<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>An investigation of the contribution of centrosomal genes to schizophrenia and cognitive function</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>Flynn, Mairéad</td>
</tr>
<tr>
<td><strong>Publication Date</strong></td>
<td>2020-03-11</td>
</tr>
<tr>
<td><strong>Publisher</strong></td>
<td>NUI Galway</td>
</tr>
<tr>
<td><strong>Item record</strong></td>
<td><a href="http://hdl.handle.net/10379/15846">http://hdl.handle.net/10379/15846</a></td>
</tr>
</tbody>
</table>
An investigation of the contribution of centrosomal genes to schizophrenia and cognitive function

By
Mairéad Flynn, B.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 Ciaran Morrison
Discipline of Biochemistry, School of Natural Sciences, National University of Ireland, Galway

August 2019
Table of Contents

Table of Contents ............................................................................................................. I
Declaration ......................................................................................................................... VI
Statement of Contribution ................................................................................................. VII
Acknowledgements .......................................................................................................... VIII
List of Figures ..................................................................................................................... X
List of Tables ....................................................................................................................... XII
List of Supplementary Tables ........................................................................................... XIV
List of Abbreviations ......................................................................................................... XVI
Abstract ............................................................................................................................. XXI

1 Introduction ..................................................................................................................... 1

1.1 Schizophrenia .............................................................................................................. 1
  1.1.1 Overview of schizophrenia .................................................................................. 1
  1.1.2 Prevalence and socioeconomic burden .............................................................. 2
  1.1.3 Current therapies ............................................................................................... 3
  1.1.4 Known schizophrenia aetiology ......................................................................... 4

1.2 Genetic Studies .......................................................................................................... 6
  1.2.1 Genetic variation in the human genome ............................................................... 6
  1.2.2 Linkage disequilibrium ....................................................................................... 7
  1.2.3 Schizophrenia heritability .................................................................................. 8
  1.2.4 Linkage studies ................................................................................................... 8
  1.2.5 Candidate gene studies ...................................................................................... 9
  1.2.6 Genome-wide association studies ....................................................................... 9
  1.2.7 CNVs ................................................................................................................ 11
  1.2.8 Exome sequencing ............................................................................................. 11

1.3 Cognition ................................................................................................................... 12
  1.3.1 Cognitive dysfunction in schizophrenia ............................................................... 12
  1.3.2 Schizophrenia and cognitive endophenotypes ..................................................... 13
  1.3.3 Genetics of cognition ......................................................................................... 16
  1.3.4 Genetic overlap between schizophrenia and cognition ..................................... 17

1.4 Centrosomes .............................................................................................................. 17
  1.4.1 Centrosome structure ......................................................................................... 18
  1.4.2 Centrosome functions ....................................................................................... 23
  1.4.3 Centrosome cycle ............................................................................................... 26
1.5 Primary Cilia .................................................................................................................. 31
  1.5.1 Primary cilia structure .......................................................................................... 31
  1.5.2 Functions of the primary cilia ............................................................................. 36
  1.5.3 Primary cilia assembly and disassembly ............................................................... 38
1.6 The role of the centrosome/cilia in schizophrenia and cognition ......................... 41
1.7 Project aims ................................................................................................................. 43

2 Materials and Methods ................................................................................................. 44
  2.1 Materials .................................................................................................................. 44
    2.1.1 Chemicals and solutions ................................................................................... 44
    2.1.2 Molecular Biology Reagents .......................................................................... 46
    2.1.3 Antibodies ........................................................................................................ 48
  2.2 Biological materials and tissue culture reagents ..................................................... 50
    2.2.1 Bacterial strains ................................................................................................. 50
    2.2.2 Tissue culture reagents .................................................................................... 51
    2.2.3 Cell lines and culture conditions ...................................................................... 51
  2.3 Identifying SZ risk genes with centrosomal function .............................................. 52
  2.4 Neuropsychological testing of SZ risk SNPs within centrosomal genes ............... 52
    2.4.1 Study participants and neuropsychological tests ............................................. 52
    2.4.2 Genotyping and statistical analysis .................................................................. 54
    2.4.3 Replication ........................................................................................................ 55
  2.5 Enrichment Analysis ............................................................................................... 56
    2.5.1 GWAS Data ...................................................................................................... 56
    2.5.2 Gene-set Analysis ............................................................................................. 56
    2.5.3 Analysis of gene-sets using rare variant data ................................................... 57
  2.6 Cell biology methods ............................................................................................... 58
    2.6.1 Cell maintenance and proliferation analysis ...................................................... 58
    2.6.2 Stable cell line generation ............................................................................... 59
    2.6.3 Transient transfections .................................................................................... 59
    2.5.5 Serum starvation .............................................................................................. 59
    2.6.6 Microscopy methods ........................................................................................ 60
  2.7 Protein methods ....................................................................................................... 61
    2.7.1 Protein extraction and preparation ................................................................... 61
    2.7.2 SDS-PAGE ....................................................................................................... 61
    2.7.3 Semi-dry transfer .............................................................................................. 62
    2.7.4 Immunoblotting technique .............................................................................. 63
2.8 Nucleic acid methods ........................................................................................................... 63
  2.8.1 RNA isolation and cDNA synthesis ............................................................................. 63
  2.8.2 Genomic DNA extraction ............................................................................................ 64
  2.8.3 Plasmid DNA prep ....................................................................................................... 64
  2.8.4 Polymerase Chain Reaction (PCR) .............................................................................. 65
  2.8.5 Restriction digestion of DNA ...................................................................................... 66
  2.8.6 Modification of DNA ends ........................................................................................... 67
  2.8.7 Ligation ........................................................................................................................ 67
  2.8.7 Assembly of CRISPR plasmids ..................................................................................... 68
  2.8.8 Analysis of cloned DNA and sequencing ................................................................. 68
  2.8.9 Agarose gel electrophoresis ....................................................................................... 68
  2.8.10 DNA purification methods ......................................................................................... 69
  2.8.11 Preparation and transformation of chemically competent E.coli ......................... 69
2.9 Migration analysis ............................................................................................................... 70
2.10 Differentiation of SH-SY5Y cells ..................................................................................... 70
2.9 RNA-Seq............................................................................................................................... 71
  2.9.1 Cell culture conditions and RNA extraction .............................................................. 71
  2.9.2 cDNA library construction and sequencing ............................................................. 71
  2.9.3 Mapping and counting ............................................................................................... 71
  2.9.4 Differential expression analysis ............................................................................... 72
  2.9.5 Pathway Analysis ........................................................................................................ 72
2.10 Computer programmes and Bioinformatics tools ......................................................... 72

