | 9.74E-01 | -0.027 | 0.012 | 9.88E-01 |
| Esophagus_Mucosa | -0.018 | 0.008 | 9.86E-01 | -0.023 | 0.008 | 9.98E-01 |
| Pancreas | -0.021 | 0.009 | 9.89E-01 | -0.021 | 0.009 | 9.88E-01 |
| Cervix_Endocervix | -0.029 | 0.013 | 9.90E-01 | -0.037 | 0.013 | 9.98E-01 |
| Spleen | -0.020 | 0.008 | 9.95E-01 | -0.013 | 0.008 | 9.53E-01 |
| Cervix_Ectocervix | -0.035 | 0.013 | 9.96E-01 | -0.046 | 0.013 | 1.00E+00 |
| Liver | -0.019 | 0.007 | 9.97E-01 | -0.017 | 0.007 | 9.93E-01 |
| Artery_Aorta | -0.032 | 0.011 | 9.98E-01 | -0.023 | 0.011 | 9.80E-01 |
| Vagina | -0.034 | 0.012 | 9.98E-01 | -0.041 | 0.012 | 1.00E+00 |
| Colon_Transverse | -0.040 | 0.013 | 9.99E-01 | -0.037 | 0.013 | 9.98E-01 |
| Thyroid | -0.034 | 0.011 | 9.99E-01 | -0.027 | 0.010 | 9.96E-01 |
| Kidney_Cortex | -0.031 | 0.010 | 9.99E-01 | -0.028 | 0.010 | 9.98E-01 |
| Prostate | -0.044 | 0.013 | 1.00E+00 | -0.038 | 0.013 | 9.98E-01 |
| Small_Intestine_Terminal_Ileum | -0.036 | 0.010 | 1.00E+00 | -0.032 | 0.010 | 9.99E-01 |
| Fallopian_Tube | -0.045 | 0.013 | 1.00E+00 | -0.056 | 0.013 | 1.00E+00 |
| Kidney_Medulla | -0.035 | 0.010 | 1.00E+00 | -0.028 | 0.010 | 9.97E-01 |

| | ***Resilience*** | | | **RT** | | |
|---|---|---|---|---|---|---|
| **Tissue** | **Beta** | **SE** | **P value** | **Beta** | **SE** | **P value** |
| Stomach | -0.047 | 0.013 | 1.00E+00 | -0.043 | 0.013 | 1.00E+00 |
| Artery_Coronary | -0.048 | 0.013 | 1.00E+00 | -0.047 | 0.013 | 1.00E+00 |
| Minor_Salivary_Gland | -0.039 | 0.010 | 1.00E+00 | -0.032 | 0.010 | 9.99E-01 |
| Bladder | -0.056 | 0.014 | 1.00E+00 | -0.057 | 0.014 | 1.00E+00 |
| Adipose_Subcutaneous | -0.051 | 0.012 | 1.00E+00 | -0.046 | 0.012 | 1.00E+00 |
| Adipose_Visceral_Omentum | -0.060 | 0.012 | 1.00E+00 | -0.063 | 0.012 | 1.00E+00 |
| Breast_Mammary_Tissue | -0.068 | 0.014 | 1.00E+00 | -0.067 | 0.014 | 1.00E+00 |
| Lung | -0.047 | 0.010 | 1.00E+00 | -0.040 | 0.010 | 1.00E+00 |

Total genes = 17234.Padj = 2.9E-6, significant associations are shown in bold

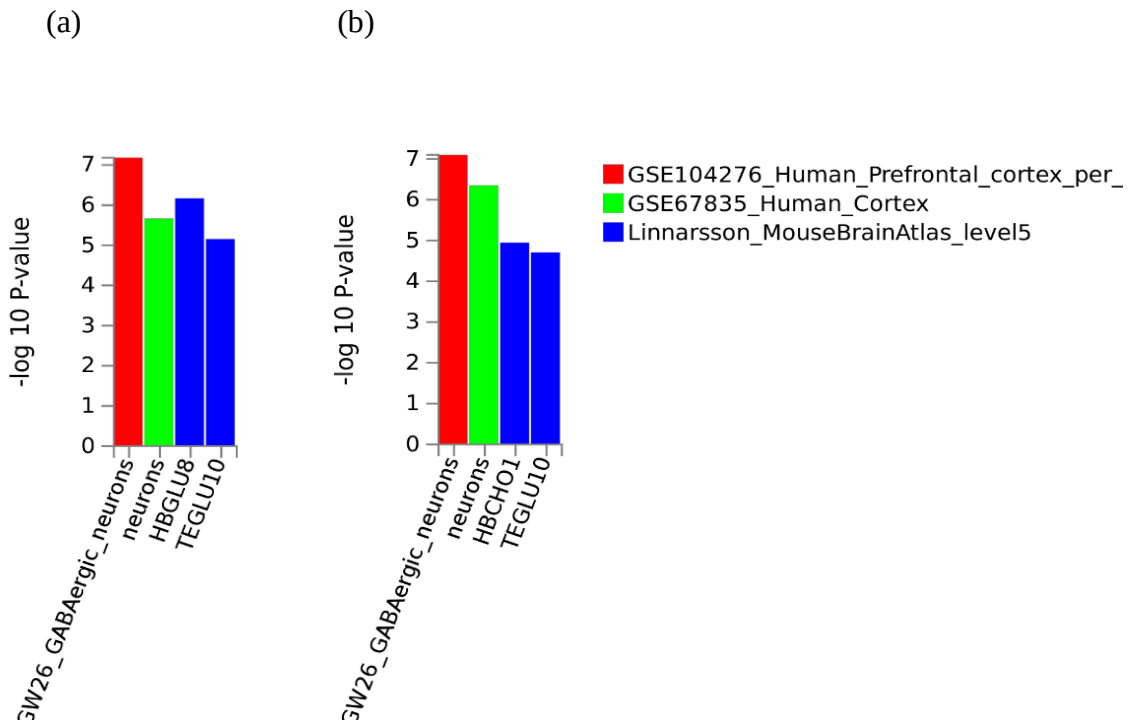



**Figure 5.3: Comparison of single cell enrichment**
Note: (a) *Resilience* (b) RT. The y-axis shows the -log 10 transformed P-values of the GWGAS and the x-axis shows the cell type.

#### 5.2.1.4 LD regression analysis

LD score regression analysis showed a positive correlation for RT with cognitive phenotypes (Intelligence: $r_g$= 0.18, $P$ = 1.87 x $10^{-11}$; Educational attainment: $r_g$= 0.12, $P$ = 9.54 x $10^{-7}$) which contrasts with that of *Resilience.* The RT GWAS was performed on all suitable individuals in the UKB, whereas the *Resilience* GWAS was generated from individuals with lower-than-average EA. Examining correlations with psychiatric and neurological phenotypes shows a significant small negative correlation with schizophrenia ($r_g$= -0.19, $P$ = 1.89 x $10^{-14}$) and bipolar disorder ($r_g$= -0.10, $P$ = 2.53 x $10^{-3}$). *Resilience* had a similar correlation with schizophrenia and a larger negative correlation with bipolar disorder. Unipolar depression was not significantly correlated with RT (rg = -0.03, P = 2.59x 10-1) whereas it showed positive correlation with *Resilience* (for a comparison of this analysis see Table 5.6).

**Table 5.6: Comparison of LDSR between *Resilience* and RT for cognitive, psychiatric, and neurological phenotypes**

| | | **Resilience** | | | | | **Reaction time** | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Trait** | N | Symbol | Z score | [a]$r_g$ | SE | P value | Z score | [a]$r_g$ | SE | P value |
| **Cognitive** | | | | | | | | | | |
| Intelligence | 269867 | INT | -8.54 | -0.26 | 0.03 | **1.29E-17** | 6.72 | 0.18 | 0.03 | **1.87E-11** |
| Educational attainment | 766345 | EA | -15.84 | -0.45 | 0.03 | **1.64E-56** | 4.90 | 0.12 | 0.02 | **9.54E-07** |
| **Psychiatric and Neurological** | | | | | | | | | | |
| Amyotrophic lateral sclerosis | 36052 | ALS | 2.45 | 0.21 | 0.09 | 1.44E-02 | 1.36 | 0.12 | 0.09 | 1.73E-01 |
| Alzheimer's Disease | 455258 | AD | 0.92 | 0.05 | 0.05 | 3.58E-01 | -2.16 | -0.11 | 0.05 | 3.05E-02 |
| Unipolar Depression | 358000 | MDD | 6.22 | 0.17 | 0.03 | **4.99E-10** | -1.13 | -0.03 | 0.03 | 2.59E-01 |
| Schizophrenia | 105318 | SCZ | -7.10 | -0.18 | 0.03 | **1.24E-12** | -7.66 | -0.19 | 0.03 | **1.89E-14** |
| Bipolar disorder | 35802 | BIP | -5.21 | -0.18 | 0.03 | **1.84E-07** | -3.02 | -0.10 | 0.03 | **2.53E-03** |
| Parkinsons | 482730 | PK | -2.00 | -0.08 | 0.04 | 4.58E-02 | -0.33 | -0.01 | 0.04 | 7.45E-01 |
| Stroke | 520000 | STK | 1.87 | 0.08 | 0.04 | 1.89E-02 | -1.75 | -0.08 | 0.00 | 8.01E-02 |
| Neurotism | 329000 | NEU | 1.51 | 0.04 | 0.03 | 1.31E-01 | -2.21 | -0.06 | 0.03 | 2.74E-02 |

Note: Significant results highlighted in bold

Examining correlations with brain imaging data showed similar results to *Resilience* where significant correlations were found for white matter volumes - a small positive correlation was found between RT and global white matter volume ($r_g$ = 0.13, $P$ = 2.8 x $10^{-3}$), and the volume of cerebral white matter in the left ($r_g$ = 0.15, $P$ = 8.29 x $10^{-04}$) and right hemisphere ($r_g$ = 0.16, $P$ = 2.18 x $10^{-04}$). For a comparison of this analysis see Table 5.7.

**Table 5.7: Comparison of LDSR of brain volumes between *Resilience* and RT**

| | *Resilience* | | | RT | | |
|---|---|---|---|---|---|---|
| | $^{d}r_g$ | SE | P value | $^{d}r_g$ | SE | P value |
| Volume of ventricular cerebrospinal fluid [a] | -0.11 | 0.05 | 2.27E-02 | -0.08 | 0.05 | 7.45E-02 |
| Volume of grey matter [a] | 0.03 | 0.05 | 5.16E-01 | 0.03 | 0.05 | 6.17E-01 |
| **Volume of white matter [a]** | **0.15** | **0.05** | **1.19E-03** | **0.13** | **0.04** | **2.80E-03** |
| Volume of left hippocampus | 0.02 | 0.05 | 6.63E-01 | 0.03 | 0.06 | 5.67E-01 |
| Volume of right hippocampus | 0.11 | 0.06 | 7.83E-02 | 0.11 | 0.06 | 7.03E-02 |
| Volume of left amygdala | 0.09 | 0.06 | 1.14E-01 | 0.12 | 0.06 | 4.96E-02 |
| Volume of right amygdala | 0.03 | 0.07 | 6.51E-01 | 0.02 | 0.07 | 7.34E-01 |
| Volume of left accumbens | 0.04 | 0.06 | 5.24E-01 | 0.08 | 0.06 | 2.45E-01 |
| Volume of right accumbens | 0.06 | 0.06 | 3.58E-01 | 0.05 | 0.06 | 4.07E-01 |
| **Volume of CerebralWhiteMatter in the left hemisphere [b]** | **0.15** | **0.05** | **1.74E-03** | **0.15** | **0.04** | **8.29E-04** |
| **Volume of CerebralWhiteMatter in the right hemisphere[b]** | **0.16** | **0.05** | **7.34E-04** | **0.16** | **0.04** | **2.18E-04** |

Note: a = normalised for head size, b = generated by subcortical volumetric segmentation. P bon= 3.0E-3. Significant results highlighted in bold. $^{d}r_g$ =the genetic correction of the trait under examination and Resilience or RT.

#### 5.2.1.5 Mendelian randomisation

Using GSMR, I observed a significant unidirectional causal effect of RT on schizophrenia ($bxy$ = -0.05, $P$ = 2.38 x $10^{-253}$). Due to the removal of a high level of pleiotropic SNPs there were insufficient SNPs to calculate the effect in the opposite direction. I had observed a significant bidirectional causal effect of *Resilience* on schizophrenia ($bxy$ = -0.25, $P$ = 7.02 x $10^{-9}$) and schizophrenia on *Resilience* ($bxy$ = -0.07, $P$ = 3.80 x $10^{-7}$). See Table 5.8.

**Table 5.8: Generalised Summary-data-based Mendelian Randomisation on RT**

| Exposure | Outcome | [a]$b_{xy}$ | se | p | [b]NSNP | [c]HEIDI filtered SNPs |
|---|---|---|---|---|---|---|
| RT | Schizophrenia | -0.045 | 0.001 | **2.38E-253** | 61 | 131 |
| Schizophrenia | RT | [d]nan | nan | nan | 2 | 156 |
| RT | Bipolar disorder | 0.004 | 0.003 | 2.07E-01 | 23 | 15 |
| Bipolar disorder | RT | nan | nan | nan | 7 | 6 |

Note: Bidirectional GSMR results with various phenotypes and Resilience. [a]bxy = the estimate effect coefficient, [b]NSNP = Number of independent significant SNPs after clumping, [c]HEIDI = is an outlier method to remove horizontal pleiotropic SNPs, [d]nan= not applicable as NSNP is less that the threshold value, P bon= 4.2E-3. Significant results highlighted in bold.

Examining the hypergeometric analysis of reported genes from GWAS-catalog shows an overlap of schizophrenia genes with *Resilience* (22 in total), and RT (45 in total), however, only 15 of these genes were shared between the two GWAS.

In conclusion:

While there were similarities in the GWAS and *Resilience* and RT for example tissue and cell type enrichment analysis, I showed a number of differences in associated genetic loci, mapped genes and pathway enrichment analysis.

The genetic correlation between Resilience and RT is strong because this is an RT-based resilience phenotype; however, the top associated SNPs for RT were not being detected here. Instead, I detected SNPs associated with faster than average RT in individuals that previously showed below average EY, i.e., the resilience phenotype in this study. This phenotype represents those individuals in the UKB who preserved their capability to process information and respond over a 40 years' time span and reflects the genetic difference between these individuals and those who showed diminishing processing speed.

### 5.2.2 Effect of large locus on chromosome 3

Almost 30% of the mapped genes and over 50% of prioritized genes can be attributed to a single large locus on chromosome 3: 49385350 - 50250837 (see Figure 4.17 and Figure 4.29). I was concerned that his locus might have an inflated influence on the functional analysis of *Resilience*.

To investigate this, I extracted all SNPs in this locus from the GWAS of *Resilience* and reprocessed the GWAS through FUMA. Figure 5.4 shows a Manhattan plot Manhattan plot of Resilience minus locus 3. Figure 5.5 compares the summary data of Resilience and Resilience minus chromosome 3 showing the removal of the SNPs and mapped genes.

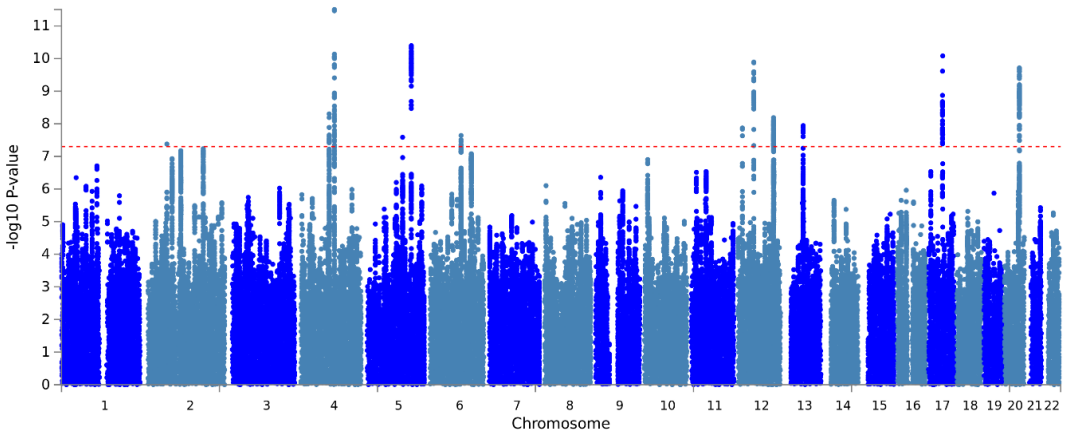
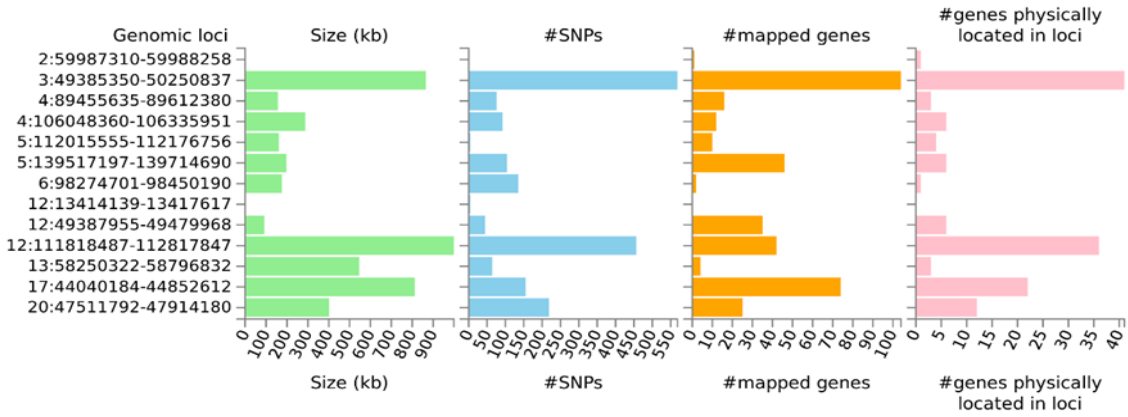



**Figure 5.4: Manhattan plot of Resilience minus Locus 3 (3: 49385350-50250837)**
The Y axis shows the -log10 transformed P-values of each SNP from the GWAS. The x axis shows the base pair position along the chromosomes. The dotted read line shows the Bonferroni corrected P-value (P<5.0E-8).

(a)




(b)

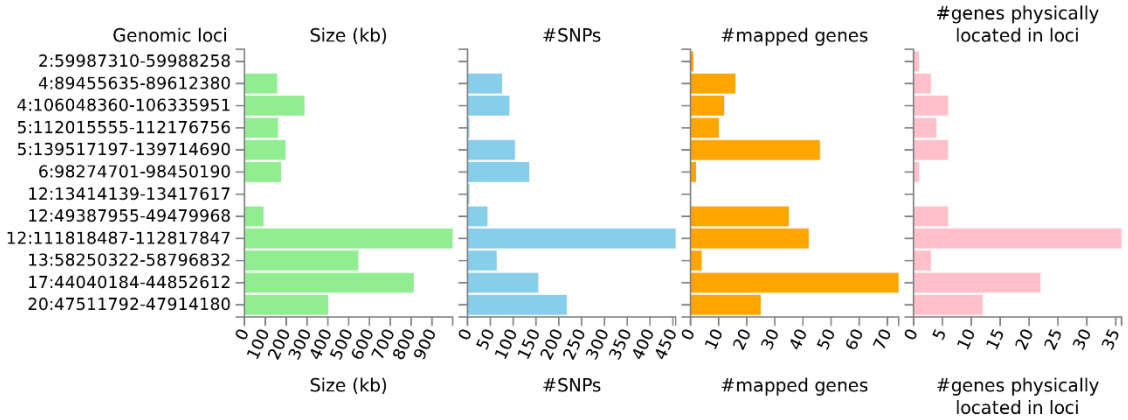



**Figure 5.5: Summary per genomic risk loci**
Figure (a) is the summary for Resilience and Figure (b) is the summary of the same GWAS minus the region 3:49385350-50250837.