3 Investigation of centrosomal genes for contribution to SZ and cognition .......... 75
  3.1 Introduction ...................................................................................................................... 75
  3.2 Identification of SZ risk genes with centrosomal function ........................................ 76
  3.3 Prioritised genes ............................................................................................................. 84
  3.4 Neuropsychological analysis ......................................................................................... 86
  3.5 Replication ...................................................................................................................... 89
  3.6 Analysis of centrosomal genes in neurodevelopmental disorders and cognition ....... 89
    3.6.1 Analysis of centrosomal genes using common variant data ................................... 90
    3.6.2 Analysis of centrosomal genes using rare variant data .......................................... 93
  3.7 Cloning and overexpression of prioritised genes ......................................................... 93
  3.8 Discussion ..................................................................................................................... 96
  3.9 Supplementary data .................................................................................................... 99

4 Functional Analysis of SDCCAG8 ................................................................................. 103
# Table of Contents

4.1 Introduction .............................................................................................................. 103
4.2 Bioinformatic analysis of SDCCAG8 ........................................................................ 104
4.3 Localisation of SDCCAG8 throughout the cell cycle ............................................ 110
4.4 Genome editing of the $SDCCAG8$ locus in hTERT-RPE1 and SH-SY5Y cells using CRISPR/Cas9 ................................................................................................. 111
4.5 Generation of RPE1 rescue cell lines and proliferation analysis .......................... 114
4.6 Reduction in PCM and satellite recruitment in SDCCAG8 deficient cells .......... 116
4.7 SDCCAG8-deficient cells show no centriole duplication defects ......................... 120
4.8 Dynamics of SDCCAG8 localisation in ciliated cells .......................................... 121
4.9 Loss of SDCCAG8 reduces ciliation frequency in human cells ......................... 122
4.10 Localisation of key ciliary regulators in the absence of SDCCAG8 ................. 124
4.11 Disruption of ciliary signalling in the absence of SDCCAG8 ......................... 128
4.12 Loss of SDCCAG8 significantly impacts cell migration ................................. 128
4.13 Loss of SDCCAG8 reduces SH-SY5Y cells’ ability to differentiate ................. 129
4.14 Discussion .............................................................................................................. 132

5 Transcriptional dysregulation of genes associated with SZ, IQ and EA in SDCCAG8-deficient cells .............................................................................................................. 137
5.1 Introduction .............................................................................................................. 137
5.2 Quality assessment of RPE1 RNA-Seq data ............................................................ 138
5.3 Differentially expressed genes in SDCCAG8-deficient RPE1 cells .................... 140
5.4 Pathway analysis of DEGs in SDCCAG8 deficient RPE1 cells ............................. 142
5.5 Generation and quality assessment of SH-SY5Y RNA-Seq data ......................... 146
5.6 Differentially expressed genes in SDCCAG8 deficient SH-SY5Y cells ............... 148
5.7 Pathway Analysis of DEGs in SDCCAG8 deficient SH-SY5Y cells .................... 150
5.8 Overlap of DEGs upon SDCCAG8 disruption ...................................................... 154
5.9 Gene-set analysis of GO terms enriched for SDCCAG8 dysregulated genes .... 157
5.10 Discussion .............................................................................................................. 161
5.11 Supplementary data ............................................................................................. 165

6. General Discussion .................................................................................................... 168
6.1 Summary of main findings ..................................................................................... 168
6.2 Strengths and weaknesses ..................................................................................... 169
6.2.1 GWAS .................................................................................................................. 169
6.2.2 Neuropsychological analysis .......................................................................... 170
6.2.3 Model systems .................................................................................................. 171
6.2.4 RNA-seq ............................................................................................................ 172
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.3 Mechanistic insight</td>
<td>172</td>
</tr>
<tr>
<td>6.4 Future directions</td>
<td>174</td>
</tr>
<tr>
<td>6.5 Translational applications and drug discovery</td>
<td>175</td>
</tr>
<tr>
<td>6.6 Concluding remarks</td>
<td>176</td>
</tr>
<tr>
<td>References</td>
<td>178</td>
</tr>
<tr>
<td>Appendices</td>
<td>209</td>
</tr>
<tr>
<td>Appendix A: List of Presentations and Awards</td>
<td>209</td>
</tr>
<tr>
<td>Oral presentations:</td>
<td>209</td>
</tr>
<tr>
<td>Poster presentations:</td>
<td>209</td>
</tr>
<tr>
<td>Awards:</td>
<td>210</td>
</tr>
<tr>
<td>Appendix B: Submissions</td>
<td>210</td>
</tr>
<tr>
<td>Appendix C: Study Funding and Acknowledgements</td>
<td>210</td>
</tr>
</tbody>
</table>
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:
Mairéad Flynn
Statement of Contribution

Colleagues in the Cognitive Genetics & Cognitive Therapy (Coggene) Group, now in National University Ireland, Galway and previously in Trinity College Dublin, collected and prepared all genotype and phenotype information for the Neuropsychological testing described in detail in Sections 2.4 and 3.4 prior to my arrival to the group.

For RNA-Seq experiments described in Sections 2.9, 5.2 and 5.5, the Genomics Core Technology Unit at Queen’s University Belfast, sequenced the DNA-libraries, aligned all reads to the genome, aggregated reads by gene and assessed the quality of the data generated.
Acknowledgements

“Nothing of me is original. I am the combined effort of everyone I've ever known.” Chuck Palahniuk

Although at times I found my PhD challenging, as many do, I believe I am only better off for the time I have spent in Galway. That is largely down to the people I have met along the way, I found it very difficult to put into words just how thankful I am to everyone, but below I gave it a shot.

First and foremost I would like to thank my two supervisors Derek Morris and Ciaran Morrison for giving me the opportunity to work with them. I consider myself extremely lucky to have had two supervisors that were so approachable, understanding, patient and fair. I am so grateful for the support that ye offered. I will certainly miss working alongside ye and wish you both the best of luck with your future research.

I’d like to thank everyone in the CCB, all the members of CogGene including Prof. Garry Donohoe, the Morris and Morrison lab groups including; Donna, Laura W., Marc, Joan, Laura F., Yetunde and Ebtissal. Everyone made it such a great atmosphere to work in. Thanks for listening to my endless moaning and for all your guidance and support over the years.

To the board game crew; Tobias, Anna, Karen, Sebastian, Eoghan and Janna. Thank you very much for introducing me to the world of board games and the deliciousness that is Spaetzle. I will miss the many evenings of food and fun. A special thanks to Janna who managed to keep me well fed through my PhD, for which I am very grateful.

Thank you to all the children and staff I met at Galway Autism Partnership, GAP house is a very special place where I really loved spending time. The people I met there brought a huge amount of joy to the time I spent in Galway.
Perhaps the hardest to say goodbye to and the people I owe a huge thanks to is the Friday night crew especially; Michael, Rachel, Aisling, Adrian, Higgins and Corrado. You are all such amazing people and I definitely wouldn’t have made it through my time in Galway without you guys. Thanks for making me laugh so often and indulging me in a boogie every now and again. I already miss you guys and the craic we had.

Thank you to all my friends from home, in particular Ali. Thanks for visiting so often and making sure I was having fun. Thank you for your support and for being so amazing. Ali your work ethic and positive outlook is inspiring and definitely helped keep me going.

I would like to thank my family, my Mam and Dad; Patricia and Adrian and my two younger sisters Nèidin and Aisling. Thanks for always supporting me in whatever I choose to do and pushing me to be the best I can be. Thanks for teaching me to work hard for what I want. Most of all I’m thankful for the love and incredible craic that are in abundance every time we all get together.

Finally a massive thank you to my partner, Colman. Thank you for your endless supply of patience, positivity, support and love. Thank you for always being there and being ready to listen or even at times accompany me to the lab. You never failed to make everything feel better when it was needed and I will always be grateful to you. My partner in crime, love you.
List of Figures
Figure 1.1: Centrosome and centriole structure 19
Figure 1.2: The centrosome cycle 27
Figure 1.3: Hierarchal assembly of distal and subdistal appendages 30
Figure 1.4: Structure of the primary cilium 32
Figure 1.5: Hedgehog signalling at the primary cilium 37
Figure 1.6: Physical stages in the intracellular pathway of ciliogenesis 39
Figure 3.1: Identification of schizophrenia risk genes with centrosomal function. 76
Figure 3.2: MAD1L1 Regional Plot 78
Figure 3.3 PRKDI Regional Plot 79
Figure 3.4 MPHOSPH9 Regional Plot 80
Figure 3.5 GIGYF2 Regional Plot 81
Figure 3.6 SDCCAG8 Regional Plot 82
Figure 3.7: MAPK3 Regional Plot 83
Figure 3.8: Gene set analysis of centrosomal proteins in SZ, EA and IQ 91
Figure 3.9: Cloning full length cDNA of prioritised genes into pEGFP-N1 94
Figure 3.10: Transient expression of full-length human MAD1L1, MAPK3, PRKDI, SDCCAG8 and GIGYF2 cDNA 95
Figure 4.1: SDCCAG8 genomic structure. 105
Figure 4.2: Comparative analysis of human SDCCAG8 and its orthologues 109
Figure 4.3: Conserved regions of SDCCAG8 in vertebrates 110
Figure 4.4: SDCCAG8 localisation throughout the cell cycle 111
Figure 4.5: Generation of SDCCAG8 knockout hTERT-RPE1 and SH-SY5Y cell lines 112
Figure 4.6: Sequencing confirmation of SDCCAG8 KO clones 114
Figure 4.7: Stable re-expression of SDCCAG8 to generate rescue RPE1 cell line. 115
Figure 4.8: SDCCAG8-deficient and rescue cell lines show no proliferation defect 116
Figure 4.9: Reduction in the recruitment of PCM and satellite components in G1 hTERT-RPE1 cells deficient of SDCCAG8. 118
Figure 4.10: Loss of SDCCAG8 has no effect on localisation and composition of centrosomal proteins in mitotic hTERT-RPE1 cells. 119
Figure 4.11 Regulation of centriole duplication remains intact in SDCCAG8 depleted cells 120
Figure 4.12: SDCCAG8 localisation in ciliated cells 121
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.13</td>
<td>Loss of SDCCAG8 significantly impacts cilia formation</td>
<td>123</td>
</tr>
<tr>
<td>4.14</td>
<td>Loss of SDCCAG8 significantly impacts ciliary length</td>
<td>124</td>
</tr>
<tr>
<td>4.15</td>
<td>Localisation of ciliary regulators remains intact in SDCCAG8 deficient cells</td>
<td>126</td>
</tr>
<tr>
<td>4.16</td>
<td>CP110 abundance is affected by SDCCAG8 deficiency</td>
<td>127</td>
</tr>
<tr>
<td>4.17</td>
<td>Defective Smo localisation in residual SDCCAG8 deficient cilia</td>
<td>128</td>
</tr>
<tr>
<td>4.18</td>
<td>Defective migration in SDCCAG8 depleted SH-SY5Y cells</td>
<td>129</td>
</tr>
<tr>
<td>4.19</td>
<td>Loss of SDCCAG8 affects cells ability to differentiate</td>
<td>131</td>
</tr>
<tr>
<td>5.1</td>
<td>Data generated in RNA-Seq experiments of RPE1 cells is of good quality</td>
<td>139</td>
</tr>
<tr>
<td>5.2</td>
<td>Principal component analysis of RPE1 expression data</td>
<td>140</td>
</tr>
<tr>
<td>5.3</td>
<td>Canonical pathways enriched for SDCCAG8 dysregulated genes in RPE1 cells under baseline conditions</td>
<td>142</td>
</tr>
<tr>
<td>5.4</td>
<td>Canonical pathways enriched for SDCCAG8 dysregulated genes in RPE1 cells following serum starvation</td>
<td>143</td>
</tr>
<tr>
<td>5.5</td>
<td>SH-SY5Y RNA-Seq data is of high quality</td>
<td>147</td>
</tr>
<tr>
<td>5.6</td>
<td>Principal component analysis of SH-SY5Y expression data</td>
<td>148</td>
</tr>
<tr>
<td>5.7</td>
<td>Canonical pathways enriched for DEGS in SDCCAG8 deficient SH-SY5Y cells at baseline conditions</td>
<td>150</td>
</tr>
<tr>
<td>5.8</td>
<td>Canonical pathways enriched for DEGs in SDCCAG8 deficient SH-SY5Y cells after serum starvation</td>
<td>151</td>
</tr>
<tr>
<td>5.9</td>
<td>Overlap of DEGs upon SDCCAG8 disruption</td>
<td>154</td>
</tr>
<tr>
<td>5.10</td>
<td>Gene expression pathways impacted by SDCCAG8 deficiency affect EA</td>
<td>158</td>
</tr>
<tr>
<td>5.11</td>
<td>Gene expression pathways impacted by SDCCAG8 deficiency affect IQ</td>
<td>159</td>
</tr>
<tr>
<td>5.12</td>
<td>Gene expression pathways impacted by SDCCAG8 deficiency affect SZ</td>
<td>159</td>
</tr>
</tbody>
</table>
List of Tables

Table 1.1: Members of MKS, NPHP and CEP290 complexes 35
Table 2.1: Common buffers and reagents 44
Table 2.2: Plasmids used in the study 47
Table 2.3: Kits used in this study 47
Table 2.4: List of primary antibodies used for this study 48
Table 2.5: List of secondary antibodies used for this study 49
Table 2.6: Cell lines used in this study 51
Table 2.7: Participant demographics – neuropsychological test sample 53
Table 2.8: SDS-PAGE gel composition used for protein detection 62
Table 2.9: PCR conditions for KOD Hot Start 65
Table 2.10: PCR programme cycling conditions 65
Table 2.11: List of oligonucleotides used in this study 66
Table 2.12: Bioinformatic tools and computer programmes used in this study 73
Table 3.1: Regression analysis of neuropsychological variable test scores in all broad psychosis cases and controls 87
Table 3.2: UK Replication Data for Episodic Memory 89
Table 3.3: Enrichment analysis of the centrosomal list for genes harbouring de novo variants in SZ, ASD and ID 93
Table 4.1: Comparative analysis of predicted human SDCCAG8 isoforms 104
Table 4.2 Mutations of SDCCAG8 known to cause Bardet-Biedl syndrome or Senior-Løken syndrome 106
Table 5.1: Differentially expressed genes in RPE1 cells depleted of SDCCAG8 141
Table 5.2: GO terms that are enriched for DEGs in SDCCAG8 deficient RPE1 cells at baseline conditions 144
Table 5.3: GO terms that are enriched for DEGs in SDCCAG8 deficient RPE1 cells after serum starvation 145
Table 5.4: Genes differentially expressed in SH-SY5Y cells upon SDCCAG8 depletion 149
Table 5.5: GO terms that are enriched for DEGs in SDCCAG8 deficient SH-SY5Y cells at baseline conditions 152
Table 5.6: GO terms that are enriched for DEGs in SDCCAG8 deficient SH-SY5Y cells after serum starvation 153
Table 5.7: Gene-set analysis of GO terms impacted by SDCCAG8 deficiency and associated with IQ, EA and SZ in control GWAS