I repeated our functional analysis of GWAS results minus locus 3 and found that all enrichments in tissues (Table 5.9) cell types (Figure 5.6) and GO terms (Table 5.10) previously identified, remained significant.

**Table 5.9: Comparison of tissue enrichment of *Resilience* with and without locus 3**

| FULL_NAME | *Resilience* minus locus 3 P | *Resilience* With locus 3P |
|---|---|---|
| Brain_Frontal_Cortex_BA9 | 1.28E-11 | 2.26E-11 |
| Brain_Cortex | 6.00E-11 | 8.48E-11 |
| Brain_Cerebellar_Hemisphere | 1.47E-10 | 1.81E-10 |
| Brain_Anterior_cingulate_cortex_BA24 | 2.29E-10 | 4.11E-10 |
| Brain_Cerebellum | 3.63E-10 | 4.07E-10 |
| Brain_Nucleus_accumbens_basal_ganglia | 1.10E-08 | 1.73E-08 |
| Brain_Hypothalamus | 3.73E-08 | 5.38E-08 |
| Brain_Amygdala | 6.26E-08 | 1.01E-07 |
| Brain_Caudate_basal_ganglia | 1.07E-07 | 1.67E-07 |
| Brain_Hippocampus | 1.96E-07 | 3.04E-07 |
| Brain_Putamen_basal_ganglia | 2.27E-07 | 3.88E-07 |
| Brain_Substantia_nigra | 4.70E-05 | 6.71E-05 |
| Brain_Spinal_cord_cervical_c-1 | 6.60E-03 | 7.98E-03 |

(a) (b)

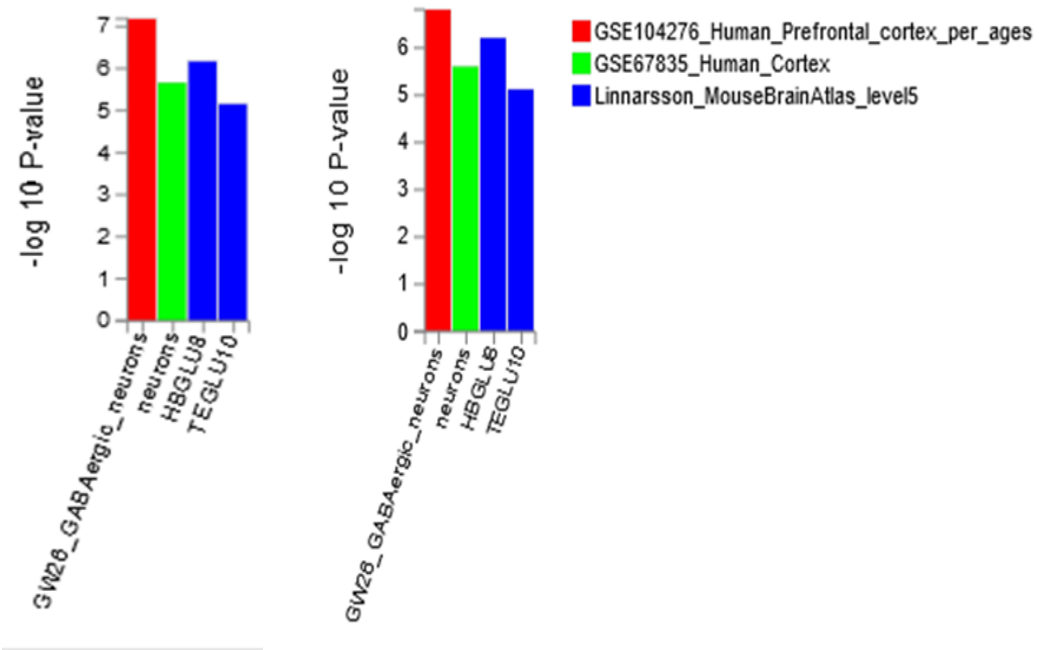



**Figure 5.6: Comparison of single cell type enrichment**
Note: (a) Resilience (b) Resilience minus locus 3. The y-axis shows the -log 10 transformed P-values of the GWGAS and the x-axis shows the cell type.

**Table 5.10: Comparison of GO terms**

| **Go Term** | ***Resilience* with Locus 3** | | ***Resilience* Minus Locus 3** | |
|---|---|---|---|---|
| | nSNPs | P | nSNPs | P |
| **GO_bp:go_neuron_differentiation** | 1277 | 9.73E-07 | 1273 | 9.58E-07 |
| **GO_cc:go_synapse_part** | 890 | 2.14E-06 | 887 | 1.13E-06 |
| **GO_bp:go_synaptic_vesicle_localization** | 156 | 3.25E-06 | 156 | 3.01E-06 |
| **GO_cc:go_synapse** | 1116 | 4.10E-06 | 1111 | 3.37E-06 |
| **GO_bp:go_neuron_development** | 1040 | 4.27E-06 | 1036 | 4.23E-06 |
| **GO_cc:go_wnt_signalosome** | 12 | 4.75E-06 | 12 | 4.48E-06 |
| **GO_bp:go_neurogenesis** | 1515 | 4.96E-06 | 1511 | 4.91E-06 |
| **GO_bp:go_synaptic_signaling** | 681 | 6.50E-06 | 679 | 4.57E-06 |
| **GO_bp:go_signal_release** | 442 | 6.81E-06 | 442 | 5.82E-06 |
| **GO_bp:go_developmental_cell_growth** | 201 | 7.80E-06 | 200 | 1.36E-05 |

In addition, the genetic correlation with schizophrenia ($r_g$ -0.1796, $P = 1.29 \times 10^{-12}$) and white matter ($r_g = 0.158$, $P = 9.5E-04$) remained significant.

This analysis demonstrates that the large locus on chromosome 3 does not have an exaggerated role in the functional analysis of *Resilience.*

### 5.2.3 Exploration of HRS to verify findings

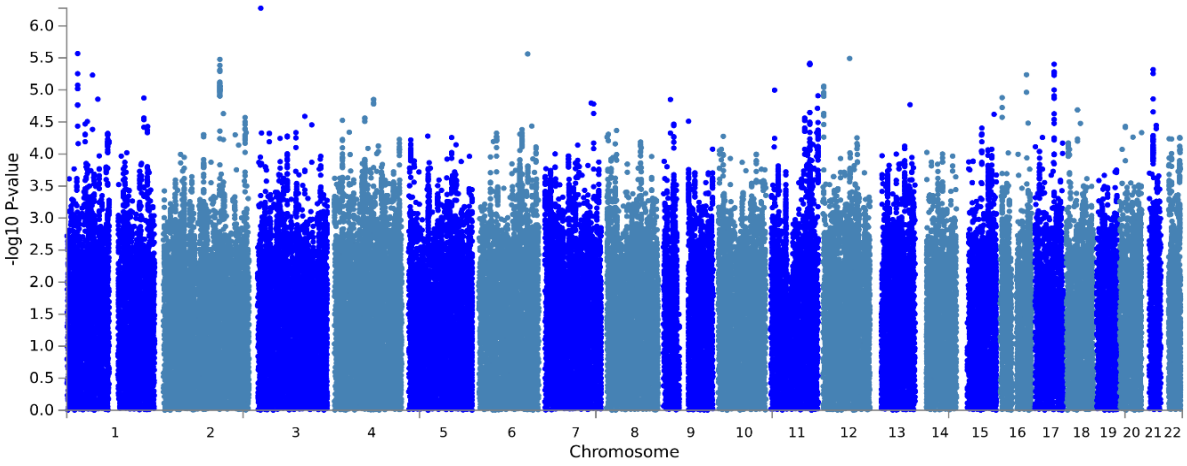
In Chapter 3, I described how I extracted a GWAS of Resilience using SEM. I showed this method was robust by first using a discovery sample and repeating the analysis in an independent sample where I replicated the results. However, given the fact that I employed proxy measures for past cognitive performance, I examined ways in which I could verify our results in longitudinal datasets with direct cognitive measures.

In Chapter 1, I discussed the very limited longitudinal data available to study cognitive resilience and the lack of significant findings generated using these datasets. The largest available dataset with cognitive measures over a long period of time is the HRS. I decided to use this dataset to specifically examine the influence of the 13 loci identified in the UKB.

#### 5.2.3.1 Generation of GWAS of cogntive change in HRS

In Section 2.2.2, I describe the HRS dataset and the cogntive measures available. Unfortunately, there is no processing speed measure, so I examined both the total cognitive variable (COGTOT) and the immediate work recall variable (Recall). The Manhattan plots for the two cogntive measures have no significant genetic loci, however, there are more SNPs with $P< 1 \times 10^{-5}$ in the larger Recall GWAS. There were 291 candidate GWAS SNPs in Recall and 205 in COGTOT. The correlation between the two phenotypes was high ($r_g = 0.8$, $P <.001$), however, the genetic correlation of the two GWAS was lower and was not significant ($r_g = -0.46$, $P = 7.10 \times 10^{-01}$). Examining the Q-Q plots for both GWAS shows a slightly better profile for Recall than COGTOT (Figure 5.8) indicating that the increase in sample size improves the outcome.

COGTOT (n=5,345)




Recall (n=9526)

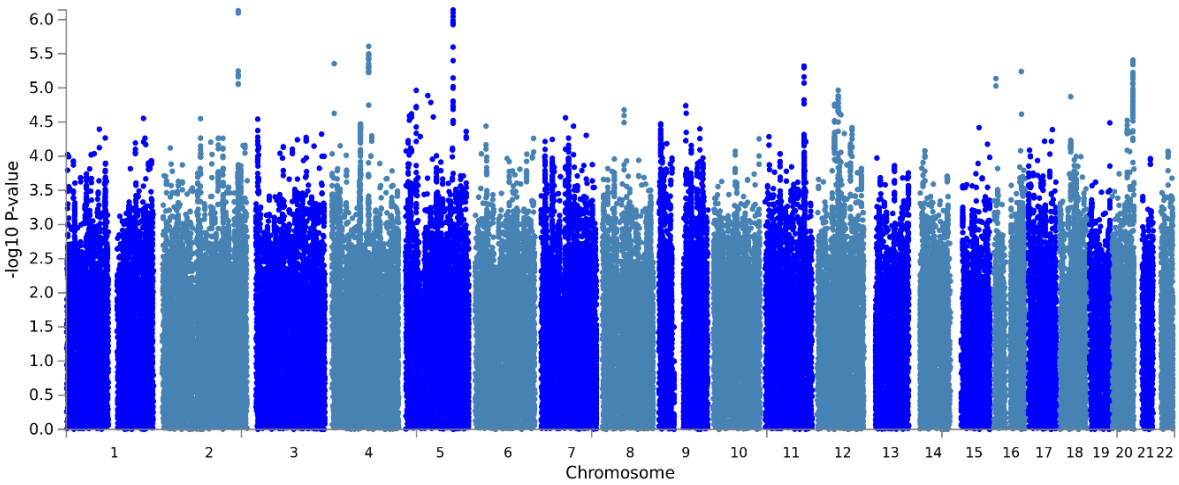



**Figure 5.7; Manhattan plots of COGTOT and Recall**
The Y axis shows the -log10 transformed P-values of each SNP from the GWAS. The x axis shows the base pair position along the chromosomes. The dotted read line shows the Bonferroni corrected P-value (P<5.0E-8).

(a) (b)

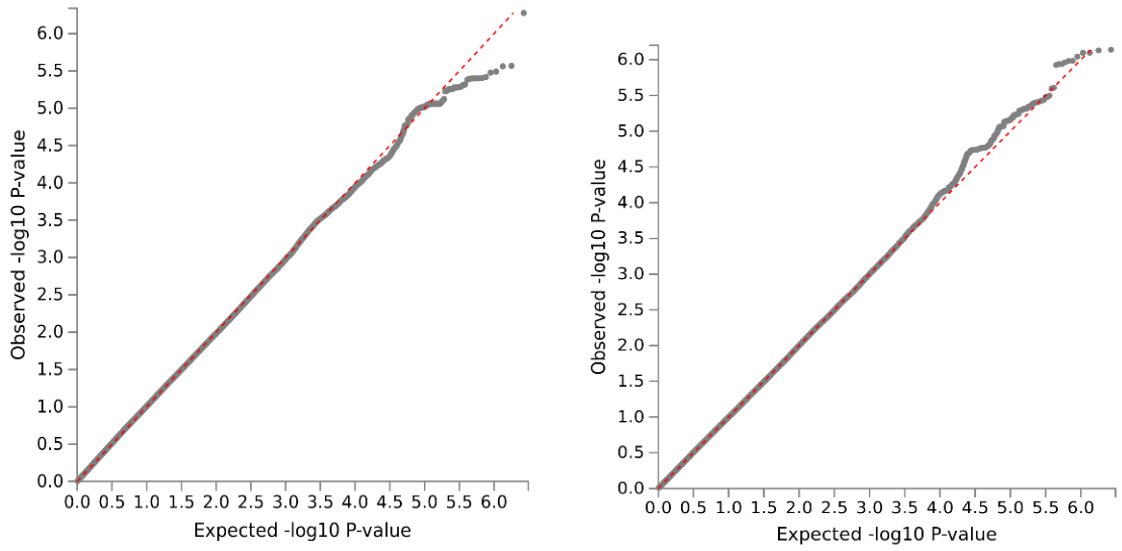



**Figure 5.8: Q-Q plots of (a) COGTOT and (b) Recall**
The x-axis shows the expected distribution of p-values from the GWAS across all SNPs, and the y-axis shows the observed p-values. The genomic inflation value (Lambda) calculated by LDSC is 1.0 for both plots.

### 5.2.3.2 Examining the relationship of the HRS GWAS with *Resilience*

#### 5.2.3.2.1 Comparison of P and beta values

Firstly, I looked at the 13 genetic loci identified in Resilience and compared their beta and P value to that of both COGTOT and Recall (Table 5.11). Due to the discreet origins of the two datasets (UKB and HRS) there were several SNPs that were not present in both datasets. When this occurred, I used a proxy SNPs that was in high LD with the lead SNP using Haploreg v4.1 (Table 2.1) employing a LD limit of $r^2 > 0.4$. Only one of the thirteen SNPs was nominally significant in Recall; rs2883059, chromosome 3, $P = 2.53 \times 10^{-2}$ which was not significant when it was corrected for multiple testing (Bonferroni P value threshold = 0.05/13 = 0.004). None of the SNPs were significantly associated in COGTOT. Examining the Betas values, 7 of 11 Recall beta values were in the opposite direction to the *Resilience* values and 4 were in the same direction (2 SNPs could not be matched). Looking at the COGTOT Beta values, 8 of 12 were in the opposite direction while 4 were in the same direction.

**Table 5.11: Comparison of Beta and P values between genetic loci associated with *Resilience* and the two HRS GWAS**

| **UKB** | | | | | | | **HRS** | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Resilience n=111,316 | | | | | | Recall n=9526 | | Tot cog n=5345 | |
| **SNP** | | Chr | A1 | A2 | Beta | P value | A1 | Beta | | | |
| rs12474507 | 59988258 | 2 | T | C | -0.060 | 4.19E-08 | T | 0.005 | 2.82E-01 | -0.003 | 6.07E-01 |
| rs2352974 | 49890613 | 3 | T | C | 0.080 | 9.27E-15 | | | | | |
| ***rs2883059*** | ***49902160*** | ***3*** | ***C*** | ***T*** | ***0.065*** | ***2.39E-10*** | ***C*** | ***-0.009*** | ***2.53E-02*** | *0.006* | *2.89E-01* |
| rs2189234 | 106075498 | 4 | T | G | 0.070 | 3.14E-12 | T | -0.006 | 1.47E-01 | -0.003 | 5.92E-01 |
| rs6857847 | 89514572 | 4 | A | G | -0.080 | 4.99E-09 | A | -0.005 | 3.17E-01 | -0.008 | 2.92E-01 |
| rs56335290 | 112036634 | 5 | A | C | -0.070 | 2.59E-08 | | | | | |
| *rs459552* | *112176756* | *5* | *T* | *A* | *-0.064* | *1.09E-07* | *T* | *0.001* | *7.66E-01* | *-0.013* | *6.03E-02* |
| rs6870103 | 139692515 | 5 | G | T | 0.070 | 4.03E-11 | G | 0.000 | 9.39E-01 | -0.007 | 2.10E-01 |
| rs7747481 | *98315696* | 6 | T | C | 0.060 | 2.29E-08 | T | -0.003 | 4.89E-01 | -0.008 | 1.75E-01 |
| rs1029388 | 111926901 | 12 | C | T | 0.070 | 6.50E-09 | | | | | |
| *rs7134740* | *111927739* | *12* | *A* | *G* | *0.070* | *2.56E-08* | *A* | *0.008* | *1.27E-01* | *0.001* | *8.58E-01* |
| rs2417261 | 13414139 | 12 | T | G | 0.090 | 1.35E-08 | T | 0.000 | 9.65E-01 | -0.005 | 5.83E-01 |
| rs6580699 | 49478812 | 12 | G | T | -0.070 | 1.29E-10 | G | -0.001 | 9.05E-01 | -0.009 | 1.11E-01 |
| rs9569811 | 58646190 | 13 | A | C | -0.090 | 1.14E-08 | A | 0.009 | 1.51E-01 | 0.008 | 3.26E-01 |
| rs62074125 | 44852612 | 17 | C | A | -0.080 | 8.31E-11 | | | | | |
| *rs7225002* | *44189067* | *17* | *G* | *A* | *0.065* | *2.44E-10* | *G* | *0.005* | *2.20E-01* | *0.000* | *9.36E-01* |
| rs4810896 | 47535298 | 20 | A | C | 0.070 | 1.94E-10 | | | | | |
| *rs2426132* | *47723127* | *20* | *C* | *G* | *-0.062* | *1.54E-09* | *C* | *0.001* | *7.25E-01* | *0.002* | *7.18E-01* |

Note: SNPs in italics are proxy SNPs for the SNP immediately above. One SNP rs2883059 is nominally significant in HRS recall and is in bold.

To examine this relationship further I looked at a larger sample in the *Resilience* GWAS where I reset the level of cut off to recognise lead SNPs from a P value of 5 x $10^{-8}$ to 1 x $10^{-5}$ in FUMA. Using the 109 lead SNPs in this analysis (Table 5.14) at the end of this chapter, I compared the P and beta values of *Resilience* with Recall. There were no proxys with LD $r^2$ values > 0.4 for 12 SNPs. For the remaining 97 SNPs, I found that there were 8 nominally significant SNPs in Recall listed in Table 5.12. None of these SNPs were significant after correcting for multiple testing (P < 0.0005).

**Table 5.12: Nominally significant SNPs in Recall**

| | | | **Resilience** | | **Recall** | |
|---|---|---|---|---|---|---|
| **SNP** | Chr | Position | Beta | P | Beta | P |
| **rs10125362** | 1 | 183546061 | 0.052 | 1.61E-06 | -0.010 | 1.88E-02 |
| **rs1029388** | 1 | 45209661 | -0.063 | 9.01E-06 | 0.016 | 8.29E-03 |
| **rs143875052** | 4 | 106150555 | -0.064 | 3.22E-05 | 0.013 | 3.75E-02 |
| **rs184738312** | 4 | 161505867 | -0.052 | 1.00E-04 | -0.011 | 5.00E-02 |
| **rs3754028** | 1 | 112282873 | 0.056 | 7.30E-06 | -0.010 | 4.76E-02 |
| **rs2883059** | 3 | 49902160 | 0.065 | 2.39E-10 | -0.009 | 2.53E-02 |
| **rs10775404** | 1 | 77949129 | 0.051 | 8.23E-07 | 0.010 | 2.17E-02 |
| **rs35428069** | 9 | 78579131 | 0.062 | 9.65E-04 | -0.025 | 1.30E-03 |

Analysis of beta values showed 49 SNPs that had opposite direction of effect and 48 that had the same direction of effect. This result shows that the direction of effect of the beta values is random and there is no significant relationship between *Resilience* and Recall (sign test *P* = 0.5).