160
List of Supplementary Tables

Supplementary Table 3.1: Centrosomal gene list 99
Supplementary Table 3.2: Gene-set analysis of centrosome genes in IQ, EA and SZ and post-hoc analysis in control phenotypes 102
Supplementary Table 5.1: Genes that are differentially expressed upon SDCCAG8 ablation in RPE1 cells at baseline conditions CD
Supplementary Table 5.2: Genes that are differentially expressed upon SDCCAG8 ablation in RPE1 cells after serum starvation. CD
Supplementary Table 5.3: Canonical pathways enriched for DEGS in SDCCAG8 deficient RPE1 cells at baseline conditions CD
Supplementary Table 5.4: Canonical pathways enriched for DEGS in SDCCAG8 deficient RPE1 cells after serum starvation CD
Supplementary Table 5.5: GO terms that are enriched for DEGs in SDCCAG8 deficient RPE1 cells at baseline conditions CD
Supplementary Table 5.6: GO terms that are enriched for DEGs in SDCCAG8 deficient RPE1 cells after serum starvation CD
Supplementary Table 5.7: Genes that are differentially expressed upon SDCCAG8 ablation in SHSY5Y cells at baseline conditions CD
Supplementary Table 5.8: Genes that are differentially expressed upon SDCCAG8 ablation in SHSY5Y cells after serum starvation CD
Supplementary Table 5.9: Canonical pathways enriched for DEGS in SDCCAG8 deficient SH-SY5Y cells at baseline CD
Supplementary Table 5.10: Canonical pathways enriched for DEGS in SDCCAG8 deficient SH-SY5Y cells after serum starvation CD
Supplementary Table 5.11: GO terms that are enriched for DEGs in SDCCAG8 deficient SHSY5Y cells at baseline conditions CD
Supplementary Table 5.12: GO terms that are enriched for DEGs in SDCCAG8 deficient SHSY5Y cells after serum starvation CD
Supplementary Table 5.13: Gene-set analysis of GO terms impacted by SDCCAG8 deficiency in SZ 165
Supplementary Table 5.14: Gene-set analysis of GO terms impacted by SDCCAG8 deficiency in IQ 166
Supplementary Table 5.15: Gene-set analysis of GO terms impacted by SDCCAG8 deficiency in EA 167
List of Abbreviations

aa Amino acids
AD Alzheimer’s disease
ADHD Attention deficient hyperactivity disorder
APS Ammonium persulphate
Arl13B ADP-ribosylation factor-like 13B
ATM Ataxia telangiectasia, mutated
ASD Autism spectrum disorder
ATP Adenosine-5’-triphosphate
ATR ATM-Rad3 related
BBS Bardet-Biedl syndrome
BD Bipolar disorder
BLAST Basic local alignment search tool
BMI Body mass index
bp Base pair(s)
BRCA1 Breast cancer associated gene 1
BRCA2 Breast cancer associated gene 2
BSA Bovine serum albumin
CA Centrosome amplification
CAD Cardiovascular disease
CANTAB Cambridge neuropsychological test automated battery
Cas9 CRISPR associated protein 9
CD Crohn’s disease
CDC Cell division cycle
CDK Cyclin-dependent kinase
CDK5RAP2 Cyclin-dependent kinase5 regulatory associated protein 2
cDNA Complementary DNA
Centrobin Centrosomal BRCA2 interacting protein
Cep Centrosomal protein
CHK Checkpoint kinase
List of Abbreviations

C-Nap1 Centrosomal Nek2-associated protein 1
CNV Copy number variant
CPAP Centrosomal P4.1-associated Protein
CPT Continuous Performance Task
CRISPR Clustered regularly interspaced short palindromic repeat
CRT Cognitive remediation therapy
CV Ciliary vesicle DA distal appendages
DALY Disability adjusted life years
DDR DNA damage response
DFPLC Dorsolateral prefrontal cortex
DNM De novo mutations
DMSO Dimethylsulfoxide
dNTP Deoxyribonucleotide-5"-triphosphate
DSM5 Diagnostic and Statistical Manual of Mental Disorders 5th edition
EA educational attainment
ECL Enhanced chemiluminesence
EDTA Ethylenediaminetetraacetic acid
EGTA Ethylene glycol tetraacetic acid
FBS Foetal bovine serum
FITC Fluorescein isothiocyanate
γ-TURC γ-tubulin ring complex
γ-TUSC γ-tubulin small complex
GFP Green fluorescent protein
GSA gene-set analysis
GWAS genome-wide association studies
Hh Hedgehog
hTERT-RPE1 Human telomerase reverse transcriptase- retinal pigment epithelial
hiPSC Human induced pluripotent stem cells
HVLT Hopkins verbal learning test
ID Intellectual disability
List of Abbreviations

IF Immunofluorescence microscopy
IFT Intraflagellar transport
Indel Insertions and deletions
IPSAQ Internal, personal and situational attribution questionnaire
IQ Intelligence quotient
IR Ionizing radiation
Kb Kilobase pair(s)
LB Luria-Bertani medium
LD linkage disequilibrium
LM logical memory
LNS Letter-number sequencing
MAF minor allele frequency
MAP Microtubule associated protein
MDD major depressive disorder
MTAG Multi-trait analysis of GWAS
MHC major histocompatibility complex
MT Microtubule
MTOC Microtubule-organising centre
NCBI National Center for Biotechnology Information
NEK NIMA-related kinase
NPC neural progenitor cell
NS non-significant
OCD obsessive compulsive disorder
ODF2 Outer dense fiber protein 2
OFD1 Oral-facial-digital syndrome 1
PBS Phosphate buffered saline
PC Polycystin
PCM Pericentriolar material
PCR Polymerase chain reaction
Pen/Strep Penicillin/streptomycin
List of Abbreviations