#### 5.2.3.2.2 Genetic correlation between *Resilience* and HRS GWAS

I examined the genetic correlation using LDSC (2.1.11) between *Resilience* and the two HRS GWAS, COGTOT and Recall and sample size appears to play a role in the sensitivity of this analysis Table 5.13. There was a small non-significant negative correlation between COGTOT and *Resilience* ($r_g$ = -0.05, *P* = 7.3 x $10^{-1}$), however, the genetic correlation between Resilience and Recall was moderate and significant ($r_g$ = -0.64, *P* = 1.5 x $10^{-2}$). The reason this result is negative is that Recall is examining cognitive change whereas *Resilience* is looking at resistance to change.

I further examined the genetic correlations between Recall and RT and intelligence and compared the results to other cognitive genetic correlation previously found for *Resilience.* The genetic correlation of Recall with RT ($r_g$ = -0.57, *P* = 4.0E x $10^{-3}$) low RT ($r_g$ = 0.33, *P* = 1.7E-01) and intelligence ($r_g$ = 0.60, *P* = 4.7E x $10^{-3}$) are all in opposite directions to those

found for *Resilience* which is consistent with Recall and Resilience measuring opposite effects.

**Table 5.13: Genetic correlations *Resilience* and Recall and COGTOT**

| | | **Recall** | | **COGTOT** | | **Resilience** | |
|---|---|---|---|---|---|---|---|
| | Source | R | P | R | P | | |
| **Resilience** | In house | -0.64 | 1.50E-02 | -0.05 | 7.30E-01 | | |
| **RT** | In house | -0.57 | 4.00E-03 | | | 0.80 | 2.4E-62 |
| **Low RT** | Davies et al 2018* | 0.33 | 1.7E-01 | | | -0.84 | 4.0E-102 |
| **Intelligence** | Savage et al 2018* | 0.60 | 4.70E-03 | | | -0.25 | 1.29E-17 |
| **COGTOT** | In house | -0.46 | 7.10E-01 | | | | |

Note: *See Table 2.6 for further information

#### 5.2.3.2.3 Conclusion of Comparison of Resilience in the HRS

The HRS dataset is limited by sample size, but it has good longitudinal data on cognitive performance over a long period of time. Here I performed an analysis on two phenotypes total cognition (COGTOT, n=5,345) and word recall (Recall, n=9,526) using linear mixed modelling to define cognitive change over time. This analysis shows that sample size is important in that stronger associations with the cognitive change phenotypes are found in the larger GWAS shown by the number of significantly associated SNPs and an improved Q-Q plot for Recall. While phenotypic correlation is high the genetic correlation diminishes with sample size.

Comparison between *Resilience* with the two HRS phenotypes shows a significant negative correlation with Recall ($r_g$ = -0.64, $P$ = 1.5 x $10^{-2}$) and a small negative correlation with COGTOT which was not significant. However, only nominally significant overlap was seen at the SNPs level and the agreement between the direction of effect of the beta values was random. A larger dataset is needed to truly confirm these finding. This is discussed further in Chapter 6.

## 5.3 Conclusion to Chapter 5

In this chapter I set out to examine questions that arose during functional analysis of *Resilience* in Cation, or a variation based solely on high/low RT?

By comparing GWAS analysis of RT with *Resilience*, I showed that while there is a strong correlation between the two phenotypes (as *Resilience* is an RT-based phenotype), functional analysis shows that there are unique attributes. The top associated SNPs for RT were not being detected in *Resilience*. This phenotype represents those individuals in the UKB who preserved their capability to process information and respond over a 40 years' time span and reflects the genetic difference between these individuals and those who showed diminishing processing speed.

The second question I asked was whether the large genetic locus on chromosome 3 was overrepresented in the findings of the functional analysis. Processing a GWAS without this locus through FUMA did indeed show a substantial decrease in candidate SNPs and mapped genes, however, the main findings of the functional analysis remained unchanged.

The third question was whether I could replicate our findings in an external dataset. Using the HRS dataset, I showed a significant correlation between cognitive change and *Resilience*. Given the size of this dataset I found only one of the 13 genetic loci identified in *Resilience* to be nominally significant in the cognitive change phenotype, Recall. Larger datasets are needed to further confirm these findings.

The findings of the three questions in this chapter will be discussed further in Chapter 6.

**Table 5.14: Comparison of Beta and P values between genetic loci associated with *Resilience* and Recall using 112 lead SNPs**

| **Resilience** | | | | | **Recall** | | **Proxy Recall** | | | | **Proxy Resilience** | | **Sign** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **SNP** | **Chr** | **Position** | **Beta** | **P** | **Beta** | **P** | **SNP** | **Position** | **Beta** | **P** | **Beta** | **P** | |
| **rs10125362** | 9 | 128490151 | 0.048 | 3.41E-06 | | | rs7023828 | 128498594 | 0.003 | 4.94E-01 | 0.046 | 8.42E-06 | yes |
| **rs1029388** | 12 | 111926901 | 0.073 | 6.50E-09 | | | rs7134740 | 111927739 | 0.008 | 1.27E-01 | 0.070 | 2.56E-08 | yes |
| **rs10432338** | 2 | 147806785 | 0.061 | 3.24E-06 | | | No proxy | | | | | | |
| **rs10521241** | 16 | 51357286 | -0.057 | 2.49E-06 | 0.001 | 8.44E-01 | | | | | | | no |
| **rs10747478** | 1 | 96901455 | -0.050 | 1.18E-06 | 0.006 | 1.62E-01 | | | | | | | no |
| **rs10769191** | 11 | 46157060 | -0.061 | 4.97E-07 | | | rs78059714 | 46157568 | 0.009 | 1.10E-01 | -0.043 | 1.97E-04 | no |
| **rs10775404** | 17 | 44167366 | 0.080 | 1.35E-09 | | | rs117913167 | 44212362 | 0.001 | 2.75E-01 | 0.075 | 2.11E-08 | yes |
| **rs10873201** | 14 | 67966599 | -0.047 | 4.22E-06 | | | rs10133618 | 67981403 | 0.001 | 7.97E-01 | -0.043 | 1.98E-05 | no |
| **rs10906892** | 10 | 15389924 | 0.047 | 7.75E-06 | 0.002 | 6.08E-01 | | | | | | | yes |
| **rs10962543** | 9 | 16698822 | 0.097 | 4.40E-07 | -0.006 | 4.39E-01 | | | | | | | no |
| **rs11024435** | 11 | 17946606 | -0.046 | 7.86E-06 | | | rs379388 | 17944246 | 0.001 | 9.00E-01 | -0.045 | 1.13E-05 | no |
| **rs11148465** | 13 | 59930803 | -0.048 | 9.99E-06 | -0.004 | 3.25E-01 | | | | | | | yes |
| **rs112261906** | 4 | 3050001 | 0.128 | 1.49E-06 | | | No proxy | | | | | | |
| **rs11647572** | 16 | 12211417 | 0.056 | 2.22E-06 | | | rs12596703 | 12213850 | -0.003 | 5.87E-01 | 0.053 | 6.85E-06 | no |
| **rs116802139** | 3 | 180079080 | 0.096 | 2.95E-06 | 0.006 | 4.33E-01 | | | | | | | yes |
| **rs117623407** | 19 | 32204489 | -0.069 | 1.34E-06 | | | No proxy | | | | | | |
| **rs11839321** | 13 | 59974439 | 0.058 | 3.97E-06 | 0.006 | 2.84E-01 | | | | | | | yes |
| **rs12474507** | 2 | 59988258 | -0.057 | 4.19E-08 | 0.005 | 2.82E-01 | | | | | | | no |
| **rs12550380** | 8 | 87186583 | -0.051 | 7.91E-06 | -0.006 | 1.76E-01 | | | | | | | yes |
| **rs12631730** | 3 | 47129903 | 0.063 | 3.20E-06 | -0.011 | 6.14E-02 | | | | | | | no |
| **rs12753665** | **1** | **183546061** | **0.052** | **1.61E-06** | **-0.010** | **1.88E-02** | | | | | | | **no** |
| **rs12807111** | 11 | 46924174 | -0.061 | 2.91E-07 | | | rs10769225 | 46982481 | 0.002 | 6.60E-01 | -0.060 | 6.98E-07 | no |
| **rs13175613** | 5 | 92026071 | -0.053 | 6.30E-07 | | | rs7730365 | 92026606 | 0.002 | 7.35E-01 | -0.050 | 3.76E-06 | no |

| **Resilience** | | | | | **Recall** | | **Proxy Recall** | | | | **Proxy Resilience** | | **Sign** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **SNP** | **Chr** | **Position** | **Beta** | **P** | **Beta** | **P** | **SNP** | **Position** | **Beta** | **P** | **Beta** | **P** | |
| **rs13254345** | 8 | 133019596 | -0.051 | 9.12E-06 | -0.001 | 7.76E-01 | | | | | | | yes |
| **rs1351848** | 11 | 45073858 | -0.048 | 2.13E-06 | | | No proxy | | | | | | |
| **rs142186733** | 1 | 145548981 | 0.107 | 9.58E-06 | | | No proxy | | | | | | |
| **rs142342829** | 5 | 93189775 | -0.106 | 8.03E-06 | 0.019 | 7.68E-02 | | | | | | | no |
| **rs150306** | 15 | 89951979 | 0.046 | 5.93E-06 | 0.008 | 6.64E-02 | | | | | | | yes |
| **rs150817429** | 11 | 56003922 | -0.205 | 7.42E-06 | | | rs148158820 | 56174037 | -0.002 | 9.08E-01 | -0.173 | 7.52E-05 | yes |
| **rs16832210** | **1** | **45209661** | **-0.063** | **9.01E-06** | **0.016** | **8.29E-03** | | | | | | | **no** |
| **rs17035310** | **4** | **106064754** | **-0.070** | **4.95E-06** | | | **rs143875052** | **106150555** | **0.013** | **3.75E-02** | **-0.064** | **3.22E-05** | **no** |
| **rs17039735** | **4** | **161476054** | **-0.065** | **1.04E-06** | | | **rs184738312** | **161505867** | **-0.011** | **5.00E-02** | **-0.052** | **1.00E-04** | **yes** |
| **rs17225749** | 3 | 50131140 | 0.072 | 1.52E-06 | | | rs17304079 | 50085153 | -0.003 | 4.90E-01 | 0.011 | 1.05E-06 | no |
| **rs17603622** | 2 | 181442373 | 0.056 | 7.56E-06 | 0.0004 | 9.43E-01 | | | | | | | no |
| **rs17663027** | 2 | 147992667 | -0.117 | 7.29E-06 | | | No proxy | | | | | | |
| **rs1768809** | 1 | 46502836 | -0.051 | 4.49E-07 | -0.001 | 8.29E-01 | | | | | | | yes |
| **rs1815823** | 18 | 35101644 | 0.056 | 4.85E-06 | | | rs4426420 | 35104386 | -0.001 | 8.28E-01 | 0.056 | 5.88E-05 | no |
| **rs1830640** | 11 | 29778003 | -0.073 | 9.42E-06 | -0.003 | 6.08E-01 | | | | | | | yes |
| **rs189033023** | 3 | 51033710 | 0.181 | 2.42E-06 | | | rs79073578 | 50999864 | 0.005 | 7.19E-01 | 0.117 | 4.28E-05 | yes |
| **rs1970811** | 10 | 126696496 | 0.045 | 9.81E-06 | -0.005 | 2.52E-01 | | | | | | | no |
| **rs197374** | **1** | **112289983** | **0.054** | **1.95E-07** | | | **rs3754028** | **112282873** | **-0.010** | **4.76E-02** | **0.056** | **7.30E-06** | **no** |
| **rs1985721** | 4 | 40146842 | -0.046 | 7.84E-06 | | | No proxy | | | | | | |
| **rs201281950** | 12 | 49672500 | -0.058 | 1.27E-06 | | | No proxy | | | | | | |
| **rs2027130** | 1 | 43906896 | -0.048 | 2.90E-06 | | | rs1334973 | 43921384 | -0.002 | 6.69E-01 | -0.047 | 4.83E-06 | yes |
| **rs2102065** | 16 | 5080523 | 0.048 | 5.52E-06 | | | rs17706797 | 5063336 | 0.003 | 5.76E-01 | 0.043 | 2.07E-04 | yes |
| **rs2177500** | 2 | 76413305 | -0.069 | 1.17E-07 | | | No proxy | | | | | | |
| **rs2189234** | 4 | 106075498 | 0.073 | 3.14E-12 | -0.006 | 1.47E-01 | | | | | | | no |
| **rs2240287** | 11 | 61505583 | -0.066 | 8.22E-06 | 0.0001 | 9.89E-01 | | | | | | | no |
| **rs2257063** | 2 | 104092954 | -0.055 | 6.72E-08 | | | rs6543211 | 104113747 | -0.004 | 3.28E-01 | -0.053 | 1.51E-07 | yes |
| **rs2326942** | 6 | 130603039 | 0.061 | 8.30E-08 | | | No proxy | | | | | | |

| Resilience | | | | | Recall | | Proxy Recall | | | | Proxy Resilience | | Sign |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **SNP** | **Chr** | **Position** | **Beta** | **P** | **Beta** | **P** | **SNP** | **Position** | **Beta** | **P** | **Beta** | **P** | |
| **rs2352974** | **3** | **49890613** | **0.080** | **9.27E-15** | | | **rs2883059** | **49902160** | **-0.009** | **2.53E-02** | **0.065** | **2.39E-10** | no |
| **rs2417261** | 12 | 13414139 | 0.094 | 1.35E-08 | 0.0003 | 9.65E-01 | | | | | | | yes |
| **rs2417262** | 12 | 13407132 | -0.064 | 1.65E-07 | -0.003 | 5.17E-01 | | | | | | | yes |
| **rs254776** | 5 | 88006893 | -0.049 | 8.17E-06 | 0.004 | 3.95E-01 | | | | | | | no |
| **rs2702042** | 2 | 24546542 | 0.047 | 9.88E-06 | -0.007 | 1.25E-01 | | | | | | | no |
| **rs28370374** | 7 | 70676309 | -0.046 | 6.84E-06 | -0.007 | 1.04E-01 | | | | | | | yes |
| **rs2863007** | 3 | 177203524 | 0.047 | 4.02E-06 | | | rs10428125 | 177194714 | -0.005 | 2.20E-01 | 0.103 | 3.25E-05 | no |
| **rs2944826** | 7 | 71792250 | -0.071 | 6.53E-06 | | | No proxy | | | | | | |
| **rs3095843** | 5 | 172501829 | 0.050 | 8.05E-07 | -0.005 | 2.11E-01 | | | | | | | no |
| **rs3180887** | 9 | 86553589 | -0.058 | 1.13E-06 | | | rs10115699 | 86554217 | 0.000 | 9.75E-01 | -0.057 | 1.31E-06 | yes |
| **rs33957528** | 16 | 28307940 | -0.050 | 1.09E-06 | | | rs9939450 | 28301487 | 0.001 | 8.13E-01 | -0.044 | 1.36E-06 | no |
| **rs34372833** | 2 | 175200014 | 0.049 | 3.21E-06 | | | rs35548534 | 175200903 | -0.005 | 2.36E-01 | 0.011 | 2.52E-04 | no |
| **rs35871487** | 2 | 161351758 | 0.056 | 7.11E-06 | | | rs10204185 | 161336706 | 0.005 | 2.97E-01 | 0.011 | 4.70E-04 | yes |
| **rs3819161** | 21 | 45092653 | -0.048 | 3.74E-06 | | | rs8129601 | 45119104 | -0.002 | 6.68E-01 | -0.046 | 6.82E-06 | yes |
| **rs4351477** | 9 | 76953396 | 0.067 | 2.31E-06 | | | rs72744343 | 76816724 | 0.009 | 1.01E-01 | 0.013 | 3.38E-05 | yes |
| **rs446994** | 17 | 7116853 | 0.053 | 2.91E-07 | 0.003 | 4.78E-01 | | | | | | | yes |
| **rs4810896** | 20 | 47535298 | 0.067 | 1.94E-10 | | | rs2426132 | 47723127 | 0.001 | 7.25E-01 | -0.062 | 1.54E-09 | no |
| **rs4875419** | 8 | 4825809 | 0.051 | 7.89E-07 | -0.002 | 7.06E-01 | | | | | | | no |
| **rs56335290** | 5 | 112036634 | -0.068 | 2.59E-08 | | | rs459552 | 112176756 | 0.001 | 7.66E-01 | -0.064 | 1.09E-07 | no |
| **rs588470** | 11 | 16380754 | 0.053 | 3.08E-07 | | | rs10219384 | 16328969 | 0.001 | 8.82E-01 | 0.052 | 5.90E-07 | yes |
| **rs60630276** | 2 | 222798862 | 0.056 | 7.15E-06 | 0.009 | 8.58E-02 | | | | | | | yes |
| **rs6073984** | 20 | 44630653 | 0.072 | 1.18E-06 | 0.005 | 3.77E-01 | | | | | | | yes |
| **rs6130929** | 20 | 44405013 | 0.050 | 3.09E-06 | | | rs1711203 | 44385389 | 0.006 | 1.74E-01 | 0.046 | 1.49E-05 | yes |
| **rs62074125** | 17 | 44852612 | -0.077 | 8.31E-11 | | | No proxy | | | | | | |
| **rs62308744** | 4 | 89782561 | -0.058 | 1.95E-06 | 0.005 | 3.15E-01 | | | | | | | no |
| **rs640999** | 11 | 63869425 | -0.046 | 7.46E-06 | -0.002 | 5.77E-01 | | | | | | | yes |
| **rs6561817** | 13 | 55638594 | 0.049 | 2.71E-06 | | | rs1925060 | 55616633 | -0.007 | 1.06E-01 | -0.031 | 2.60E-03 | yes |

| Resilience | | | | | Recall | | Proxy Recall | | | | Proxy Resilience | | Sign |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **SNP** | **Chr** | **Position** | **Beta** | **P** | **Beta** | **P** | **SNP** | **Position** | **Beta** | **P** | **Beta** | **P** | |
| **rs6580699** | 12 | 49478812 | -0.066 | 1.29E-10 | -0.001 | 9.05E-01 | | | | | | | yes |
| **rs6792702** | 3 | 149565982 | 0.057 | 9.43E-07 | 0.004 | 4.13E-01 | | | | | | | yes |
| **rs6850086** | 4 | 92782848 | 0.046 | 7.20E-06 | | | rs1385867 | 92750927 | 0.003 | 5.05E-01 | 0.044 | 1.47E-05 | yes |
| **rs6857847** | 4 | 89514572 | -0.076 | 4.99E-09 | -0.005 | 3.17E-01 | | | | | | | yes |
| **rs6870103** | 5 | 139692515 | 0.068 | 4.03E-11 | 0.000 | 9.39E-01 | | | | | | | no |
| **rs690371** | 17 | 72859078 | -0.053 | 6.71E-06 | 0.002 | 6.47E-01 | | | | | | | no |
| **rs6918725** | 6 | 126990392 | 0.045 | 7.90E-06 | | | rs853982 | 127046121 | 0.004 | 3.63E-01 | -0.040 | 7.23E-05 | yes |
| **rs71581523** | 5 | 139363779 | 0.133 | 5.28E-06 | -0.004 | 7.05E-01 | | | | | | | no |
| **rs7163832** | 15 | 79906344 | 0.047 | 8.25E-06 | 0.004 | 3.27E-01 | | | | | | | yes |
| **rs72750102** | 5 | 53874523 | -0.086 | 4.14E-06 | 0.0004 | 9.63E-01 | | | | | | | no |
| **rs72833334** | 10 | 63663562 | -0.047 | 7.77E-06 | | | rs58936043 | 63674934 | -0.006 | 1.79E-01 | -0.046 | 1.49E-05 | yes |
| **rs72889923** | 3 | 65137914 | 0.093 | 7.90E-06 | 0.008 | 3.28E-01 | | | | | | | yes |
| **rs7528932** | **1** | **77949129** | **0.051** | **8.23E-07** | **0.010** | **2.17E-02** | | | | | | | **yes** |
| **rs754298** | 3 | 50548895 | 0.071 | 1.81E-06 | | | rs73082948 | 50544417 | -0.008 | 1.77E-01 | 0.060 | 2.04E-05 | no |
| **rs75479062** | 11 | 45313976 | -0.085 | 9.49E-06 | 0.002 | 7.88E-01 | | | | | | | no |
| **rs75719921** | 2 | 174991914 | -0.070 | 5.83E-08 | 0.004 | 4.35E-01 | | | | | | | no |
| **rs75835456** | 6 | 68999978 | -0.057 | 1.44E-06 | | | rs7451317 | 69021802 | 0.004 | 4.16E-01 | -0.052 | 7.67E-06 | no |
| **rs76074510** | 8 | 64862463 | -0.064 | 2.77E-06 | 0.003 | 5.63E-01 | | | | | | | no |
| **rs7689919** | 4 | 36330993 | 0.052 | 1.95E-06 | -0.002 | 7.10E-01 | | | | | | | no |
| **rs7717864** | 5 | 59486768 | 0.063 | 7.40E-06 | | | rs35265720 | 59359191 | 0.0032 | 6.45E-01 | 0.060 | 1.71E-05 | yes |
| **rs77321042** | 3 | 53179543 | 0.105 | 4.75E-06 | 0.004 | 6.99E-01 | | | | | | | yes |
| **rs7747481** | 6 | 98315696 | 0.058 | 2.29E-08 | -0.003 | 4.89E-01 | | | | | | | no |
| **rs778346** | 2 | 233799736 | -0.052 | 2.63E-06 | | | rs1996342 | 233805499 | 0.005 | 2.90E-01 | -0.040 | 3.06E-05 | no |
| **rs7903084** | 10 | 10861098 | -0.069 | 1.25E-07 | -0.004 | 5.17E-01 | | | | | | | yes |
| **rs79621462** | 6 | 152254115 | 0.048 | 7.55E-06 | | | rs9479142 | 152244787 | -0.001 | 8.10E-01 | 0.050 | 9.32E-06 | no |
| **rs7988108** | 13 | 59820720 | 0.051 | 2.36E-06 | 0.002 | 6.20E-01 | | | | | | | yes |
| **rs79889335** | **9** | **78550303** | **0.087** | **6.93E-06** | | | **rs35428069** | **78579131** | **-0.025** | **1.30E-03** | **0.062** | **9.65E-04** | **no** |

| **Resilience** | | | | | **Recall** | | **Proxy Recall** | | | | **Proxy Resilience** | | **Sign** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **SNP** | **Chr** | **Position** | **Beta** | **P** | **Beta** | **P** | **SNP** | **Position** | **Beta** | **P** | **Beta** | **P** | |
| **rs79967991** | 6 | 85822584 | -0.205 | 2.08E-06 | -0.003 | 8.47E-01 | | | | | | | yes |
| **rs8019612** | 14 | 29822421 | 0.049 | 2.22E-06 | | | rs11849411 | 29845225 | 0.003 | 5.03E-01 | 0.047 | 6.39E-06 | yes |
| **rs843370** | 3 | 183875822 | -0.047 | 7.79E-06 | 0.002 | 6.88E-01 | | | | | | | no |
| **rs9569811** | 13 | 58646190 | -0.086 | 1.14E-08 | 0.009 | 1.51E-01 | | | | | | | no |
| **rs9606967** | 22 | 32899516 | -0.056 | 5.42E-06 | | | rs8142308 | 32905470 | -0.007 | 2.00E-01 | -0.050 | 3.12E-05 | yes |
| **rs9911735** | 17 | 77778226 | -0.061 | 5.89E-06 | | | rs1285248 | 77810444 | -0.009 | 1.40E-01 | -0.038 | 7.25E-03 | yes |
| **rs9955131** | 18 | 34524533 | 0.047 | 7.86E-06 | | | rs11661504 | 34518724 | -0.003 | 5.80E-01 | 0.048 | 1.94E-04 | no |

Note: All SNPs that have a nominally significant P value are in bold. No proxy is where there was no comparable SNP in either dataset with an LD R value of >0.4. The minor allele was reversed for the Resilience and HRS in rs853982 - proxy for rs6918725.

# 6 Discussion

## 6.1 Generating a resilience GWAS:

During my research for this thesis, it became obvious that the genetic contributors to cognitive resilience were poorly understood. This was the case due to (a) the polygenic nature of cognition that requires large datasets to identify associated genetic variation and (b) the long timeframe over which cognitive data needs to be collected to provide longitudinal phenotypic data spanning several years or even decades. The added complexity is gene/environment interactions that influence resilience. In chapter 1, I discuss the difficulty in operationalising cognitive resilience and lack of progress to date.

Driven by the need to understand the genetics of numerous complex human traits, biobank projects have been instigated to collect genetic and phenotypic data on very large samples of individuals. The first of these that is now publicly available for research, is the UKB where genetic and phenotypic data on approximately 500,000 participants has led to numerous publications that have improved our understanding of many human traits.

As cognitive ageing has not been subject to large scale GWAS, I used the UKB in an attempt to fill an important gap in our understanding of the molecular genetic basis for cognitive resilience. While cognitive data is available in the UKB, the longitudinal cognitive data is of short duration and is only measured in a relatively low number of participants. In the absence of sufficient longitudinal data, I examined novel approaches to use this extensive resource to study cognitive resilience. In Chapter 3, I showed how using the proxy phenotype of education years for past cognitive performance and processing speed as measured by reaction time for current cognitive performance, could generate longitudinal cognitive resilience data over a 40-year period.

The generation of a GWAS of *Resilience* proved complex due to interference from the highly heritable EY phenotype and I determined that to overcome this I needed to subtract the results of one GWAS from another – one showing resilience and EY and the other showing just EY. Fortunately, research was published around this time, using a novel bioinformatics tool, GenomicSEM, which allowed the subtraction of one GWAS from another. This method enabled me to separate out the genetic contribution to *Resilience* in a GWAS that I could bring forward for further analysis.

## 6.2 Replication and GWAS output

As has become expected in current genomic research, it is necessary to show replication of the results in an independent sample. Given the number of steps involved in extracting the *Resilience* GWAS, demonstrating replication was essential. I was able to demonstrate replication using an independent sample from within the UKB that was separate from the discovery sample. My results showed excellent agreement between the two GWAS, demonstrating replication and allowing me to proceed to functional analysis on the full sample.

Once I showed that the results replicated, I then combined the discovery and replication sample into a GWAS of the full sample. This GWAS identified 13 independent genome-wide significant loci resulting in 366 mapped genes and 33 prioritized genes for *Resilience*.

## 6.3 Functional analysis

A GWAS identifies genetic loci associated with particular traits but in order to use these findings to understand the potential biology underpinning these traits, it is necessary to use bioinformatic tools to integrate these findings with those from functional genomic datasets such as gene expression or chromatin interactions across tissue and cell types (Cano-Gamez & Trynka, 2020).

Chapter 4 of this thesis describes the various methods used to understand the biological nature of *Resilience*. Functional analysis showed significant expression of associated genes in all brain tissues, and particularly in the frontal cortex. Significant enrichment of associated genes was also found at the cellular level in both GABAergic and glutamatergic neurons indicating an excitatory/inhibitory function in the prefrontal cortex, and within biological processes related to neuron differentiation and synaptic activity.

One GO term that was nominally significant for enrichment with a very large beta value was the Wnt signalosome. There are 12 genes in this gene set that code for scaffolding proteins to form a signalosome at the cell membrane that provide a platform for interaction with effector proteins downstream and are involved in the transduction of Wnt signals (Gerlach et al., 2018). The formation of the signalosome is part of the canonical Wnt pathway. Reduced levels of canonical Wnt signalling leads to synapse disassembly resulting in compromised synaptic signalling in the ageing brain (Palomer et al., 2019).

Mapping of GWAS results, identified genes that have been previously associated with cognitive decline including *STAU1*, *SEMF3A*, *IP6K1*, *MST1*, the *ATNX2/BRAP* locus, *ALDH2* and *DDX27*, where a likely functional missense variant is highly associated. Other associated genes involved with synaptic activity and neurogenesis include *BNS*, *DAG1*, *IP6K1* and *TET2*, pointing to potential targets for improvement of cognitive resilience.

Several questions arose during this analysis that I examined in Chapter 5. The first of these is what is the influence of reaction time (RT) on *Resilience*? My study relied on RT to create the *Resilience* phenotype, which results in a strong genetic correlation between *Resilience* and RT (Section3.2.9). This reflected my study design that detected SNPs associated with faster than average RT or processing speed in individuals that previously showed below average EY. However, the majority of genes prioritized by my *Resilience* GWAS are not prioritized by the RT GWAS and vice versa. I conclude that these findings point to genes where genetic variation enhances maintenance of processing speed over the life span.

The genetic correlation between *Resilience* and RT is strong because this is an RT-based resilience phenotype. However, there are differences in the associated genes being detected. This phenotype enabled the identification of genetic differences between those individuals in the UKB who preserved or maintained their capability to process information and respond over a 40-year time period compared to individuals who showed diminishing processing speed.

The second question that arose from the functional analysis was whether the large associated locus on chromosome 3 was overrepresented in the findings? Processing a GWAS without this locus through FUMA did indeed show a substantial decrease in candidate SNPs and mapped genes, however, the main findings of the functional analysis remained unchanged. This region on chromosome 3 is gene rich and had been associated with traits associated with cognition in the past, for example educational attainment (Davies et al., 2016) (Lee et al., 2018) and intelligence (Savage et al., 2018) but not RT (Davies et al., 2018).

### 6.3 Replication of findings in an external dataset:

Although I showed robustness of the method used to determine *Resilience* in the UKB by showing equivalent results in a discovery and replication sample and finding prioritized genes that were previously connected to cognitive decline, to confirm my findings I would like to replicate them in a dataset other than the UKB. As discussed in Section 1.3.4, suitable

datasets containing large genetic data coupled with strong longitudinal cognitive data are not available. However, the US HRS had strong longitudinal data on a relatively small number of individuals (in genomic analysis terms) so I used these data (n=9,526) to see if I could replicate my findings. I showed a genetic correlation between our Resilience findings in UKB and cognitive change in HRS ($r_g$ = -0.64, $P$ = 1.5 x $10^{-2}$). However, most likely due to the low sample size, I was not able to replicate the genetic variants associated with *Resilience* genetic variants. Overcoming this issue of insufficient sample sizes available to confirm associations observed here can hopefully be addressed by various biobanks that plan new data collection in the future.

My work with the HRS dataset is a demonstration of the effect of sample size. By increasing my analysis from an initial sample size of 5,345 to 9,526, I found more SNPs with lower P values approaching significance and improved the Q-Q plot of the outcome. This confirms what was previously found with other cognitive studies (see Figure 1.11). Understanding of the genetics of cognitive resilience would be greatly enhanced by the generation of large datasets with strong longitudinal cognitive data.

An alternative approach to replication with longitudinal data in an external dataset is to replicate the method used to examine *Resilience* in the UKB, in a second large dataset using a proxy phenotype. Very recent research has used cognitive data generated within the 23andMe dataset on over 300,000 individuals with cognitive measures of digit symbol substitution (n=132,807) and the Flicker test (n=158,888), both of which contain a strong element of processing speed. The participants had an average EY of 16 years (Carey et al., 2020). There is potential to analyse these data using a similar approach to my study to augment the findings with the UKB.

## 6.4 Limitation of current analysis

A limitation to my approach in this study was the use of a proxy phenotype for past cognitive performance that is different to the measure used for current cognitive performance. This was done to allow full use of the UKB. I used the proxy phenotype of academic achievement (EY) to represent past cognitive performance in the absence of a direct measure of processing speed. In support of this approach, a study using a sample of 1,560 pupils found that information processing speed is the key predictor of number sense, fluid intelligence and working memory, which in turn predict individual difference in academic achievement (Tikhomirova et al., 2020).

A further limitation is that processing speed as measured by RT is only one component of cognition and it may not be possible to extrapolate the results of this analysis to global cognitive resilience. I explored this using a general cognitive measure ‘g’ by combining various cognitive measures in the UKB as described in Section 3.2, however, this resulted in a considerable reduction in sample size for analysis. Therefore, to maximise the sample size available, I used RT alone as this parameter was measured in nearly all participants.

The creation of binary phenotypes with high and low EY and RT had an influence on the direction of effect on the correlation of *Resilience* with other traits (see Section 3.2.9). This was driven by the lack of a direct measure of cognitive decline in the UKB and hopefully will be clarified in future datasets.

The findings from functional analysis are limited by current knowledge of the function of genes involved in neurological processes. As our knowledge increases with the development of deeper and more accurate datasets (see section 6.6), our findings may change.

The moderation effects of environmental factors on cognitive decline (see section 1.1.4) is well recognised. My research, however, did not include environmental effects and gene/environment interactions.

## 6.5 Understanding the genetic contribution to cognitive resilience

In Chapter 1 (Section 6), I listed six questions for my thesis to address. The first four have already been discussed where I have shown that individual genetic variations associated with cognitive resilience can be confirmed in the UKB and this finding is novel. I demonstrated replication within the UKB and highlighted brain regions, biological processes and biological pathways that are involved in cogntive resilience. I also showed limited confirmation of these findings in external datasets.

There are two outstanding questions for discussion. The first of these is whether the findings support the contribution of brain reserve, cognitive reserve, and brain maintenance to cognitive resilience?

In examining the findings of this research, I concluded that they support the theory that maintenance of processing speed over the life span is associated with *Resilience* to cognitive decline. Analysis of gene expression of associated genes shows enrichment in all brain regions, and most prominently in excitatory/inhibitory control in the prefrontal cortex. This coupled with a link to biological processes related to synaptic activity would suggest

enhanced synaptic activity due to strong cognitive reserve and maintenance of existing networks. However, increased synaptic activity may also be due to compensatory mechanisms where alternative pathways are activated when existing ones are damaged. Functional analysis also shows prioritized genes that have been previously linked to cognitive decline and of particular interest are genes involved with synaptic activity and neurogenesis including *BNS, DAG1, IP6K1* and *TET2*, which would support cognitive reserve and maintenance theories.

The causal link between resilience and white matter (WM) volume is interesting. WM volume increases during the life span and peaks at around 50 years of age, but then decreases from 60 years of age onwards. WM lesions are associated with cognitive decline (Liu et al., 2017) indicating a role for both brain reserve (more WM to start) and maintenance of existing WM.

The remaining question asks if there is more to superior cognitive resilience than superior intelligence?

The *Resilience* GWAS was derived by examining the common genetic variants in individuals that had lower than average education years (EY) and higher than average reaction time 40 years later. Given the correlation of EY and intelligence, this group have lower than average intelligence but have preserved cognitive function better when compared to others with higher intelligence. Seven of the thirteen genome-wide significant loci for *Resilience* are not associated with a recent GWAS of intelligence (Savage et al., 2018), indicating that factors such as reserve, compensation and maintenance may play a role over and above overall intelligence in determining resilience.

## 6.6 Future studies

There has been considerable and rapid progress in identifying the genetic architecture of cognitive performance in recent years (see Section 1.4.1.5). This has been aided, perhaps equally, by both improvements in genomic methods, and the increasing availability of data due to data sharing and cooperation. This progress has resulted in a stronger picture of the highly polygenic basis of cognitive performance, and of the multiple biological processes involved. The contribution of common genetic variation to explaining variation in cognitive performance is clear. The contribution of rare(er) variants both to intellectual disability but also cognitive variation in the general population is also clear (Ganna et al., 2016). The overlap, but also the discontinuity, between the polygenic variation underpinning cognitive

function, illness risk, and cognitive decline is also beginning to come into view. With this clarity, the need for further development of analytical and bioinformatic approaches to understand the biology of these processes has become visible. In particular, the need to more sharply identify the myriad biological pathways underpinning cognitive function is underlined (e.g., the contribution of oligodendroglial-related genetic variation to cognitive performance). Similarly, the need to model how genetic variation - both common and rare - interacts with environmental factors to predict cognitive performance is also a clear priority. Given the progress made in the past 5-10 years, furthering these objectives continues to hold significant promise for understanding cognitive ability.

With a growing understanding of cognitive function, there should be a parallel growth in the understanding of what factors constitutes cognitive decline. Hopefully, the generation of large datasets with strong longitudinal, cognitive data will assist in this understanding.

The type of study I would propose, given unlimited resources and time, to study cognitive resilience in healthy ageing is a long-term study on a large cohort (greater than 1 million) of cognitively healthy adults. The reason that a dataset of this size is needed is that the effect sizes of genetic variants associated with cognitive resilience are tiny, so very large samples are needed to detect genetic influences. The study would need to consider ethnic diversity in parallel and combined studies. Using this approach would allow us to boost our sample size and examine the divergence in findings in different groups while also allowing us to examine LD pattern differences and enhance our fine mapping findings (Fernández-Rhodes et al., 2017; Lam, Chen, et al., 2019). The participants should be in the age range of 30 to 40 as recent research shows that this is the age of peak cognitive ability (Strittmatter, Sunde, & Zegners, 2020). I would use robust cognitive measures at base line and every five subsequent years (given the slow rate of change associated with cognitive decline) until death. As well as performing full genetic testing at baseline, I would also include epigenetic markers including, histone modifications and chromatin remodelling and DNA methylation (Zhang, Qu, Liu, & Belmonte, 2020) in blood samples at each time point and *post-mortem* brain samples. In addition, environmental measures such as fitness, weight, diet, smoking status, alcohol consumption, stress markers and social interactions will be recorded at each interval to enable gene/environment modelling. Recent findings on the role of rare and ultrarare variants in cognitive variation would warrant an analysis of their effects (see Section 1.4.1.3). Full genome sequencing would be performed to allow for the study of rare and ultra-rare variants that can shed light on the biological processes involved(Singh, Neale, & Daly, 2020).

Studying non-coding RNA has shown that age related disturbances of long non-coding RNA (lncRNA) expression may affect synaptic activity and neurogenesis (Pereira Fernandes, Bitar, Jacobs, & Barry, 2018).

In the meantime, as the above proposed study would take years to generate meaningful data another alternative approach using related phenotypes such as longevity may be helpful. A study of cognition in 340 cognitively healthy Dutch centenarians with an average age of 100.5 years was published recently. They went through a battery of cognitive tests which was repeated on an average of 1.6 years later and showed no cognitive decline. *Post-mortem* brain examinations were performed on 44 individuals which showed varying loads of hallmarks of AD, but this was not associated with cognitive decline. The research concludes that resilient individuals preserve cognition to exceptional ages despite the presence of risk factors of cognitive decline (Beker et al., 2021). However, this research did not include genetic analysis. As we are an ageing population, the number of people reaching older age is increasing so there is a possibility of collecting genetic data and cognitive data on a large population of cognitively healthy individuals over 85 years of age. Performing a GWAS using this data and comparing it to results of younger cohorts could pinpoint the genetic variants that are important to resilience. The cognitive data could be repeated 5 years later.

Datasets available for functional analysis are increasing rapidly and should increase the accuracy in linking genetic findings to biological processes. Recent advancements in single-cell RNA sequencing including microfluidics techniques and combinatorial indexing, and reduction in sequencing costs has greatly increased the output of available datasets and is adding to our knowledge of the functional consequences of genetic variation. While several published cell atlases are currently available, including the mouse brain, availability of human brain cell data is lacking. However, endeavours are underway to map human cells under the initiative of the Human Cell Atlas (Lähnemann et al., 2020; Regev et al., 2018). Limited data is currently available for human tissue including the developing brain (Eze, Bhaduri, Haeussler, Nowakowski, & Kriegstein, 2021). Adult brain cell data is acquired mainly through *post-mortem* samples, and this has proven a challenge for accurate recovery of neurons. Using single nuclei has improved results (Colonna & Brioschi, 2020). In addition, improving eQTL data from tissues is ongoing but with the development of single-cell RNA-sequencing there are emerging opportunities for mapping eQTLs in dynamic processes and across different cell types. To this end the single-cell eQTLGen consortium had been established (van der Wijst et al., 2020).

Once the genetic variants associated with resilience are identified and functional analysis has indicated the biological pathways involved, I would then generate a PRS to identify, at an early stage, those individuals that are a vulnerable to cognitive decline who would benefit from either lifestyle changes or therapeutic intervention (if available). Pointing out a genetic vulnerability might persuade people to make these changes. The use of PRSs for polygenic traits is currently not advised due to lack of ethnic diversity in GWAS (Martin, Daly, Robinson, Hyman, & Neale, 2019; Martin, Kanai, et al., 2019). In addition, ethical use of PRS data needs to be considered. Consideration should also be given to including rare variant and copy number variants, and environmental effects when generating a PRS (Fries, 2020).