PFA Paraformaldehyde
PGC Psychiatric Genomics Consortium
PLK Polo-like kinase
PNK Polynucleotide kinase
PRS polygenic risk score
QC quality control
RNA Ribonucleic acid
RNAi RNA interference
RNase Ribonuclease
RT-PCR Reverse transcription-PCR
SART Sustained Attention to Response Task
SDA Subdistal appendage
SDCCAG8 Serologically defined colon cancer antigen 8
SDS Sodium dodecyl sulphate
SDS-PAGE SDS polyacrylamide gel electrophoresis
SEM Standard error of the mean
SHh Sonic hedgehog
siRNA Short interfering RNA
SLS Senior-Loken syndrome
SMO Smoothened receptor
SNP single-nucleotide polymorphism
ss Serum starved
SSTR3 Somatostatin receptor-3
ST stroke
STIL SCL/TAL1 interrupting locus
SUFU Suppressor of fused
SWM spatial working memory
SZ schizophrenia SAP Shrimp alkaline phosphatase
T2D type 2 diabetes
TAE Tris acetate EDTA
TEMED N,N,N''-tetramethylethylenediamine
TF Transition fibers
TG Tris-glycine
Tris Tris(hydroxymethyl)aminomethane
TZ Transition zone
UC ulcerative colitis
VIQ verbal IQ
VNR verbal numerical reasoning
v/v Volume per volume
Wnt Wingless-type MMTV integration site family
WT Wild-type
WAIS III Wechsler Adult Intelligence Scale 3rd Edition
WB Western blot
WMS III Wechsler Memory Scale 3rd Edition
Abstract
The centrosome is the key microtubule organising centre of animal somatic cells. Centrosomes play a crucial role in brain development, influencing cell shape, polarity, motility, and division. The older of the two barrel-shaped centrioles that form the core around which the pericentriolar material assembles can serve as the base of the primary cilium, an antenna-like structure with sensory and signal transduction functions that are essential for brain development. Mutations that affect ciliary function cause ciliopathies, heterogeneous developmental or degenerative disorders that affect multiple organs, including the brain.

Schizophrenia (SZ) is a chronic psychiatric disorder affecting approximately 1% of the population. Symptoms range from delusional thoughts and hallucinations to anhedonia and social withdrawal. Affected individuals also display cognitive deficits in IQ, memory and attention. Cognitive deficits are a key factor for explaining disability, leading to significant unemployment, homelessness and social isolation. SZ is highly heritable and significant genetic overlap between SZ and cognitive ability means that some genes contribute to both phenotypes. Genome-wide association studies (GWAS) have now identified hundreds of loci for SZ and cognition phenotypes such as human intelligence (IQ) and educational attainment (EA). The molecular mechanisms by which associated genes contribute to SZ risk and cognitive function are not well understood. However, the likelihood is that genetic variation within multiple biological pathways contributes to both illness and lower cognition function.

I have focused on the contribution of the centrosome genes to SZ and cognition. I found that, common variation in centrosome genes are enriched for association with IQ, while rare variants are enriched for association with ID and ASD. The ciliopathy gene SDCCAG8 is associated with SZ and EA and social cognition in our Irish dataset. Genome editing of SDCCAG8 caused defects in primary ciliogenesis and cilium-dependent cell signalling. These ciliary defects in the absence of SDCCAG8, I hypothesise are due to impaired protein trafficking to the ciliary transition zone (TZ), a specialised ciliary compartment that forms a gate, controlling protein entry and exit from...
the primary cilium. Transcriptomic analysis of SDCCAG8-deficient cells identified differentially expressed genes that are enriched in neurodevelopmental processes such as generation of neurons and synapse organization. These processes are enriched for genes associated with SZ, IQ and EA. Phenotypic analysis of SDCCAG8-deficient neuronal cells revealed impaired migration and neuronal differentiation, which have been implicated in SZ pathogenesis previously. These data implicate ciliary signalling in the aetiology of SZ and cognitive dysfunction.
1 Introduction

1.1 Schizophrenia

1.1.1 Overview of schizophrenia

Schizophrenia (SZ) affects 1% of adults and is a major global health problem. Patients with this chronic neuropsychiatric disorder can present with a diverse array of symptoms which are broadly categorised into positive, negative and cognitive symptoms. Positive symptoms include hallucinations, the most common being auditory and visual, where people have false experiences; hearing voices and/or seeing things that are not there. Delusions are also a positive symptom where patients can hold false beliefs that are unchanged even when presented with contradictory evidence. Negative symptoms include social withdrawal, a decrease in motivation, and an inability to express emotions or to find pleasure in things (Owen et al., 2016). Cognitive symptoms include confused and disorganised speech, deficits in IQ (Fujino et al., 2017), memory (Forbes et al., 2009), attention (Wang et al., 2005) and social functioning (Green et al., 2015). Current anti-psychotic drugs, discovered serendipitously >50 years ago, are only partially effective. They can reduce psychotic symptoms but do not target negative symptoms and cognitive deficits. SZ is ranked by the WHO as one of the top five causes of years lived with disability from ages 18-34 (Üstün et al., 1999). Cognitive deficits (poor memory, attention, and IQ) are key factors for explaining disability in SZ, leading to significant unemployment, homelessness and social isolation (Green et al., 2006). The average life expectancy for an individual with SZ is 20 years below that of the general population. This has been attributed to a number of factors such as side effects of treatments, lifestyle factors, suicide and comorbid conditions which are diagnosed much later and are ineffectively treated in patients with SZ (Laursen et al., 2014). In the public perception, SZ can often be associated with violence. However, patients with SZ are 14 times more likely to be the victim than the perpetrator of violent crime (Fitzgerald et al., 2005, Brekke et al., 2001).
SZ is highly heritable (Hilker et al., 2018) and there is molecular genetic overlap between SZ and cognitive ability (Lencz et al., 2014), meaning that some genes contribute to both phenotypes. Genome-wide association studies (GWAS) have now identified hundreds of loci for SZ (Pardiñas et al., 2018) and cognition phenotypes such as human intelligence (IQ; Savage et al., 2018) and educational attainment (EA; Lee et al., 2018). However, the molecular mechanisms by which associated genes contribute to SZ risk and cognitive function are not well understood.

1.1.2 Prevalence and socioeconomic burden

The lifetime prevalence of SZ in adults is approximately 1%, although it has been shown that this figure can vary up to three-fold depending on the diagnostic definition that is used (van Os and Kapur, 2009). A clinical diagnosis of SZ is arrived at in concordance with the Diagnostic and Statistical Manual of Mental Disorders (5th edition, DSM-5) (American Psychiatric Association, 2013), which requires the presence of at least two of the core symptoms listed above for a period of at least 6 months. SZ is more prevalent in males than females at a ratio of 1.4:1 (McGrath et al., 2008). For males, onset of symptoms is typically between 21-25 years old. Female symptom onset is slightly later, at 25-30 years, and they also have a second onset peak after 45 years old (Abel et al., 2010). Some studies also suggest that women have a better prognosis once diagnosed (Morgan et al., 2008). It has been hypothesised that estrogen may play a protective role against SZ and partially explain the sex differences that are observed (Kaneda and Ohmori, 2005, Grigoriadis and Seeman, 2002). A number of environmental factors have been shown to affect the rate of SZ, including urbanised living (Krabbendam and van Os, 2005), exposure to cannabis (Moore et al., 2007) and migrant status (Cantor-Graae and Selten, 2005).