Past candidate gene studies were generally unsuccessful as they were based on findings from limited studies and failed to replicate when larger samples were used (Chabris et al., 2012), however, with the advent of large GWAS, we have greater statistical assurance of genes involvement (see Section 1.4.1.4). Environmental parameters are particularly important in the study of cognitive decline and resilience given the proven role that lifestyle factors play in maintaining cognitive function (see Section 1.1.4) (Hasan & Afzal, 2019).

Another approach using data generated by large scale GWAS is to investigate drug-gene interactions and druggable genes using the Drug-Gene Interaction Database (DGIdb) which gathers information on drug-gene interactions (Freshour et al., 2021). The use of this tool to examine drug-genes interactions was demonstrated in a study of genes associated with depression (Howard et al., 2019).

Modelling environment interactions outputs on cognitive resilience using multifactorial analysis or SEM was discussed earlier (See section 1.3.3.2). Tools are being developed further to incorporate gene/environment interactions (Briley, Harden, Bates, & Tucker-Drob, 2015). The incorporation of GWAS into structural equation modelling is possible through the advent of GenomicSEM (Grotzinger et al., 2019) and several modelling variations are described. One of these (GBS) was used as a pivotal tool in this thesis.

Suitable lifestyle interventions to improve cognitive resilience are already well understood. However, therapeutic targets to boost resilience to cognitive decline in healthy ageing are not available. It is suggested that prevention of oxidative stress, which is associated with epigenetic changes in various systems, is a proposed strategy and has been examined in mouse models. Prevention of oxidative damage using compounds that inhibit apoptosis,

reduce reactive oxygen species or preserve chromatin structure in genes involved in learning and memory are proposed (Kandlur, Satyamoorthy, & Gangadharan, 2020).

My research highlighted a number of genes involved in synaptic activity including the Wnt/β-Catenin Signalling. Downregulation of this pathway is associated with cognitive decline in the elderly but can be attenuated by exercise in rat models (Chen et al., 2020). Restoring Wnt/β-catenin signalling is proposed as a therapeutic strategy for AD and general cognitive ageing (Jia, Piña-Crespo, & Li, 2019). In addition, this pathway has been extensively studied as it is highly activated in many human cancers which has led to the development of various Wnt signalling inhibitors for cancer therapies (Jung & Park, 2020). Perhaps antidotes to these inhibitors could make potential therapeutic targets.

A further extension of therapeutic targets is the use of pharmacogenomics where genetic variants are identified that influence how individuals respond to medications to tailor their treatment. However, currently, there is little accumulated evidence of the clinical validity and utility of pharmacogenomics testing in the medication management of older adults. More research is needed so that this can be implemented in ageing subjects with comorbid conditions (Inventor & Paun, 2021).

## 6.7 Further use of tools used in my research

The ability to subtract one GWAS from another using GenomicSEM could be extended to other areas of interest within the CogGene group at NUIG. For example, preliminary research showed that subtraction of the most recent GWAS of schizophrenia (Ripke, Walters, & O'Donovan, 2020) and bipolar disorder (Mullins et al., 2021) in both directions generated interesting results and has the potential to be a future publication.

Exploring the usefulness of FUMA as a functional analysis tool proved very useful. The ability to perform analysis on the one platform has advantages, for example, I initially looked at performing single–cell analysis using tools in R which required further data formatting, and there were limitations in dataset availability. I discovered that this analysis was available in FUMA and could be performed on my GWAS output which was already in FUMA and allowed access to a wide range of single-cell expression datasets.

## 6.8 Challenges

Given the highly polygenetic nature of cognition, the uncertainty around what constitutes intelligence (Deary & Sternberg, 2021), and the biological processes underpinning it, determining the genetics of cognitive resilience is a major challenge. I needed to be creative

in my approach to engineer longitudinal measures of cognition in large datasets and there really was no precedent to follow. In this thesis I have started to unravel the biology of cognitive resilience using big data.

Another challenge with the emergence of big data is the support services needed to store, manage, process, and analyse the data. Unless one is based in a large institute with full IT infrastructure support, this is a challenge and has proved a challenge during my thesis. Lack of background expertise and computer servers configured to accept large datasets and the capacity and software to manipulate these data was a challenge. However, this was also an opportunity to delve into these processes for a greater understanding. In the future, the switch to using cloud computing services such as Amazon Marketing Services (AMS) and others should eliminate many of these challenges. The UKB plans to switch to this format for data access as opposed to local downloads within the next couple of months.

A further challenge for me was my IT skills. Having grown up in a world when having expertise in Excel and SPSS was considered an advantage, I soon learnt that these packages were not suitable for large datasets and needed to learn how to manipulate data in Linux using bash commands and run packages in R. This is just a consequence of changing times and being part of an older demographic that evolved with different computer skills to the current generation.

A further challenge in processing genetic data is learning to navigate its complicated background. As the field is rapidly evolving and improving methods of collecting and defining phenotypic and genetic data, working with older sets can present challenges. When using research data, one needs to know the origin of the reference genome. In this thesis, I used GRCh37 as this is the reference build used in the UKB and most public datasets. In addition, how SNP alleles are defined can differ between studies. The HRS data had SNPs containing kpg identification numbers instead of rsID numbers and needed to be converted. Genetic call data can be presented as a bed file (binary files) or pgen files (probability files) and pgen files will not run-on certain platforms.

An additional challenge during this thesis was the restrictions on research caused by the Covid pandemic. Fortunately, I had spent my first two years with direct access to the facilities at NUIG and the expertise of my colleagues. This proved to be a challenge during lockdown, but I was fortunate in that my work was advanced enough to finish my analysis from home.

## 6.9 Concluding remarks

Ageing is an inevitable part of our life cycle and can be compromised by ill health and cognitive and physical impairment. As life expectancy increases, understanding ageing processes is essential to implement effective policies that promote healthy ageing. It is now well accepted that a lifetime regime of healthy eating, exercise, sociability, regular sleep, and reduced stress can maintain wellness into older age. However, the biological pathways involved are poorly understood.

Loosing cognitive ability is one of the most feared aspects of ageing and results in increased difficulty in performing tasks that require memory or rapid information processing and can have an increasingly detrimental effect on quality of life. Some people show a remarkable cognitive ability right through the ageing process while other non-pathologically healthy adults flounder. Understanding this variation could be a key to implementing therapeutic or lifestyle changes that increase cognitive resilience.

My thesis addresses one aspect of the biological basis of healthy cognitive ageing by identifying common genetic variation associated with greater or lesser resilience to cognitive decline. And in so doing this helpfully informs us about some of the genetic architecture of ageing, and the biological processes involved. It also lays the foundation for future studies in this area.

References

Abdellaoui, A., Nivard, M. G., Hottenga, J. J., Fedko, I., Verweij, K. J. H., Baselmans, B. M. L., . . . Cacioppo, J. T. (2018). Predicting Loneliness with Polygenic Scores of Social, Psychological, and Psychiatric Traits. *Genes Brain Behav*, e12472. doi:10.1111/gbb.12472

Adams, H. H., Hibar, D. P., Chouraki, V., Stein, J. L., Nyquist, P. A., Renteria, M. E., . . . Thompson, P. M. (2016). Novel genetic loci underlying human intracranial volume identified through genome-wide association. *Nat Neurosci, 19*(12), 1569-1582. doi:10.1038/nn.4398

Aiello, A., Farzaneh, F., Candore, G., Caruso, C., Davinelli, S., Gambino, C. M., . . . Accardi, G. (2019). Immunosenescence and Its Hallmarks: How to Oppose Aging Strategically? A Review of Potential Options for Therapeutic Intervention. *Frontiers in Immunology, 10*(2247). doi:10.3389/fimmu.2019.02247

Andrews, S. J., Das, D., Anstey, K. J., & Easteal, S. (2017). Association of AKAP6 and MIR2113 with cognitive performance in a population-based sample of older adults. *Genes, Brain and Behavior, 16*(4), 472-478. doi:10.1111/gbb.12368

Andrews, S. J., Das, D., Cherbuin, N., Anstey, K. J., & Easteal, S. (2016). Association of genetic risk factors with cognitive decline: the PATH through life project. *Neurobiol Aging, 41*, 150-158. doi:10.1016/j.neurobiolaging.2016.02.016

Annamneedi, A., Caliskan, G., Müller, S., Montag, D., Budinger, E., Angenstein, F., . . . Stork, O. (2018). Ablation of the presynaptic organizer Bassoon in excitatory neurons retards dentate gyrus maturation and enhances learning performance. *Brain structure & function, 223*(7), 3423-3445. doi:10.1007/s00429-018-1692-3

Arpawong, T. E., Pendleton, N., Mekli, K., McArdle, J. J., Gatz, M., Armoskus, C., . . . Prescott, C. A. (2017). Genetic variants specific to aging-related verbal memory: Insights from GWASs in a population-based cohort. *PLoS One, 12*(8), e0182448. doi:10.1371/journal.pone.0182448

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., . . . Sherlock, G. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics, 25*(1), 25-29. doi:10.1038/75556

Baddeley, A., Eysenck, M., & Anderson, M. (2014). Memory (2nd edn). *Hoboken: Taylor and Francis*.

Battle, A., Brown, C. D., Engelhardt, B. E., & Montgomery, S. B. (2017). Genetic effects on gene expression across human tissues. *Nature, 550*(7675), 204-213. doi:10.1038/nature24277

Beker, N., Ganz, A., Hulsman, M., Klausch, T., Schmand, B., Scheltens, P., . . . Holstege, H. (2021). Association of Cognitive Function Trajectories in Centenarians With Postmortem Neuropathology, Physical Health, and Other Risk Factors for Cognitive Decline. *JAMA Network Open, 4*, e2031654. doi:10.1001/jamanetworkopen.2020.31654

Benner, C., Spencer, C. C. A., Havulinna, A. S., Salomaa, V., Ripatti, S., & Pirinen, M. (2016). FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics, 32*(10), 1493-1501. doi:10.1093/bioinformatics/btw018

Bettcher, B. M., Gross, A. L., Gavett, B. E., Widaman, K. F., Fletcher, E., Dowling, N. M., . . . Mungas, D. (2019). Dynamic change of cognitive reserve: associations with changes in brain, cognition, and diagnosis. *Neurobiology of Aging, 83*, 95-104. doi:https://doi.org/10.1016/j.neurobiolaging.2019.08.016

Boldrini, M., Fulmore, C. A., Tartt, A. N., Simeon, L. R., Pavlova, I., Poposka, V., . . . Mann, J. J. (2018). Human Hippocampal Neurogenesis Persists throughout Aging. *Cell Stem Cell, 22*(4), 589-599.e585. doi:10.1016/j.stem.2018.03.015

Bonder, M. J., Luijk, R., Zhernakova, D. V., Moed, M., Deelen, P., Vermaat, M., . . . the, B. C. (2017). Disease variants alter transcription factor levels and methylation of their binding sites. *Nature Genetics, 49*(1), 131-138. doi:10.1038/ng.3721

Bott, N. T., Bettcher, B. M., Yokoyama, J. S., Frazier, D. T., Wynn, M., Karydas, A., . . . Kramer, J. H. (2017). Youthful Processing Speed in Older Adults: Genetic,

Biological, and Behavioral Predictors of Cognitive Processing Speed Trajectories in Aging. *Frontiers in Aging Neuroscience, 9*(55). doi:10.3389/fnagi.2017.00055

Brandt, J., Spencer, M., & Folstein, M. (1988). The Telephone Interview for Cognitive Status. *Cognitive and Behavioral Neurology, 1*.

Briley, D. A., Harden, K. P., Bates, T. C., & Tucker-Drob, E. M. (2015). Nonparametric Estimates of Gene × Environment Interaction Using Local Structural Equation Modeling. *Behavior Genetics, 45*(5), 581-596. doi:10.1007/s10519-015-9732-8

Briley, D. A., & Tucker-Drob, E. M. (2013). Explaining the increasing heritability of cognitive ability across development: a meta-analysis of longitudinal twin and adoption studies. *Psychological science, 24*(9), 1704-1713. doi:10.1177/0956797613478618

Brown, R. E. (2016). Hebb and Cattell: The Genesis of the Theory of Fluid and Crystallized Intelligence. *Frontiers in Human Neuroscience, 10*(606). doi:10.3389/fnhum.2016.00606

Buniello, A., MacArthur, J. A L., Cerezo, M., Harris, L. W., Hayhurst, J., Malangone, C., . . . Parkinson, H. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Research, 47*(D1), D1005-D1012. doi:10.1093/nar/gky1120

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., . . . Marchini, J. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature, 562*(7726), 203-209. doi:10.1038/s41586-018-0579-z

Cabello, R., Navarro Bravo, B., Latorre, J. M., & Fernandez-Berrocal, P. (2014). Ability of university-level education to prevent age-related decline in emotional intelligence. *Front Aging Neurosci, 6*, 37. doi:10.3389/fnagi.2014.00037

Cabeza, R., Albert, M., Belleville, S., Craik, F. I. M., Duarte, A., Grady, C. L., . . . Rajah, M. N. (2018). Maintenance, reserve and compensation: the cognitive neuroscience of healthy ageing. *Nat Rev Neurosci, 19*(11), 701-710. doi:10.1038/s41583-018-0068-2

Cadar, D., Robitaille, A., Pattie, A., Deary, I., & Muniz-Terrera, G. (2020). The Long Arm of Childhood Intelligence on Terminal Decline: Evidence From the Lothian Birth Cohort 1921. *Psychol Aging, 35*(6), 806-817. doi:10.1037/pag0000477

Cano-Gamez, E., & Trynka, G. (2020). From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases. *Frontiers in genetics, 11*(424). doi:10.3389/fgene.2020.00424

Carey, C. E., Huang, Y., Strong, R. W., Aslibekyan, S., Gentleman, R. C., Smoller, J. W., . . . Germine, L. (2020). Shared and distinct genetic influences between cognitive domains and psychiatric disorder risk based on genome-wide data. *bioRxiv*, 2020.2009.2016.297408. doi:10.1101/2020.09.16.297408

Carithers, L. J., Ardlie, K., Barcus, M., Branton, P. A., Britton, A., Buia, S. A., . . . Moore, H. M. (2015). A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project. *Biopreserv Biobank, 13*(5), 311-319. doi:10.1089/bio.2015.0032

Chabris, C. F., Hebert, B. M., Benjamin, D. J., Beauchamp, J., Cesarini, D., van der Loos, M., . . . Laibson, D. (2012). Most reported genetic associations with general intelligence are probably false positives. *Psychological science, 23*(11), 1314-1323. doi:10.1177/0956797611435528

Chan, J. S. Y., Deng, K., Wu, J., & Yan, J. H. (2019). Effects of Meditation and Mind–Body Exercises on Older Adults' Cognitive Performance: A Meta-analysis. *The Gerontologist, 59*(6), e782-e790. doi:10.1093/geront/gnz022

Chang, C. C., Chow, C. C., Tellier, L. C. A. M., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience, 4*(1). doi:10.1186/s13742-015-0047-8

Cheesman, R., Hunjan, A., Coleman, J. R. I., Ahmadzadeh, Y., Plomin, R., McAdams, T. A., . . . Breen, G. (2019). Comparison of adopted and non-adopted individuals reveals gene-environment interplay for education in the UK Biobank. *bioRxiv*, 707695. doi:10.1101/707695

Cheesman, R., Hunjan, A., Coleman, J. R. I., Ahmadzadeh, Y., Plomin, R., McAdams, T. A., . . . Breen, G. (2020). Comparison of Adopted and Nonadopted Individuals Reveals

Gene–Environment Interplay for Education in the UK Biobank. *Psychological science, 31*(5), 582-591. doi:10.1177/0956797620904450

Chen, C.-Y., Lam, M., Tian, R., Liu, J., Tsai, E., Whelan, C., . . . Runz, H. (2021). Large-Scale Exome Sequencing Identifies Novel Genes for Cognitive Function. *Biological Psychiatry, 89*(9, Supplement), S45. doi:https://doi.org/10.1016/j.biopsych.2021.02.129

Chen, D., Zhang, Y., Zhang, M., Chang, J., Zeng, Z., Kou, X., & Chen, N. (2020). Exercise Attenuates Brain Aging by Rescuing Down-Regulated Wnt/β-Catenin Signaling in Aged Rats. *Frontiers in Aging Neuroscience, 12*(105). doi:10.3389/fnagi.2020.00105

Chiang, C., Scott, A. J., Davis, J. R., Tsang, E. K., Li, X., Kim, Y., . . . Hall, I. M. (2017). The impact of structural variation on human gene expression. *Nat Genet, 49*(5), 692-699. doi:10.1038/ng.3834

Chopra, S., Shaw, M., Shaw, T., Sachdev, P. S., Anstey, K. J., & Cherbuin, N. (2018). More highly myelinated white matter tracts are associated with faster processing speed in healthy adults. *NeuroImage, 171*, 332-340. doi:https://doi.org/10.1016/j.neuroimage.2017.12.069

Coleman, J. R. I., Bryois, J., Gaspar, H. A., Jansen, P. R., Savage, J. E., Skene, N., . . . Breen, G. (2019). Biological annotation of genetic loci associated with intelligence in a meta-analysis of 87,740 individuals. *Molecular Psychiatry, 24*(2), 182-197. doi:10.1038/s41380-018-0040-6

Colonna, M., & Brioschi, S. (2020). Neuroinflammation and neurodegeneration in human brain at single-cell resolution. *Nature Reviews Immunology, 20*(2), 81-82. doi:10.1038/s41577-019-0262-0

Consortium, G. T. (2015). Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science (New York, N.Y.), 348*(6235), 648-660. doi:10.1126/science.1262110

Corley, J., Cox, S., & Deary, I. (2018). Healthy cognitive ageing in the Lothian Birth Cohort studies: marginal gains not magic bullet. *Psychological medicine, 48*(2), 187-207.

Crimmins, E. M., Kim, J. K., Langa, K. M., & Weir, D. R. (2011). Assessment of cognition using surveys and neuropsychological assessment: the Health and Retirement Study and the Aging, Demographics, and Memory Study. *J Gerontol B Psychol Sci Soc Sci, 66 Suppl 1*, i162-171. doi:10.1093/geronb/gbr048

Cunningham, C. (2013). Microglia and neurodegeneration: the role of systemic inflammation. *Glia, 61*(1), 71-90.

Daffner, K. R. (2010). Promoting successful cognitive aging: a comprehensive review. *J Alzheimers Dis, 19*(4), 1101-1122. doi:10.3233/JAD-2010-1306

Darmanis, S., Sloan, S. A., Zhang, Y., Enge, M., Caneda, C., Shuer, L. M., . . . Quake, S. R. (2015). A survey of human brain transcriptome diversity at the single cell level. *Proceedings of the National Academy of Sciences of the United States of America, 112*(23), 7285-7290. doi:10.1073/pnas.1507125112

Daselaar, S. M., Iyengar, V., Davis, S. W., Eklund, K., Hayes, S. M., & Cabeza, R. E. (2015). Less wiring, more firing: low-performing older adults compensate for impaired white matter with greater neural activity. *Cereb Cortex, 25*(4), 983-990. doi:10.1093/cercor/bht289

Dassel, K. B., & Carr, D. C. (2014). Does Dementia Caregiving Accelerate Frailty? Findings From the Health and Retirement Study. *The Gerontologist, 56*(3), 444-450. doi:10.1093/geront/gnu078

Davies, G., Armstrong, N., Bis, J. C., Bressler, J., Chouraki, V., Giddaluru, S., . . . Deary, I. J. (2015). Genetic contributions to variation in general cognitive function: a meta-analysis of genome-wide association studies in the CHARGE consortium (N=53949). *Molecular Psychiatry, 20*(2), 183-192. doi:10.1038/mp.2014.188

Davies, G., Lam, M., Harris, S. E., Trampush, J. W., Luciano, M., Hill, W. D., . . . Deary, I. J. (2018). Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. *Nature communications, 9*(1), 2098. doi:10.1038/s41467-018-04362-x

Davies, G., Marioni, R. E., Liewald, D. C., Hill, W. D., Hagenaars, S. P., Harris, S. E., . . . Deary, I. J. (2016). Genome-wide association study of cognitive functions and

educational attainment in UK Biobank (N=112 151). *Mol Psychiatry, 21*(6), 758-767. doi:10.1038/mp.2016.45

De Jager, P. L., Shulman, J. M., Chibnik, L. B., Keenan, B. T., Raj, T., Wilson, R. S., . . . Evans, D. A. (2012). A genome-wide scan for common variants affecting the rate of age-related cognitive decline. *Neurobiol Aging, 33*(5), 1017 e1011-1015. doi:10.1016/j.neurobiolaging.2011.09.033

de Leeuw, C. A., Mooij, J. M., Heskes, T., & Posthuma, D. (2015). MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLOS Computational Biology, 11*(4), e1004219. doi:10.1371/journal.pcbi.1004219

De Schryver, M., Hughes, S., Rosseel, Y., & De Houwer, J. (2016). Unreliable Yet Still Replicable: A Comment on LeBel and Paunonen (2011). *Frontiers in Psychology, 6*(2039). doi:10.3389/fpsyg.2015.02039

Deary, I. J. (2013). Intelligence. *Current Biology, 23*(16), R673-R676. doi:https://doi.org/10.1016/j.cub.2013.07.021

Deary, I. J., Corley, J., Gow, A. J., Harris, S. E., Houlihan, L. M., Marioni, R. E., . . . Starr, J. M. (2009). Age-associated cognitive decline. *British Medical Bulletin, 92*(1), 135-152. doi:10.1093/bmb/ldp033

Deary, I. J., Penke, L., & Johnson, W. (2010). The neuroscience of human intelligence differences. *Nature Reviews Neuroscience, 11*, 201. doi:10.1038/nrn2793

Deary, I. J., & Sternberg, R. J. (2021). Ian Deary and Robert Sternberg answer five self-inflicted questions about human intelligence. *Intelligence, 86*, 101539. doi:https://doi.org/10.1016/j.intell.2021.101539

Deary, I. J., Yang, J., Davies, G., Harris, S. E., Tenesa, A., Liewald, D., . . . Visscher, P. M. (2012). Genetic contributions to stability and change in intelligence from childhood to old age. *Nature, 482*(7384), 212-215. doi:10.1038/nature10781

Demange, P. A., Malanchini, M., Mallard, T. T., Biroli, P., Cox, S. R., Grotzinger, A. D., . . . Nivard, M. G. (2021). Investigating the genetic architecture of noncognitive skills

using GWAS-by-subtraction. *Nature Genetics, 53*(1), 35-44. doi:10.1038/s41588-020-00754-2

Eckert, M. A. (2011). Slowing down: age-related neurobiological predictors of processing speed. *Front Neurosci, 5*, 25. doi:10.3389/fnins.2011.00025

Eckert, M. A., Keren, N. I., Roberts, D. R., Calhoun, V. D., & Harris, K. C. (2010). Age-related changes in processing speed: unique contributions of cerebellar and prefrontal cortex. *Frontiers in Human Neuroscience, 4*, 10-10. doi:10.3389/neuro.09.010.2010

Eichler, E. E. (2019). Genetic Variation, Comparative Genomics, and the Diagnosis of Disease. *New England Journal of Medicine, 381*(1), 64-74. doi:10.1056/NEJMra1809315

Eriksson, N., Macpherson, J. M., Tung, J. Y., Hon, L. S., Naughton, B., Saxonov, S., . . . Mountain, J. (2010). Web-Based, Participant-Driven Studies Yield Novel Genetic Associations for Common Traits. *PLOS Genetics, 6*(6), 1-20. doi:10.1371/journal.pgen.1000993

Erixon-Lindroth, N., Farde, L., Robins Wahlin, T.-B., Sovago, J., Halldin, C., & Bäckman, L. (2005). The role of the striatal dopamine transporter in cognitive aging. *Psychiatry Research: Neuroimaging, 138*(1), 1-12. doi:https://doi.org/10.1016/j.pscychresns.2004.09.005

Evans, I. E. M., Llewellyn, D. J., Matthews, F. E., Woods, R. T., Brayne, C., Clare, L., & on behalf of the, C.-W. r. t. (2018). Social isolation, cognitive reserve, and cognition in healthy older people. *PLoS One, 13*(8), e0201008. doi:10.1371/journal.pone.0201008

Eze, U. C., Bhaduri, A., Haeussler, M., Nowakowski, T. J., & Kriegstein, A. R. (2021). Single-cell atlas of early human brain development highlights heterogeneity of human neuroepithelial cells and early radial glia. *Nature Neuroscience, 24*(4), 584-594. doi:10.1038/s41593-020-00794-1

Fang, J., Wheaton, A. G., Keenan, N. L., Greenlund, K. J., Perry, G. S., & Croft, J. B. (2012). Association of sleep duration and hypertension among US adults varies by age and sex. *Am J Hypertens, 25*(3), 335-341. doi:10.1038/ajh.2011.201

Faul, J., Smith, J., & Zhao, W. (2014). Health and retirement study: candidate gene and SNP data description. *Health Retire. Study, Univ. Mich., Ann Arbor, MI. http://hrsonline. isr. umich. edu/sitedocs/genetics/candidategene/CandidateGeneSNPDataDescription Google Scholar Article Location*.

Fawns-Ritchie, C., & Deary, I. J. (2020). Reliability and validity of the UK Biobank cognitive tests. *PLoS One, 15*(4), e0231627-e0231627. doi:10.1371/journal.pone.0231627

Fernández-Rhodes, L., Gong, J., Haessler, J., Franceschini, N., Graff, M., Nishimura, K. K., . . . North, K. E. (2017). Trans-ethnic fine-mapping of genetic loci for body mass index in the diverse ancestral populations of the Population Architecture using Genomics and Epidemiology (PAGE) Study reveals evidence for multiple signals at established loci. *Hum Genet, 136*(6), 771-800. doi:10.1007/s00439-017-1787-6

Feuk, L., Carson, A. R., & Scherer, S. W. (2006). Structural variation in the human genome. *Nature Reviews Genetics, 7*(2), 85-97. doi:10.1038/nrg1767

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*: sage.

Fitzgerald, J., Fahey, L., Holleran, L., Broin, P. Ó., Donohoe, G., & Morris, D. W. (2021). Thirteen independent genetic loci associated with preserved processing speed in a study of cognitive resilience in 330,097 individuals in the UK Biobank. *bioRxiv*, 2021.2001.2022.427640. doi:10.1101/2021.01.22.427640

Fitzgerald, J., Morris, D. W., & Donohoe, G. (2020). Cognitive Genomics: Recent Advances and Current Challenges. *Curr Psychiatry Rep, 22*(1), 2. doi:10.1007/s11920-019-1125-x

Fletcher, E., Gavett, B., Harvey, D., Farias, S. T., Olichney, J., Beckett, L., . . . Mungas, D. (2018). Brain volume change and cognitive trajectories in aging. *Neuropsychology, 32*(4), 436-449. doi:10.1037/neu0000447

Floresco, S. B. (2015). The nucleus accumbens: an interface between cognition, emotion, and action. *Annu Rev Psychol, 66*, 25-52. doi:10.1146/annurev-psych-010213-115159

Freshour, S. L., Kiwala, S., Cotto, K. C., Coffman, A. C., McMichael, J. F., Song, J. J., . . . Wagner, A. H. (2021). Integration of the Drug–Gene Interaction Database (DGIdb 4.0) with open crowdsource efforts. *Nucleic Acids Research, 49*(D1), D1144-D1151. doi:10.1093/nar/gkaa1084

Fries, G. R. (2020). Polygenic risk scores and their potential clinical use in psychiatry: are we there yet? *Braz J Psychiatry, 42*(5), 459-460. doi:10.1590/1516-4446-2020-0865

Ganna, A., Genovese, G., Howrigan, D. P., Byrnes, A., Kurki, M., Zekavat, S. M., . . . Neale, B. M. (2016). Ultra-rare disruptive and damaging mutations influence educational attainment in the general population. *Nat Neurosci, 19*(12), 1563-1565. doi:10.1038/nn.4404

Gardiner, S. L., Trompet, S., Sabayan, B., Boogaard, M. W., Jukema, J. W., Slagboom, P. E., . . . Aziz, N. A. (2019). Repeat variations in polyglutamine disease–associated genes and cognitive function in old age. *Neurobiology of Aging, 84*, 236.e217-236.e228. doi:https://doi.org/10.1016/j.neurobiolaging.2019.08.002

Geary, D. C. (2019). The Spark of Life and the Unification of Intelligence, Health, and Aging. *Current Directions in Psychological Science, 28*(3), 223-228. doi:10.1177/0963721419829719

Gerlach, J. P., Jordens, I., Tauriello, D. V. F., van ‘t Land-Kuper, I., Bugter, J. M., Noordstra, I., . . . Maurice, M. M. (2018). TMEM59 potentiates Wnt signaling by promoting signalosome formation. *Proceedings of the National Academy of Sciences, 115*(17), E3996. doi:10.1073/pnas.1721321115

Gontier, G., Iyer, M., Shea, J. M., Bieri, G., Wheatley, E. G., Ramalho-Santos, M., & Villeda, S. A. (2018). Tet2 Rescues Age-Related Regenerative Decline and Enhances Cognitive Function in the Adult Mouse Brain. *Cell reports, 22*(8), 1974-1981. doi:10.1016/j.celrep.2018.02.001

González, H. M., Tarraf, W., Harrison, K., Windham, B. G., Tingle, J., Alonso, A., . . . Mosley, T. H. (2018). Midlife cardiovascular health and 20-year cognitive decline: Atherosclerosis Risk in Communities Study results. *Alzheimer's & Dementia, 14*(5), 579-589. doi:https://doi.org/10.1016/j.jalz.2017.11.002

Gordon, R., Smith-Spark, J. H., Newton, E. J., & Henry, L. A. (2018). Executive Function and Academic Achievement in Primary School Children: The Use of Task-Related Processing Speed. *Frontiers in Psychology, 9*(582). doi:10.3389/fpsyg.2018.00582

Grady, C. (2012). The cognitive neuroscience of ageing. *Nature Reviews Neuroscience, 13*(7), 491-505. doi:10.1038/nrn3256

Grotzinger, A. D., Rhemtulla, M., de Vlaming, R., Ritchie, S. J., Mallard, T. T., David, H. W., . . . Tucker-Drob, E. M. (2019). Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nature Human Behaviour, 3*(5), 513-525. doi:http://dx.doi.org/10.1038/s41562-019-0566-x

Guerra-Carrillo, B., Katovich, K., & Bunge, S. A. (2017). Does higher education hone cognitive functioning and learning efficacy? Findings from a large and diverse sample. *PLoS One, 12*(8), e0182276. doi:10.1371/journal.pone.0182276

Gupta, S., Fua, K., Pautler, D., & Farber, I. (2013). *Designing Serious Games for Elders*.

Harada, C. N., Natelson Love, M. C., & Triebel, K. (2013). Normal Cognitive Aging. *Clinics in geriatric medicine, 29*(4), 737-752. doi:10.1016/j.cger.2013.07.002

Harman, M. F., & Martín, M. G. (2020). Epigenetic mechanisms related to cognitive decline during aging. *Journal of Neuroscience Research, 98*(2), 234-246. doi:https://doi.org/10.1002/jnr.24436

Harrison, S. L., Sajjad, A., Bramer, W. M., Ikram, M. A., Tiemeier, H., & Stephan, B. C. (2015). Exploring strategies to operationalize cognitive reserve: A systematic review of reviews. *Journal of clinical and experimental neuropsychology, 37*(3), 253-264.

Hasan, A., & Afzal, M. (2019). Gene and environment interplay in cognition: Evidence from twin and molecular studies, future directions and suggestions for effective candidate gene x environment (cGxE) research. *Multiple Sclerosis and Related Disorders, 33*, 121-130. doi:https://doi.org/10.1016/j.msard.2019.05.005

Haynes, B. I., Bauermeister, S., & Bunce, D. (2017). A Systematic Review of Longitudinal Associations Between Reaction Time Intraindividual Variability and Age-Related

Cognitive Decline or Impairment, Dementia, and Mortality. *J Int Neuropsychol Soc, 23*(5), 431-445. doi:10.1017/s1355617717000236

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods, 50*(3), 1166-1186. doi:10.3758/s13428-017-0935-1

Hedman, A. M., van Haren, N. E., Schnack, H. G., Kahn, R. S., & Hulshoff Pol, H. E. (2012). Human brain changes across the life span: a review of 56 longitudinal magnetic resonance imaging studies. *Hum Brain Mapp, 33*(8), 1987-2002. doi:10.1002/hbm.21334

Henry, A., Katsoulis, M., Masi, S., Fatemifar, G., Denaxas, S., Acosta, D., . . . Dale, C. E. (2019). The relationship between sleep duration, cognition and dementia: a Mendelian randomization study. *International journal of epidemiology, 48*(3), 849-860. doi:10.1093/ije/dyz071

Hill, A., Loh, P.-R., Bharadwaj, R. B., Pons, P., Shang, J., Guinan, E., . . . Jelinsky, S. A. (2017). Stepwise Distributed Open Innovation Contests for Software Development: Acceleration of Genome-Wide Association Analysis. *GigaScience, 6*(5). doi:10.1093/gigascience/gix009

Hill, W. D., Arslan, R. C., Xia, C., Luciano, M., Amador, C., Navarro, P., . . . Penke, L. (2018). Genomic analysis of family data reveals additional genetic effects on intelligence and personality. *Molecular Psychiatry, 23*(12), 2347-2362. doi:10.1038/s41380-017-0005-1

Hill, W. D., Davies, G., Harris, S. E., Hagenaars, S. P., Liewald, D. C., Penke, L., . . . Deary, I. J. (2016). Molecular genetic aetiology of general cognitive function is enriched in evolutionarily conserved regions. *Transl Psychiatry, 6*(12), e980. doi:10.1038/tp.2016.246

Hill, W. D., Marioni, R. E., Maghzian, O., Ritchie, S. J., Hagenaars, S. P., McIntosh, A. M., . . . Deary, I. J. (2018). A combined analysis of genetically correlated traits identifies 187 loci and a role for neurogenesis and myelination in intelligence. *Mol Psychiatry*. doi:10.1038/s41380-017-0001-5

Hipkiss, A. R. (2006). Accumulation of altered proteins and ageing: causes and effects. *Exp Gerontol, 41*(5), 464-473. doi:10.1016/j.exger.2006.03.004

Holland, N., Robbins, T. W., & Rowe, J. B. (2021). The role of noradrenaline in cognition and cognitive disorders. *Brain*. doi:10.1093/brain/awab111

Howard, D. M., Adams, M. J., Clarke, T.-K., Hafferty, J. D., Gibson, J., Shirali, M., . . . Major Depressive Disorder Working Group of the Psychiatric Genomics, C. (2019). Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nature Neuroscience, 22*(3), 343-352. doi:10.1038/s41593-018-0326-7

Huddleston, J., Chaisson, M. J. P., Steinberg, K. M., Warren, W., Hoekzema, K., Gordon, D., . . . Eichler, E. E. (2017). Discovery and genotyping of structural variation from long-read haploid genome sequence data. *Genome Research, 27*(5), 677-685.

Huguet, G., Schramm, C., Douard, E., Jiang, L., Labbe, A., Tihy, F., . . . Consortium, I. (2018). Measuring and Estimating the Effect Sizes of Copy Number Variants on General Intelligence in Community-Based Samples. *JAMA Psychiatry, 75*(5), 447-457. doi:10.1001/jamapsychiatry.2018.0039

Inventor, B. R., & Paun, O. (2021). Pharmacogenomics in Older Adults: An Integrative Review. *Research in Gerontological Nursing*, 1-10. doi:http://dx.doi.org/10.3928/19404921-20210428-01

Jackson, J. J., Hill, P. L., Payne, B. R., Parisi, J. M., & Stine-Morrow, E. A. L. (2020). Linking openness to cognitive ability in older adulthood: The role of activity diversity. *Aging & Mental Health, 24*(7), 1079-1087. doi:10.1080/13607863.2019.1655705

Jackson, T. C., Kotermanski, S. E., & Kochanek, P. M. (2017). Whole-transcriptome microarray analysis reveals regulation of Rab4 by RBM5 in neurons. *Neuroscience, 361*, 93-107. doi:10.1016/j.neuroscience.2017.08.014

Jansen, I. E., Savage, J. E., Watanabe, K., Bryois, J., Williams, D. M., Steinberg, S., . . . Posthuma, D. (2019). Genome-wide meta-analysis identifies new loci and functional

pathways influencing Alzheimer's disease risk. *Nature Genetics, 51*(3), 404-413. doi:10.1038/s41588-018-0311-9

Jia, L., Piña-Crespo, J., & Li, Y. (2019). Restoring Wnt/β-catenin signaling is a promising therapeutic strategy for Alzheimer's disease. *Molecular brain, 12*(1), 104. doi:10.1186/s13041-019-0525-5

Johnson, W., Nijenhuis, J. t., & Bouchard, T. J. (2008). Still just 1 g: Consistent results from five test batteries. *Intelligence, 36*(1), 81-95. doi:https://doi.org/10.1016/j.intell.2007.06.001

Juarez, E. J., Castrellon, J. J., Green, M. A., Crawford, J. L., Seaman, K. L., Smith, C. T., . . . Samanez-Larkin, G. R. (2019). Reproducibility of the correlative triad among aging, dopamine receptor availability, and cognition. *Psychol Aging, 34*(7), 921-932. doi:10.1037/pag0000403

Jung, Y.-S., & Park, J.-I. (2020). Wnt signaling in cancer: therapeutic targeting of Wnt signaling beyond β-catenin and the destruction complex. *Experimental & Molecular Medicine, 52*(2), 183-191. doi:10.1038/s12276-020-0380-6

Kandlur, A., Satyamoorthy, K., & Gangadharan, G. (2020). Oxidative Stress in Cognitive and Epigenetic Aging: A Retrospective Glance. *Frontiers in Molecular Neuroscience, 13*(41). doi:10.3389/fnmol.2020.00041

Karrer, T. M., Josef, A. K., Mata, R., Morris, E. D., & Samanez-Larkin, G. R. (2017). Reduced dopamine receptors and transporters but not synthesis capacity in normal aging adults: a meta-analysis. *Neurobiology of Aging, 57*, 36-46. doi:https://doi.org/10.1016/j.neurobiolaging.2017.05.006

Katzman, R., Terry, R., DeTeresa, R., Brown, T., Davies, P., Fuld, P., . . . Peck, A. (1988). Clinical, pathological, and neurochemical changes in dementia: A subgroup with preserved mental status and numerous neocortical plaques. *Annals of Neurology, 23*(2), 138-144. doi:https://doi.org/10.1002/ana.410230206

Kendall, K. M., Bracher-Smith, M., Fitzpatrick, H., Lynham, A., Rees, E., Escott-Price, V., . . . Kirov, G. (2019). Cognitive performance and functional outcomes of carriers of

pathogenic copy number variants: analysis of the UK Biobank. *The British Journal of Psychiatry, 214*(5), 297-304. doi:10.1192/bjp.2018.301

Kendall, K. M., Rees, E., Escott-Price, V., Einon, M., Thomas, R., Hewitt, J., . . . Kirov, G. (2017). Cognitive Performance Among Carriers of Pathogenic Copy Number Variants: Analysis of 152,000 UK Biobank Subjects. *Biological Psychiatry, 82*(2), 103-110. doi:https://doi.org/10.1016/j.biopsych.2016.08.014

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., & Haussler, D. (2002). The human genome browser at UCSC. *Genome Research, 12*(6), 996-1006. doi:10.1101/gr.229102

Kim, M.-G., Zhang, S., Park, H., Park, S. J., Kim, S., & Chung, C. (2020). Inositol hexakisphosphate kinase-1 is a key mediator of prepulse inhibition and short-term fear memory. *Molecular brain, 13*(1), 72-72. doi:10.1186/s13041-020-00615-3