The socioeconomic burden of the disorder is extremely high due to the early onset of the disorder, poor life expectancy, increased unemployment and the chronic nature of the disorder. The economic burden of SZ varies from 0.02 to 5.46% of the gross domestic product depending on the country being analysed (Chong et al., 2016). In 2006, the cost of SZ in Ireland was estimated at €460.6 million. A large fraction of these costs are indirect costs due to absenteeism, lack of employment or the inability to work, premature
death and informal care provided by families (Behan et al., 2008). The World Health Organisation uses a measure of disease burden known as disability adjusted life years (DALYs), a cumulative measure of years lost due to ill-health and/or early death. SZ accounts for 1.1% of all DALYs, and 2.8% of years lived with disability. SZ is the 8th leading cause of DALYs worldwide in the 15-44 age group (Rössler et al., 2005).

1.1.3 Current therapies

Pharmacological treatments for SZ antagonize dopamine D2 receptors. The neuroleptic, chlorpromazine, was first synthesised in France as a result of research on compounds with antihistaminic properties. Further behavioural experiments led to it being recommended for psychosis in 1961 (López-Muñoz et al., 2005). Antipsychotics were later shown to act by dampening dopamine neuron excitability (Seeman and Lee, 1975). The antagonism of dopamine D2 receptor in the mesolimbic system is thought to be responsible for the reduction in positive symptoms (Walter et al., 2009). However, these compounds have a lack of selectivity, leading to numerous side-effects. The most frequently reported side effects are extrapyramidal effects; unwanted movement, tremors, rigidity and muscle breakdown. Side effects related to the antihistaminic function of antipsychotics include sedation, drowsiness, vertigo, sleep disturbances, agitation, dementia memory loss and depression (Stępnicki et al., 2018). Blurred vision and dry mouth are also possible side effects as well as weight gain which contributes to the increased risk of metabolic issues that are experienced by patients with SZ including increased risk of heart disease, type 2 diabetes and high cholesterol (De Hert et al., 2009).

Second-generation antipsychotics were developed in the hope that they may have improved side-effect profiles. Clozapine was the first of these compounds developed. These second-generation antipsychotics have a lower affinity for the dopamine D2 receptor, and have a higher affinity for serotonin 5-HT2A receptors and are classed as atypical antipsychotics. Some, but not all, second-generation antipsychotics have been shown to be more effective than first-generation compounds (Leucht et al., 2013). Patients with SZ are categorised as treatment resistant if they are nonresponsive to at least two trials of antipsychotic medication of adequate dose and duration. Approximately 30% of patients with SZ are treatment-resistant, with the remaining 70% experiencing a
reduction in positive symptoms (Meltzer, 1997). Clozapine is effective in reducing positive symptoms in 30% of patients with treatment-resistant SZ, to whom clozapine is usually restricted because of the risk of agranulocytosis (Kane et al., 1988). In addition to the large side-effect profiles of antipsychotics, there is also evidence that they cause dopamine supersensitivity where long term administration of D_2 agonists leads to an upregulation of D_2 receptors. If patients then decide to discontinue their medication, dopamine supersensitivity causes a rapid relapse of positive symptoms and new or more severe psychotic symptoms. Dopamine supersensitivity can also lead to tolerance to previously used therapeutic doses of antipsychotics (Chouinard et al., 2017). It has been established by brain imaging studies that there are reductions to both white matter and grey matter volume in patients with SZ that correlate with functional impairment (Dietsche et al., 2017, Ho et al., 2003). Perhaps most alarming of all, chronic exposure to antipsychotic medication in macaque monkeys led to significant shrinkage across brain regions (Dorph-Petersen et al., 2005), suggesting that the gross brain abnormalities observed in patients with SZ may be partially caused by chronic antipsychotic exposure. There is a substantial need for newer treatments that not only target positive but also negative and cognitive symptoms.

1.1.4 Known schizophrenia aetiology

SZ has been historically considered a progressive brain disease, a view that was reinforced by MRI findings demonstrating progressive brain tissue loss (Zipursky et al., 2013). However, as discussed above, these brain changes can be explained in part by chronic use of antipsychotics. In recent years, neurodevelopmental mechanisms of SZ have gained more traction. It is now thought that variation in susceptibility genes and environmental insults during early neurodevelopment initiate neurophysiological changes that in combination with aberrant postnatal brain maturation culminate in the onset of disease (Jaaro-Peled et al., 2009). One of the first neurophysiological changes associated with SZ was dysfunction of the dopaminergic system, after antipsychotics were shown to attenuate this pathway (Davis et al., 1991, Lee and Seeman, 1980). The dopamine hypothesis of SZ proposes that hyperactive transmission in the mesolimbic areas result in positive symptoms, while hypoactive dopamine transmission in the prefrontal cortex contributes to negative symptoms (Brisch et al., 2014). There has been
intensive research on the link between dopamine and SZ. Imaging studies have shown increased dopamine synthesis and release in the striatum of SZ patients (Howes et al., 2015). However, these studies have failed to show whether the observed imbalances are a consequence of SZ itself or due to antipsychotic treatment. In addition to this there are patients who do not respond to dopamine receptor agonists at all, and patients who experience relapses while faithfully taking their medication (Robinson et al., 1999). This suggests that there are further neurophysiological changes contributing to SZ.

Glutamatergic mechanisms of SZ were also hypothesised. The earliest reports of glutamatergic abnormalities in living patients with SZ was in 1980 (Kim et al., 1980). The prevailing glutamatergic theory is that NMDA receptor dysfunction contributes to SZ pathogenesis (reviewed by Howes et al., 2015). The roles of both glutamate and dopamine dysfunction in SZ have been supported by results from large genetic studies (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014). These studies also highlighted a greater and more complex aetiology in SZ pathophysiology than can be explained by glutamate and dopamine dysfunction alone. A more recent hypothesis has suggested that synaptic pruning is altered in SZ. There is an extensive elimination of synapses in the human cerebral cortex during late adolescence and early adulthood (Petanjek et al., 2011). This period is typically when individuals begin to present with SZ symptoms. Patients with SZ have been shown to have abnormal functional connectivity and a reduced number of dendritic spines (Konopaske et al., 2014, Glausier and Lewis, 2013, Stephan et al., 2009, Lawrie et al., 2002). In addition to this a recent publication has shown an increase in synapse elimination by microglia in SZ patient-derived models of synaptic pruning (Sellgren et al., 2019). The engulfing of synapses by microglia was partially explained by an increase in neuronal complement deposition. Genetic alteration of components of the complement system has been linked to SZ risk previously (Sekar et al., 2016).