Kircher, M., Witten, D. M., Jain, P., O'Roak, B. J., Cooper, G. M., & Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics, 46*(3), 310-315. doi:10.1038/ng.2892

Kobayashi, L. C., Glymour, M. M., Kahn, K., Payne, C. F., Wagner, R. G., Montana, L., . . . Berkman, L. F. (2017). Childhood deprivation and later-life cognitive function in a population-based study of older rural South Africans. *Social Science & Medicine, 190*, 20-28. doi:https://doi.org/10.1016/j.socscimed.2017.08.009

Koller, O. M., Hill, N. L., Mogle, J., & Bhang, I. (2019). Relationships Between Subjective Cognitive Impairment and Personality Traits: A Systematic Review. *J Gerontol Nurs, 45*(2), 27-34. doi:10.3928/00989134-20190111-04

Konar, A., Singh, P., & Thakur, M. K. (2016). Age-associated Cognitive Decline: Insights into Molecular Switches and Recovery Avenues. *Aging and Disease, 7*(2), 121-129. doi:10.14336/AD.2015.1004

Kong, A., Thorleifsson, G., Frigge, M. L., Vilhjalmsson, B. J., Young, A. I., Thorgeirsson, T. E., . . . Stefansson, K. (2018). The nature of nurture: Effects of parental genotypes. *Science, 359*(6374), 424. doi:10.1126/science.aan6877

Kuhn, H. G., Toda, T., & Gage, F. H. (2018). Adult Hippocampal Neurogenesis: A Coming-of-Age Story. *The Journal of Neuroscience, 38*(49), 10401. doi:10.1523/JNEUROSCI.2144-18.2018

La Corte, V., Sperduti, M., Malherbe, C., Vialatte, F., Lion, S., Gallarda, T., . . . Piolino, P. (2016). Cognitive Decline and Reorganization of Functional Connectivity in Healthy Aging: The Pivotal Role of the Salience Network in the Prediction of Age and Cognitive Performances. *Frontiers in Aging Neuroscience, 8*(204). doi:10.3389/fnagi.2016.00204

Lähnemann, D., Köster, J., Szczurek, E., McCarthy, D. J., Hicks, S. C., Robinson, M. D., . . . Schönhuth, A. (2020). Eleven grand challenges in single-cell data science. *Genome Biology, 21*(1), 31. doi:10.1186/s13059-020-1926-6

Lam, M., Chen, C. Y., Li, Z., Martin, A. R., Bryois, J., Ma, X., . . . Huang, H. (2019). Comparative genetic architectures of schizophrenia in East Asian and European populations. *Nat Genet, 51*(12), 1670-1678. doi:10.1038/s41588-019-0512-x

Lam, M., Hill, W. D., Trampush, J. W., Yu, J., Knowles, E., Davies, G., . . . Lencz, T. (2019). Pleiotropic Meta-Analysis of Cognition, Education, and Schizophrenia Differentiates Roles of Early Neurodevelopmental and Adult Synaptic Pathways. *The American Journal of Human Genetics, 105*(2), 334-350. doi:10.1016/j.ajhg.2019.06.012

Lastres-Becker, I., Nonis, D., Nowock, J., & Auburger, G. (2019). New alternative splicing variants of the ATXN2 transcript. *Neurological Research and Practice, 1*(1), 22. doi:10.1186/s42466-019-0025-1

Lee, C., & Scherer, S. W. (2010). The clinical context of copy number variation in the human genome. *Expert Rev Mol Med, 12*, e8. doi:10.1017/S1462399410001390

Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., . . . Social Science Genetic Association, C. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics, 50*(8), 1112-1121. doi:10.1038/s41588-018-0147-3

Leritz, E. C., McGlinchey, R. E., Kellison, I., Rudolph, J. L., & Milberg, W. P. (2011). Cardiovascular Disease Risk Factors and Cognition in the Elderly. *Current cardiovascular risk reports, 5*(5), 407-412. doi:10.1007/s12170-011-0189-x

Levitin, D. (2020). *The Changing Mind: A Neuroscientist's Guide to Ageing Well*: Penguin UK.

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., & Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell systems, 1*(6), 417-425. doi:10.1016/j.cels.2015.12.004

Lievre, A., Alley, D., & Crimmins, E. M. (2008). Educational differentials in life expectancy with cognitive impairment among the elderly in the United States. *J Aging Health, 20*(4), 456-477. doi:10.1177/0898264308315857

Lindenberger, U., Nagel, I., Chicherio, C., Li, S.-C., Heekeren, H., & Bäckman, L. (2008). Age-related decline in brain resources modulates genetic effects on cognitive functioning. *Frontiers in Neuroscience, 2*(39). doi:10.3389/neuro.01.039.2008

Linghui, D., Shi, Q., Chi, C., Xiaolei, L., Lixing, Z., Zhiliang, Z., & Birong, D. (2020). The Association Between Leukocyte Telomere Length and Cognitive Performance Among the American Elderly. *Frontiers in Aging Neuroscience, 12*(354). doi:10.3389/fnagi.2020.527658

Lisman, J., Buzsáki, G., Eichenbaum, H., Nadel, L., Ranganath, C., & Redish, A. D. (2017). Viewpoints: how the hippocampus contributes to memory, navigation and cognition. *Nat Neurosci, 20*(11), 1434-1447. doi:10.1038/nn.4661

Litwin, H., & Stoeckel, K. J. (2015). Social Network, Activity Participation, and Cognition: A Complex Relationship. *Research on Aging, 38*(1), 76-97. doi:10.1177/0164027515581422

Liu, H., Yang, Y., Xia, Y., Zhu, W., Leak, R. K., Wei, Z., . . . Hu, X. (2017). Aging of cerebral white matter. *Ageing Research Reviews, 34*, 64-76. doi:https://doi.org/10.1016/j.arr.2016.11.006

Lövdén, M., Fratiglioni, L., Glymour, M. M., Lindenberger, U., & Tucker-Drob, E. M. (2020). Education and Cognitive Functioning Across the Life Span. *Psychological Science in the Public Interest, 21*(1), 6-41. doi:10.1177/1529100620920576

Low, D. V., Wu, M. N., & Spira, A. P. (2019). Sleep Duration and Cognition in a Nationally Representative Sample of U.S. Older Adults. *The American Journal of Geriatric Psychiatry, 27*(12), 1386-1396. doi:https://doi.org/10.1016/j.jagp.2019.07.001

Luciano, M., Hagenaars, S. P., Davies, G., Hill, W. D., Clarke, T.-K., Shirali, M., . . . Deary, I. J. (2018). Association analysis in over 329,000 individuals identifies 116 independent variants influencing neuroticism. *Nature Genetics, 50*(1), 6-11. doi:10.1038/s41588-017-0013-8

Lyall, D. M., Cullen, B., Allerhand, M., Smith, D. J., Mackay, D., Evans, J., . . . Pell, J. P. (2016). Cognitive Test Scores in UK Biobank: Data Reduction in 480,416 Participants and Longitudinal Stability in 20,346 Participants. *PLoS One, 11*(4), e0154222. doi:10.1371/journal.pone.0154222

Lyall, L. M., Wyse, C. A., Graham, N., Ferguson, A., Lyall, D. M., Cullen, B., . . . Smith, D. J. (2018). Association of disrupted circadian rhythmicity with mood disorders, subjective wellbeing, and cognitive function: a cross-sectional study of 91 105 participants from the UK Biobank. *The Lancet Psychiatry*. doi:10.1016/s2215-0366(18)30139-1

Madden, D. J., Bennett, I. J., Burzynska, A., Potter, G. G., Chen, N.-k., & Song, A. W. (2012). Diffusion tensor imaging of cerebral white matter integrity in cognitive aging. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease, 1822*(3), 386-400. doi:https://doi.org/10.1016/j.bbadis.2011.08.003

Malek-Ahmadi, M., Lu, S., Chan, Y., Perez, S. E., Chen, K., & Mufson, E. J. (2017). Static and Dynamic Cognitive Reserve Proxy Measures: Interactions with Alzheimer's Disease Neuropathology and Cognition. *J Alzheimers Dis Parkinsonism, 7*(6). doi:10.4172/2161-0460.1000390

Malik, R., Chauhan, G., Traylor, M., Sargurupremraj, M., Okada, Y., Mishra, A., . . . Consortium, M. (2018). Multiancestry genome-wide association study of 520,000

subjects identifies 32 loci associated with stroke and stroke subtypes. *Nature Genetics, 50*(4), 524-537. doi:10.1038/s41588-018-0058-3

Malinowski, P., & Shalamanova, L. (2017). Meditation and Cognitive Ageing: the Role of Mindfulness Meditation in Building Cognitive Reserve. *Journal of Cognitive Enhancement, 1*(2), 96-106. doi:10.1007/s41465-017-0022-7

Mallard, T. T., Linnér, R. K., Grotzinger, A. D., Sanchez-Roige, S., Seidlitz, J., Okbay, A., . . . Harden, K. P. (2020). Multivariate GWAS of psychiatric disorders and their cardinal symptoms reveal two dimensions of cross-cutting genetic liabilities. *bioRxiv*, 603134. doi:10.1101/603134

Manning, K. J., Chan, G., & Steffens, D. C. (2017). Neuroticism Traits Selectively Impact Long Term Illness Course and Cognitive Decline in Late-Life Depression. *Am J Geriatr Psychiatry, 25*(3), 220-229. doi:10.1016/j.jagp.2016.10.006

Marques, P., Moreira, P., Magalhaes, R., Costa, P., Santos, N., Zihl, J., . . . Sousa, N. (2016). The functional connectome of cognitive reserve. *Hum Brain Mapp, 37*(9), 3310-3322. doi:10.1002/hbm.23242

Martin, A. R., Daly, M. J., Robinson, E. B., Hyman, S. E., & Neale, B. M. (2019). Predicting Polygenic Risk of Psychiatric Disorders. *Biological Psychiatry, 86*(2), 97-109. doi:https://doi.org/10.1016/j.biopsych.2018.12.015

Martin, A. R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B. M., & Daly, M. J. (2019). Clinical use of current polygenic risk scores may exacerbate health disparities. *Nature Genetics, 51*(4), 584-591. doi:10.1038/s41588-019-0379-x

McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A. R., Teumer, A., . . . Haplotype Reference, C. (2016). A reference panel of 64,976 haplotypes for genotype imputation. *Nature Genetics, 48*(10), 1279-1283. doi:10.1038/ng.3643

McClellan, J. M., Lehner, T., & King, M.-C. (2017). Gene Discovery for Complex Traits: Lessons from Africa. *Cell, 171*(2), 261-264. doi:https://doi.org/10.1016/j.cell.2017.09.037

McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., . . . Cunningham, F. (2016). The Ensembl Variant Effect Predictor. *Genome Biology, 17*(1), 122. doi:10.1186/s13059-016-0974-4

Medaglia, J. D., Pasqualetti, F., Hamilton, R. H., Thompson-Schill, S. L., & Bassett, D. S. (2017). Brain and cognitive reserve: Translation via network control theory. *Neurosci Biobehav Rev, 75*, 53-64. doi:10.1016/j.neubiorev.2017.01.016

Melzer, D., Pilling, L. C., & Ferrucci, L. (2020). The genetics of human ageing. *Nature Reviews Genetics, 21*(2), 88-101. doi:10.1038/s41576-019-0183-6

Mise, A., Yoshino, Y., Yamazaki, K., Ozaki, Y., Sao, T., Yoshida, T., . . . Ueno, S.-i. (2017). TOMM40 and APOE Gene Expression and Cognitive Decline in Japanese Alzheimer's Disease Subjects. *Journal of Alzheimer's Disease, 60*, 1107-1117. doi:10.3233/JAD-170361

Mitteldorf, J., & Sagan, D. (2016). *Cracking the Aging Code: The New Science of Growing Old-and What It Means for Staying Young*: Macmillan.

Mullins, N., Forstner, A. J., O'Connell, K. S., Coombes, B., Coleman, J. R. I., Qiao, Z., . . . Psychiatry, H. A.-I. (2021). Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nature Genetics, 53*(6), 817-829. doi:10.1038/s41588-021-00857-4

Nagel, M., Jansen, P. R., Stringer, S., Watanabe, K., de Leeuw, C. A., Bryois, J., . . . andMe Research, T. (2018). Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. *Nature Genetics, 50*(7), 920-927. doi:10.1038/s41588-018-0151-7

Nalls, M. A., Blauwendraat, C., Vallerga, C. L., Heilbron, K., Bandres-Ciga, S., Chang, D., . . . Zhang, F. (2019). Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *The Lancet Neurology, 18*(12), 1091-1102. doi:10.1016/S1474-4422(19)30320-5

Nave, G., Jung, W. H., Karlsson Linnér, R., Kable, J. W., & Koellinger, P. D. (2019). Are Bigger Brains Smarter? Evidence From a Large-Scale Preregistered Study. *Psychological science, 30*(1), 43-54. doi:10.1177/0956797618808470

Nilsson, M. I., & Tarnopolsky, M. A. (2019). Mitochondria and Aging—The Role of Exercise as a Countermeasure. *Biology, 8*(2), 40.

Nivard, M. (2019). GenomicSem. Retrieved from https://github.com/MichelNivard/GenomicSEM/wiki

Nowakowska, B. (2017). Clinical interpretation of copy number variants in the human genome. *Journal of applied genetics, 58*(4), 449-457. doi:10.1007/s13353-017-0407-4

Okbay, A., Beauchamp, J. P., Fontana, M. A., Lee, J. J., Pers, T. H., Rietveld, C. A., . . . Benjamin, D. J. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature, 533*(7604), 539-542. doi:10.1038/nature17671

Orellana, C., Ferreira, D., Muehlboeck, J. S., Mecocci, P., Vellas, B., Tsolaki, M., . . . Westman, E. (2016). Measuring Global Brain Atrophy with the Brain Volume/Cerebrospinal Fluid Index: Normative Values, Cut-Offs and Clinical Associations. *Neurodegener Dis, 16*(1-2), 77-86. doi:10.1159/000442443

Palomer, E., Buechler, J., & Salinas, P. C. (2019). Wnt Signaling Deregulation in the Aging and Alzheimer's Brain. *Frontiers in Cellular Neuroscience, 13*(227). doi:10.3389/fncel.2019.00227

Panzanelli, P., Früh, S., & Fritschy, J.-M. (2017). Differential role of GABAA receptors and neuroligin 2 for perisomatic GABAergic synapse formation in the hippocampus. *Brain Structure and Function, 222*(9), 4149-4161. doi:10.1007/s00429-017-1462-7

Papenberg, G., Lindenberger, U., & Bäckman, L. (2015). Aging-related magnification of genetic effects on cognitive and brain integrity. *Trends in Cognitive Sciences, 19*(9), 506-514. doi:https://doi.org/10.1016/j.tics.2015.06.008

Pardiñas, A. F., Holmans, P., Pocklington, A. J., Escott-Price, V., Ripke, S., Carrera, N., . . . Walters, J. T. R. (2018). Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nature Genetics, 50*(3), 381-389. doi:10.1038/s41588-018-0059-2

Park, D. C., & Reuter-Lorenz, P. (2008). The Adaptive Brain: Aging and Neurocognitive Scaffolding. *Annu Rev Psychol, 60*(1), 173-196. doi:10.1146/annurev.psych.59.103006.093656

Pausova, Z., Paus, T., Abrahamowicz, M., Bernard, M., Gaudet, D., Leonard, G., . . . Veillette, S. (2017). Cohort Profile: The Saguenay Youth Study (SYS). *International journal of epidemiology, 46*(2), e19-e19. doi:10.1093/ije/dyw023

Pereira Fernandes, D., Bitar, M., Jacobs, F. M. J., & Barry, G. (2018). Long Non-Coding RNAs in Neuronal Aging. *Non-Coding RNA, 4*(2), 12.

Petersen, R. C., Doody, R., Kurz, A., Mohs, R. C., Morris, J. C., Rabins, P. V., . . . Winblad, B. (2001). Current Concepts in Mild Cognitive Impairment. *Archives of Neurology, 58*(12), 1985-1992. doi:10.1001/archneur.58.12.1985

Plassman, B. L., Williams, J. W., Jr., Burke, J. R., Holsinger, T., & Benjamin, S. (2010). Systematic Review: Factors Associated With Risk for and Possible Prevention of Cognitive Decline in Later Life. *Annals of Internal Medicine, 153*(3), 182-193. doi:10.7326/0003-4819-153-3-201008030-00258

Plomin, R., & Deary, I. J. (2014). Genetics and intelligence differences: five special findings. *Molecular Psychiatry, 20*, 98. doi:10.1038/mp.2014.105

Plomin, R., DeFries, J. C., McClearn, G. E., & McGuffin, P. . (2001). *Behavioral genetics* (4th ed.). New York: Worth Publishers.

Plomin, R., & Spinath, F. (2004). Intelligence: Genetics, Genes, and Genomics. *Journal of personality and social psychology, 86*, 112-129. doi:10.1037/0022-3514.86.1.112

Plomin, R., & von Stumm, S. (2018). The new genetics of intelligence. *Nature Reviews Genetics, 19*, 148. doi:10.1038/nrg.2017.104

Ramus, F. (2006). Genes, brain, and cognition: A roadmap for the cognitive scientist. *Cognition, 101*(2), 247-269. doi:https://doi.org/10.1016/j.cognition.2006.04.003

Rea, I. M. (2017). Towards ageing well: Use it or lose it: Exercise, epigenetics and cognition. *Biogerontology, 18*(4), 679-691. doi:10.1007/s10522-017-9719-3

Reed, B. R., Dowling, M., Tomaszewski Farias, S., Sonnen, J., Strauss, M., Schneider, J. A., . . . Mungas, D. (2011). Cognitive activities during adulthood are more important than education in building reserve. *Journal of the International Neuropsychological Society : JINS, 17*(4), 615-624. doi:10.1017/S1355617711000014

Reed, B. R., Mungas, D., Farias, S. T., Harvey, D., Beckett, L., Widaman, K., . . . DeCarli, C. (2010). Measuring cognitive reserve based on the decomposition of episodic memory variance. *Brain, 133*(Pt 8), 2196-2209. doi:10.1093/brain/awq154

Regev, A., Teichmann, S., Rozenblatt-Rosen, O., Stubbington, M., Ardlie, K., Amit, I., . . . Committee, H. (2018). *The Human Cell Atlas White Paper*.

Reynolds, C. A., & Finkel, D. (2015). A meta-analysis of heritability of cognitive aging: minding the "missing heritability" gap. *Neuropsychol Rev, 25*(1), 97-112. doi:10.1007/s11065-015-9280-2

Rietveld, C. A., Esko, T., Davies, G., Pers, T. H., Turley, P., Benyamin, B., . . . Koellinger, P. D. (2014). Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. *Proceedings of the National Academy of Sciences of the United States of America, 111*(38), 13790-13794. doi:10.1073/pnas.1404623111

Ripke, S., Walters, J. T. R., & O'Donovan, M. C. (2020). Mapping genomic loci prioritises genes and implicates synaptic biology in schizophrenia. *medRxiv*, 2020.2009.2012.20192922. doi:10.1101/2020.09.12.20192922

Ritchie, S. J., Hill, W. D., Marioni, R. E., Davies, G., Hagenaars, S. P., Harris, S. E., . . . Deary, I. J. (2019). Polygenic predictors of age-related decline in cognitive ability. *Molecular Psychiatry*. doi:10.1038/s41380-019-0372-x

Robertson, I. H. (2013). A noradrenergic theory of cognitive reserve: implications for Alzheimer's disease. *Neurobiol Aging, 34*(1), 298-308. doi:10.1016/j.neurobiolaging.2012.05.019

Robertson, I. H. (2014). Right hemisphere role in cognitive reserve. *Neurobiol Aging, 35*(6), 1375-1385. doi:10.1016/j.neurobiolaging.2013.11.028

Royle, N. A., Booth, T., Valdes Hernandez, M. C., Penke, L., Murray, C., Gow, A. J., . . . Wardlaw, J. M. (2013). Estimated maximal and current brain volume predict cognitive ability in old age. *Neurobiol Aging, 34*(12), 2726-2733. doi:10.1016/j.neurobiolaging.2013.05.015

Rozycka, A., & Liguz-Lecznar, M. (2017). The space where aging acts: focus on the GABAergic synapse. *Aging cell, 16*(4), 634-643. doi:10.1111/acel.12605

Salthouse, T. (1996). The processing-speed theory of adult age differences in cognition. *Psychological review, 103 3*, 403-428.

Salthouse, T. A. (2015). Implications of the Flynn Effect for Age-Cognition Relations. *Intelligence, 48*, 51-57. doi:10.1016/j.intell.2014.10.007

Salthouse, T. A., & Ferrer-Caja, E. (2003). What needs to be explained to account for age-related effects on multiple cognitive variables? *Psychology and aging, 18*(1), 91.

Samrani, G., Bäckman, L., & Persson, J. (2017). Age-differences in the temporal properties of proactive interference in working memory. *Psychology and aging, 32*(8), 722.

Savage, J. E., Jansen, P. R., Stringer, S., Watanabe, K., Bryois, J., de Leeuw, C. A., . . . Posthuma, D. (2018). Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nature Genetics, 50*(7), 912-919. doi:10.1038/s41588-018-0152-6

Schacter, D. L. (2019). Implicit Memory, Constructive Memory, and Imagining the Future: A Career Perspective. *Perspectives on Psychological Science, 14*(2), 256-272. doi:10.1177/1745691618803640

Scheiber, C., Chen, H., Kaufman, A. S., & Weiss, L. G. (2017). How Much Does WAIS-IV Perceptual Reasoning Decline Across the 20 to 90-Year Lifespan When Processing Speed is Controlled? *Applied Neuropsychology: Adult, 24*(2), 116-131. doi:10.1080/23279095.2015.1107564

Schubert, A.-L., Nunez, M. D., Hagemann, D., & Vandekerckhove, J. (2019). Individual Differences in Cortical Processing Speed Predict Cognitive Abilities: a Model-Based

Cognitive Neuroscience Account. *Computational Brain & Behavior, 2*(2), 64-84. doi:10.1007/s42113-018-0021-5

Schumann, G., Loth, E., Banaschewski, T., Barbot, A., Barker, G., Büchel, C., . . . the, I. c. (2010). The IMAGEN study: reinforcement-related behaviour in normal brain function and psychopathology. *Molecular Psychiatry, 15*(12), 1128-1139. doi:10.1038/mp.2010.4

Schwartz, C. E., Rapkin, B. D., & Healy, B. C. (2016). Reserve and Reserve-building activities research: key challenges and future directions. *BMC Neurosci, 17*(1), 62. doi:10.1186/s12868-016-0297-0

Shafee, R., Nanda, P., Padmanabhan, J. L., Tandon, N., Alliey-Rodriguez, N., Kalapurakkel, S., . . . Robinson, E. B. (2018). Polygenic risk for schizophrenia and measured domains of cognition in individuals with psychosis and controls. *Translational Psychiatry, 8*(1), 78. doi:10.1038/s41398-018-0124-8

Shao, Y., Yan, G., Xuan, Y., Peng, H., Huang, Q.-J., Wu, R., & Xu, H. (2015). Chronic social isolation decreases glutamate and glutamine levels and induces oxidative stress in the rat hippocampus. *Behavioural brain research, 282*, 201-208.

Singh, T., Neale, B. M., & Daly, M. J. (2020). Exome sequencing identifies rare coding variants in 10 genes which confer substantial risk for schizophrenia. *medRxiv*, 2020.2009.2018.20192815. doi:10.1101/2020.09.18.20192815

Smith, B. H., Campbell, H., Blackwood, D., Connell, J., Connor, M., Deary, I. J., . . . Morris, A. D. (2006). Generation Scotland: the Scottish Family Health Study; a new resource for researching genes and heritability. *BMC Medical Genetics, 7*(1), 74. doi:10.1186/1471-2350-7-74

Smith, S. M., Douaud, G., Chen, W., Hanayik, T., Alfaro-Almagro, F., Sharp, K., & Elliott, L. T. (2020). Enhanced Brain Imaging Genetics in UK Biobank. *bioRxiv*, 2020.2007.2027.223545. doi:10.1101/2020.07.27.223545

Sniekers, S., Stringer, S., Watanabe, K., Jansen, P. R., Coleman, J. R. I., Krapohl, E., . . . Posthuma, D. (2017). Genome-wide association meta-analysis of 78,308 individuals

identifies new loci and genes influencing human intelligence. *Nat Genet, 49*(7), 1107-1112. doi:10.1038/ng.3869

Sonnega, A., Faul, J. D., Ofstedal, M. B., Langa, K. M., Phillips, J. W. R., & Weir, D. R. (2014). Cohort Profile: the Health and Retirement Study (HRS). *International journal of epidemiology, 43*(2), 576-585. doi:10.1093/ije/dyu067

Staal, M. A., Bolton, A., Yaroush, R., & Bourne Jr, L. (2008). Cognitive performance and resilience to stress. *Biobehavioral resilience to stress*, 259-299.

Stahl, E. A., Breen, G., Forstner, A. J., McQuillin, A., Ripke, S., Trubetskoy, V., . . . the Bipolar Disorder Working Group of the Psychiatric Genomics, C. (2019). Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nature Genetics, 51*(5), 793-803. doi:10.1038/s41588-019-0397-8

Stefansson, H., Meyer-Lindenberg, A., Steinberg, S., Magnusdottir, B., Morgen, K., Arnarsdottir, S., . . . Stefansson, K. (2013). CNVs conferring risk of autism or schizophrenia affect cognition in controls. *Nature, 505*, 361. doi:10.1038/nature12818

https://www.nature.com/articles/nature12818#supplementary-information

Stefansson, H., Meyer-Lindenberg, A., Steinberg, S., Magnusdottir, B., Morgen, K., Arnarsdottir, S., . . . Stefansson, K. (2014). CNVs conferring risk of autism or schizophrenia affect cognition in controls. *Nature, 505*(7483), 361-366. doi:10.1038/nature12818

Steptoe, A., Breeze, E., Banks, J., & Nazroo, J. (2013). Cohort Profile: The English Longitudinal Study of Ageing. *International journal of epidemiology, 42*(6), 1640-1648. doi:10.1093/ije/dys168

Stern, Y. (2002). What is cognitive reserve? Theory and research application of the reserve concept. *Journal of the International Neuropsychological Society, 8*(3), 448-460. doi:10.1017/S1355617702813248

Stern, Y. (2012). Cognitive reserve in ageing and Alzheimer's disease. *Lancet neurology, 11*(11), 1006-1012. doi:10.1016/S1474-4422(12)70191-6

Stern, Y. (2017). An approach to studying the neural correlates of reserve. *Brain Imaging Behav, 11*(2), 410-416. doi:10.1007/s11682-016-9566-x

Stern, Y., Arenaza-Urquijo, E. M., Bartrés-Faz, D., Belleville, S., Cantilon, M., Chetelat, G., . . . Vuoksimaa, E. (2018). Whitepaper: Defining and investigating cognitive reserve, brain reserve, and brain maintenance. *Alzheimer's & Dementia*. doi:https://doi.org/10.1016/j.jalz.2018.07.219

Stillman, C. M., Weinstein, A. M., Marsland, A. L., Gianaros, P. J., & Erickson, K. I. (2017). Body-Brain Connections: The Effects of Obesity and Behavioral Interventions on Neurocognitive Aging. *Frontiers in Aging Neuroscience, 9*, 115-115. doi:10.3389/fnagi.2017.00115

Strittmatter, A., Sunde, U., & Zegners, D. (2020). Life cycle patterns of cognitive performance over the long run. *Proceedings of the National Academy of Sciences, 117*(44), 27255. doi:10.1073/pnas.2006653117

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., . . . Collins, R. (2015). UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med, 12*(3), e1001779. doi:10.1371/journal.pmed.1001779

Suh, S. W., Han, J. W., Lee, J. R., Byun, S., Kwon, S. J., Oh, S. H., . . . Kim, K. W. (2018). Sleep and cognitive decline: A prospective nondemented elderly cohort study. *Annals of Neurology, 83*(3), 472-482. doi:https://doi.org/10.1002/ana.25166

Tangestani Fard, M., & Stough, C. (2019). A Review and Hypothesized Model of the Mechanisms That Underpin the Relationship Between Inflammation and Cognition in the Elderly. *Frontiers in Aging Neuroscience, 11*(56). doi:10.3389/fnagi.2019.00056

Tasaki, S., Gaiteri, C., Mostafavi, S., Yu, L., Wang, Y., De Jager, P. L., & Bennett, D. A. (2018). Multi-omic Directed Networks Describe Features of Gene Regulation in Aged Brains and Expand the Set of Genes Driving Cognitive Decline. *Frontiers in genetics, 9*, 294-294. doi:10.3389/fgene.2018.00294

Taylor, A. M., Pattie, A., & Deary, I. J. (2018). Cohort Profile Update: The Lothian Birth Cohorts of 1921 and 1936. *International journal of epidemiology, 47*(4), 1042-1042r. doi:10.1093/ije/dyy022

Tikhomirova, T., Malykh, A., & Malykh, S. (2020). Predicting Academic Achievement with Cognitive Abilities: Cross-Sectional Study across School Education. *Behavioral Sciences, 10*(10), 158.

Timmers, P. R., Mounier, N., Lall, K., Fischer, K., Ning, Z., Feng, X., . . . Joshi, P. K. (2019). Genomics of 1 million parent lifespans implicates novel pathways and common diseases and distinguishes survival chances. *eLife, 8*, e39856. doi:10.7554/eLife.39856

Toepper, M. (2017). Dissociating Normal Aging from Alzheimer's Disease: A View from Cognitive Neuroscience. *J Alzheimers Dis, 57*(2), 331-352. doi:10.3233/JAD-161099

Trampush, J. W., Yang, M. L., Yu, J., Knowles, E., Davies, G., Liewald, D. C., . . . Lencz, T. (2017). GWAS meta-analysis reveals novel loci and genetic correlates for general cognitive function: a report from the COGENT consortium. *Mol Psychiatry, 22*(3), 336-345. doi:10.1038/mp.2016.244

Turley, P., Walters, R. K., Maghzian, O., Okbay, A., Lee, J. J., Fontana, M. A., . . . Social Science Genetic Association, C. (2018). Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nature Genetics, 50*(2), 229-237. doi:10.1038/s41588-017-0009-4

van der Wijst, M. G. P., de Vries, D. H., Groot, H. E., Trynka, G., Hon, C. C., Bonder, M. J., . . . Franke, L. (2020). The single-cell eQTLGen consortium. *eLife, 9*, e52155. doi:10.7554/eLife.52155

Van Hout, C. V., Tachmazidou, I., Backman, J. D., Hoffman, J. X., Ye, B., Pandey, A. K., . . . Baras, A. (2019). Whole exome sequencing and characterization of coding variation in 49,960 individuals in the UK Biobank. *bioRxiv*, 572347. doi:10.1101/572347

van Rheenen, W., Shatunov, A., Dekker, A. M., McLaughlin, R. L., Diekstra, F. P., Pulit, S. L., . . . Veldink, J. H. (2016). Genome-wide association analyses identify new risk

variants and the genetic architecture of amyotrophic lateral sclerosis. *Nature Genetics, 48*(9), 1043-1048. doi:10.1038/ng.3622

Vanherle, L., Matuskova, H., Don-Doncow, N., Uhl, F. E., & Meissner, A. (2020). Improving Cerebrovascular Function to Increase Neuronal Recovery in Neurodegeneration Associated to Cardiovascular Disease. *Frontiers in Cell and Developmental Biology, 8*(53). doi:10.3389/fcell.2020.00053

Vecchio, L. M., Meng, Y., Xhima, K., Lipsman, N., Hamani, C., & Aubert, I. (2018). The Neuroprotective Effects of Exercise: Maintaining a Healthy Brain Throughout Aging. *Brain Plasticity, 4*, 17-52. doi:10.3233/BPL-180069

Wang, D., Liu, S., Warrell, J., Won, H., Shi, X., Navarro, F. C. P., . . . Gerstein, M. B. (2018). Comprehensive functional genomic resource and integrative model for the human brain. *Science, 362*(6420). doi:10.1126/science.aat8464

Wang, P., Geng, J., Gao, J., Zhao, H., Li, J., Shi, Y., . . . Chen, L. (2019). Macrophage achieves self-protection against oxidative stress-induced ageing through the Mst-Nrf2 axis. *Nature communications, 10*(1), 755-755. doi:10.1038/s41467-019-08680-6

Ward, D. D., Summers, M. J., Saunders, N. L., & Vickers, J. C. (2015). Modeling cognitive reserve in healthy middle-aged and older adults: the Tasmanian Healthy Brain Project. *International Psychogeriatrics, 27*(4), 579-589. doi:http://dx.doi.org/10.1017/S1041610214002075

Warland, A., Kendall, K. M., Rees, E., Kirov, G., & Caseras, X. (2019). Schizophrenia-associated genomic copy number variants and subcortical brain volumes in the UK Biobank. *Molecular Psychiatry*. doi:10.1038/s41380-019-0355-y

Warne, R. T., & Burningham, C. (2019). Spearman's g found in 31 non-Western nations: Strong evidence that g is a universal phenomenon. *Psychological Bulletin, 145*(3), 237-272. doi:10.1037/bul0000184

Watanabe, K., Taskesen, E., van Bochoven, A., & Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. *Nature communications, 8*(1), 1826. doi:10.1038/s41467-017-01261-5

Watanabe, K., Umićević Mirkov, M., de Leeuw, C. A., van den Heuvel, M. P., & Posthuma, D. (2019). Genetic mapping of cell type specificity for complex traits. *Nature communications, 10*(1), 3222. doi:10.1038/s41467-019-11181-1

Welsh, K. A., Breitner, J. C. S., & Magruder-Habib, K. M. (1993). Detection of dementia in the elderly using telephone screening of cognitive status. *Cognitive and Behavioral Neurology, 6*(2), 103-110.

Westra, H.-J., Peters, M. J., Esko, T., Yaghootkar, H., Schurmann, C., Kettunen, J., . . . Franke, L. (2013). Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nature Genetics, 45*(10), 1238-1243. doi:10.1038/ng.2756

Wingo, A. P., Dammer, E. B., Breen, M. S., Logsdon, B. A., Duong, D. M., Troncosco, J. C., . . . Wingo, T. S. (2019). Large-scale proteomic analysis of human brain identifies proteins associated with cognitive trajectory in advanced age. *Nature communications, 10*(1), 1619. doi:10.1038/s41467-019-09613-z

World Health, O. (2015). *World report on ageing and health*. Geneva: World Health Organization.

Wu, N. N., & Ren, J. (2019). Aldehyde Dehydrogenase 2 (ALDH2) and Aging: Is There a Sensible Link? *Adv Exp Med Biol, 1193*, 237-253. doi:10.1007/978-981-13-6260-6_15

Xia, C., Amador, C., Huffman, J., Trochet, H., Campbell, A., Porteous, D., . . . Haley, C. S. (2016). Pedigree- and SNP-Associated Genetics and Recent Environment are the Major Contributors to Anthropometric and Cardiometabolic Trait Variation. *PLOS Genetics, 12*(2), e1005804. doi:10.1371/journal.pgen.1005804

Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet, 88*(1), 76-82. doi:10.1016/j.ajhg.2010.11.011

Zahodne, L. B., Manly, J. J., Brickman, A. M., Narkhede, A., Griffith, E. Y., Guzman, V. A., . . . Stern, Y. (2015). Is residual memory variance a valid method for quantifying

cognitive reserve? A longitudinal application. *Neuropsychologia, 77*, 260-266. doi:10.1016/j.neuropsychologia.2015.09.009

Zaitlen, N., Kraft, P., Patterson, N., Pasaniuc, B., Bhatia, G., Pollack, S., & Price, A. L. (2013). Using Extended Genealogy to Estimate Components of Heritability for 23 Quantitative and Dichotomous Traits. *PLOS Genetics, 9*(5), e1003520. doi:10.1371/journal.pgen.1003520

Zeisel, A., Hochgerner, H., Lönnerberg, P., Johnsson, A., Memic, F., van der Zwan, J., . . . Linnarsson, S. (2018). Molecular Architecture of the Mouse Nervous System. *Cell, 174*(4), 999-1014.e1022. doi:10.1016/j.cell.2018.06.021

Zhan, Y., Clements, M. S., Roberts, R. O., Vassilaki, M., Druliner, B. R., Boardman, L. A., . . . Hägg, S. (2018). Association of telomere length with general cognitive trajectories: a meta-analysis of four prospective cohort studies. *Neurobiology of Aging, 69*, 111-116. doi:https://doi.org/10.1016/j.neurobiolaging.2018.05.004

Zhang, C., & Pierce, B. L. (2014). Genetic susceptibility to accelerated cognitive decline in the US Health and Retirement Study. *Neurobiol Aging, 35*(6), 1512 e1511-1518. doi:10.1016/j.neurobiolaging.2013.12.021

Zhang, J., Rane, G., Dai, X., Shanmugam, M. K., Arfuso, F., Samy, R. P., . . . Sethi, G. (2016). Ageing and the telomere connection: An intimate relationship with inflammation. *Ageing Res Rev, 25*, 55-69. doi:10.1016/j.arr.2015.11.006

Zhang, W., Qu, J., Liu, G.-H., & Belmonte, J. C. I. (2020). The ageing epigenome and its rejuvenation. *Nature Reviews Molecular Cell Biology, 21*(3), 137-150. doi:10.1038/s41580-019-0204-5

Zhong, S., Zhang, S., Fan, X., Wu, Q., Yan, L., Dong, J., . . . Wang, X. (2018). A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. *Nature, 555*(7697), 524-528. doi:10.1038/nature25980

Zhu, Z., Zheng, Z., Zhang, F., Wu, Y., Trzaskowski, M., Maier, R., . . . Yang, J. (2018). Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nature communications, 9*(1), 224. doi:10.1038/s41467-017-02317-2

### Oral presentations

November 2018, Threeses competition, NUIG, “What role do our genes play in cognitive resilience”.

December 2018, Galway Neuroscience centre (GNC) “ Investigating the genetics of cognitive resilience in healthy ageing”.

April 2019, British Neuroscience Association, “ Investigating the genetics of cognitive resilience in healthy ageing*”.

September 2019, Irish Society of Human Genetics ( same title as above*).

December 2019, GNC (same title as above*)

May 2020, Center for Chromosome Biology, NUIG ( same title as above*).

September 2020, Irish Society of Human Genetics ( same title as above*).

May 2021, Big Data in Biomedical Research - Using the UKB in Ireland. “Identification of 13 genetic loci associated with cognitive resilience in the UK Biobank (N=330,0098)”.

### Poster presentations

April 2019, British Neurpscience Association “ Investigating the genetics of cognitive resilience in healthy ageing”.

April 2021, Welcome Trust, Genomics of Brain Disorders conference “ Identification of 13 independent genetic loci associated with cognitive resilience in healthy aging in 330,097 individuals in the UK Biobank”.

### Awards

December 2018, Runner up , best rapid fire oral presentation at GNC research day.

September 2019, Irish Society of Human Genetics Young Investigator Award for Best Postgraduate Oral Presentation (Belfast). The award included a scholarship to the European Society of Human genetics in 2020.

May 2021, Runner up, best oral presentation, Big Data in Biomedical Research - Using the UKB in Ireland.

## Publications

Fitzgerald, J., Morris, D.W. & Donohoe, G. Cognitive Genomics: Recent Advances and Current Challenges. Curr Psychiatry Rep 22, 2 (2020). https://doi.org/10.1007/s11920-019-1125-x

Draft of paper "Identification of 13 independent genetic loci associated with cognitive resilience in healthy aging in 330,097 individuals in the UK Biobank"– currently going through submission, logged with Bioarchive

https://www.biorxiv.org/content/10.1101/2021.01.22.427640v1

## Funding

My thesis was self funded and I did not apply for supporting grants during the period.