The role of environmental factors is also highlighted by the neurodevelopmental model of SZ. In line with this prenatal exposure to viral infection, low birth weight, hypoxia and obstetric complications have all been associated with SZ (Murray et al., 2017). Later in life, childhood adversity including sexual abuse, physical abuse, and neglect, have been
shown to increase the risk of psychosis (Varese et al., 2012), as has parental separation (Morgan et al., 2007). Income inequality, urban living, use of potent cannabis and being part of a minority ethnic group have all been associated with increased SZ risk (Burns et al., 2014, Moore et al., 2007, Krabbendam and van Os, 2005, Cantor-Graae and Selten, 2005).

1.2 Genetic Studies

1.2.1 Genetic variation in the human genome

A typical human genome differs from the reference genome at approximately 3.5 to 5.0 million sites (The 1000 Genomes Project consortium, 2015). This variation can be broadly broken into two categories: (1) Structural variants that change a large number of nucleotides (>1 Kilobase (Kb)). These include copy number variants (CNVs) that are produced by insertions, deletions and duplications of sections of the genome, and translocations. (2) Sequence variants that make small changes to or do not alter nucleotide number; single nucleotide polymorphisms (SNPs), and small insertions and deletions (indels).

SNPs are single-base pair changes in the DNA sequence. They generally have two alternative alleles and are located in exonic, intronic and intergenic regions of the genome (Frazer et al., 2009). More than 99.9% of sequence variation from the reference genome is due to SNPs and short indels (The 1000 Genomes Project consortium, 2015). Functional consequences linked to SNPs include amino acid changes and alteration to mRNA stability and transcription factor binding affinity. SNP frequency is presented in terms of minor allele frequency (MAF); the frequency of the least common allele in the population. SNPs with a MAF greater than 5% are considered common variants, those with a MAF between 1-5% are low frequency and those with a MAF below 1% are considered rare (Frazer et al., 2009).
A typical genome also contains 2,100 to 2,500 structural variants. Collectively these variants affect ~20 million base pairs (bp) (The 1000 Genomes Project consortium, 2015). CNVs can contribute to disease via gene dosage, gene fusion and gene disruption effects. However CNVs can also be benign and can be an important mechanism for driving genome evolution (Zhang et al., 2009). Like SNPs, CNVs can vary in their frequency within the population from ones that are common to CNVs that are unique to a single individual. When the frequency of a CNV is lower than 1% it is considered rare (Nowakowska et al., 2017). Robust detection of CNVs continues to be problematic (Teo et al., 2012). Unidentified CNVs are likely to contribute to some of the missing heritability in disorders like SZ (Avramopoulos, 2018).

Solitary variants can exist in the human genome. However, it is more often the case that sequence and structural variants are related due to linkage disequilibrium (LD).

1.2.2 Linkage disequilibrium

LD is the non-random association of alleles at different loci in the population (Slatkin, 2008). SNPs tend to co-inherited with other SNPs in close proximity. This LD means that an allele occurs more often than expected with other alleles on the same chromosome (Psychiatric GWAS Consortium Coordinating Committee, 2009). Genetic recombination has a strong influence on LD, leading to a variation in LD rates across the genome and between populations. Regions with low rates of recombination maintain high rates of LD even when loci are quite distant from each other. These LD rates can be mapped, and the genome can be segregated into groups of SNPs that exhibit high LD with each other. LD can be used to map disease genes: if a SNP is associated with a condition then SNPs in high LD with that SNP become indirectly associated with the condition (Psychiatric GWAS Consortium Coordinating Committee, 2009). Knowledge of LD patterns and the sequencing of 1000s of human genomes, have led to the development of comprehensive SNP arrays that capture a high proportion of common sequence variation in the human genome for analysis in GWAS (Discussed in Section 1.2.6). The role that linkage studies have played in SZ will be discussed in Section 1.2.4.
1.2.3 Schizophrenia heritability

Heritability (h$^2$) is an estimate of the degree of variation in a phenotype that is directly attributable with genetic variation. Heritability estimates can therefore be useful in determining the degree to which environmental factors contribute to a condition relative to genetic contribution. SZ heritability is estimated to be 79% (Hilker et al., 2018). This figure was calculated by combining information from the Danish Twin Register and the Danish Psychiatric Research Register. In this same study for monozygotic twins, when one twin is diagnosed with SZ, the risk to the second twin is 33%. For dizygotic twins who only share approximately half of their DNA, the concordance rate was much lower at 7% (Hilker et al., 2018). Children whose two biological parents have a SZ diagnosis carry a 27% risk of developing the disorder themselves (Gottesman et al., 2010). Even for adopted children with a biological parent with SZ, their risk of developing the disorder is higher than that of the general population, further highlighting the important contribution of genetic variation in SZ (Shih et al., 2004). Despite the consistent evidence for the role of genetic variation in SZ aetiology, the search for specific genes that are robustly associated with SZ has proved challenging.

1.2.4 Linkage studies

One of the earliest strategies for identifying genetic risk loci for psychiatric conditions was through genetic linkage. In 1996, it was predicted that some of the most important loci for complex diseases would be uncovered using linkage studies (Risch and Merikangas, 1996). Linkage studies are best performed in large families where multiple family members present with a common condition. An attempt is made to identify genetic markers that co-segregate with the disorder, due to linkage with the casual variant (co-location on the same chromosome). Linkage analysis highlights a genomic region and further analysis is needed to pinpoint the exact casual variant. Linkage studies are best suited to Mendelian disorders and have led to the discovery of mutations in 1,600 different conditions (Psychiatric GWAS Consortium Coordinating Committee, 2009). However, despite considerable efforts, linkage studies have not reliably identified risk loci for psychiatric conditions (Bray et al., 2019). This lack of success indicates that the genetic contribution to these conditions does not adhere to simple monogenic models.
1.2.5 Candidate gene studies

Another important method of gene identification is based on genetic association. The aim is to identify susceptibility variants that alone are not sufficient to be causal. The most common design compares allele frequencies in affected individuals versus control individuals. Association studies use a chi-square statistic ($X^2$) to test for differences in allele frequencies. A statistically significant difference in allele frequency of a SNP indicates that one allele is associated with an increased risk of a disease, not causality (Visscher et al., 2017). Technological limitations meant that the early association studies focused on a small number of variants within candidate genes. Candidate genes were selected on the basis of their known biological roles. This hypothesis-driven approach led to numerous candidate genes being reported as associated with SZ risk. In 2008 the SZ Gene database reported on studies that identified over 700 candidate genes (Allen et al., 2008), a figure that increased to 1,008 by 2012 (http://www.szgene.org/). However, the reproducibility of these studies was very poor. This can be explained in retrospect by the small sample sizes, the low probability that a selected candidate allele will be truly associated with the phenotype of interest, the small effects that common variants have on psychiatric disease susceptibility and the lack of understanding of SZ pathophysiology making the selection candidate genes difficult (Bray et al., 2018).

1.2.6 Genome-wide association studies

Genome-wide association studies (GWAS) compare genome-wide SNP data between disease cases and controls in an attempt to identify a statistically significant differences in allele frequencies at SNPs that would identify them as associated with disease. The development of genotyping arrays has made it possible to simultaneously genotype 100,000s of SNPs in a cost-effective manner in a large number of individuals. Imputation, allows for the numbers of SNPs analysed to increase to millions, by using a reference panel of SNPs identified by sequencing and knowledge of LD to impute or infer genotypes at SNPs that have not been directly genotyped by the SNP array. Studying variants throughout the entire genome is a hypothesis-free approach to identifying risk variants (Visscher et al., 2017). Each SNP is tested for association with the phenotypic trait being analysed, this creates a large multiple test burden. In GWAS, a Bonferroni correction is often applied ($0.05/k$), where $k$ represents the number of tests performed.
(Dunn 1961, 1959). This correction leads to a stringent genome-wide threshold of \( P < 5 \times 10^{-8} \). It is now clear that common variants in the population individually confer a small increase in risk for neuropsychiatric disorders, which, in combination with the stringent P-value, means that very large samples are generally required to detect significant associations. Progress in psychiatric genetics has been accelerated due to international collaborative efforts namely by the Psychiatric Genomics Consortium (PGC). A landmark study published by the PGC identified 108 independent risk loci associated with SZ (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014). Most recently a meta-analysis involving an additional 11,260 cases and 24,542 controls bringing the total to 40,675 cases and 63,643 controls identified a further 50 loci a genome-wide significance (Pardiñas et al., 2018). While GWAS studies have been successful in identifying risk loci, further functional studies are usually required to understand how genes within these loci are contributing to risk.

In GWAS there are many variants that do not reach genome wide significance but exert a weak effect on SZ risk. The highly polygenic nature of SZ was first appreciated in 2009, when the combining of thousands of variants that had weak effects on risk individually, was shown to account for a significant proportion of SZ risk in an independent sample (International Schizophrenia Consortium, 2009). This summation of multiple variants generates what is known as a polygenic risk score (PRS). For SZ, the estimate of risk variance explained by PRS (Nagelkerke \( R^2 \)) is 0.184 (Schizophrenia Working Group of the Psychiatric Genomics, 2014). Although PRS is not of diagnostic utility for SZ at the moment, it has been useful for examining the validity of intermediate phenotypes and exploring genetic relationship between different neuropsychiatric disorders. Common variation contribution to SZ overlaps with autism spectrum disorder (ASD), obsessive-compulsive disorder, bipolar disorder, attention deficit hyperactivity disorder and major depressive disorder (O’Donovan and Owen, 2016). Furthermore, individuals with a higher SZ PRS have also been shown to perform more poorly on cognitive tasks (Riglin et al., 2017) and SZ PRS has also been associated with brain volume (Terwisscha van Scheltinga et al., 2013).
1.2.7 CNVs

In addition to common variants that individually exert weak effects on risk, the role of rarer variants that potentially exert a much greater effect on risk of neuropsychiatric disorders has now been recognised. In 1999 it was shown that adults with DiGeorge syndrome, a condition that results from a large deletion on chromosome 22q11.2, had an increased risk of developing SZ (Murphy et al., 1999). This is now recognised as the first example of a CNV that is associated with SZ risk. DiGeorge syndrome is a developmental disorder that affects 1 in 4,000 births, and an average of 40 genes are typically affected by the large deletion. (McDonald-McGinn et al., 2011). More recent genome-wide CNV studies have revealed that individuals with ASD and SZ harbour rare or de novo CNVs at a significantly increased rate compared to controls (Sebat et al., 2007, Kirov et al., 2011). In addition, CNVs were shown to be causative in approximately 14% of cases of idiopathic intellectual disability in children (Cooper et al., 2011). Due to the rare nature of pathogenic CNVs, large sample sizes are required to investigate their role in SZ. A study of 21,094 SZ cases associated 8 CNV loci with SZ risk at genome-wide significance, including the DiGeorge locus (Marshall et al., 2016). For many of these loci, multiple genes are deleted or duplicated. The exception is deletions of chromosome 2p16.3 which specifically disrupt NRXN1, a gene which encodes a synaptic cell-adhesion molecule. The effect sizes to these CNVs are much greater than common variants and the odds ratios, measuring risk, vary between 2 and 60 (Marshall et al., 2016), compared to approximately 1.1 or less for common variants (Pardiñas et al., 2018). However many of these CNVs have been identified in healthy controls so it appears that even for carriers of rare high risk CNVs, additional genetic and environmental factors influence phenotypic presentation (Tansey et al., 2015).

1.2.8 Exome sequencing

There has also been an effort to capture rare point mutations in SZ. This has been done through sequence studies that focus on the coding regions of the genome; the exome. The hypothesised benefits of these studies are: (1) exonic point mutations are associated with a single gene, (2) the potential consequences of mutations that affect the amino acid sequence are easier to predict and model and (3) rare coding mutations are likely to be associated with high odds ratios (Bray et al., 2018). On average, individuals have 1 -2 de
1.3 Cognition

General cognitive ability (g) is a quantitative trait, which has a normal distribution. There are diverse measures of cognitive ability; verbal ability, spatial ability, memory and information processing speed. These measures correlate substantially with one another and g is the commonality of these diverse measures (Plomin, 1999). g is reliable and stable following childhood and it predicts outcomes such as education and occupational level better than other behavioural traits (Gottfredson, 1997).

1.3.1 Cognitive dysfunction in schizophrenia

Cognitive dysfunction is a core symptom of SZ. Impairments in working memory, attention, processing speed, visual and verbal learning, reasoning, planning and problem solving have all been well documented in SZ (Bora et al., 2009, Heinrichs and Zakzanis,1998). It is estimated that 98.1% of patients with SZ fail to reach their expected level of cognitive function based on parental education and pre-morbid intelligence (Keefe et al., 2005). Cognitive deficits can even be observed in the prodromal phase of SZ, but are more pronounced during the first episode of psychosis, following which they remain stable throughout the condition even in the absence of positive symptoms (Woodberry et al., 2008). There is a significant association between functional capacity or the ability to function independently in daily life, and neurocognitive test scores (Green...
et al., 2015). Social cognition relates to the ability to process information carried by social stimuli and use this information to generate socially acceptable responses. Social cognition in patients with SZ is strongly linked to community functioning (Fett et al., 2011). Cognitive dysfunction strongly predicts rates of employment, independent living and interpersonal relationships in SZ (Bowie et al., 2010, Green, 2006). In addition to this self-perceived cognitive deficits in SZ patients are associated with internalised stigma and quality of life (Shin et al., 2016). In response to these studies that show that functional capacity is heavily influenced by cognitive function, cognitive remediation therapy (CRT) has been developed for patients with SZ. CRTs are behavioural training interventions that aim to improve cognitive functioning. A meta-analysis of CRTs concluded that small to moderate effects on cognitive outcomes were achieved after intervention completion and in follow-up assessments in patients with SZ (Wykes et al., 2011).

1.3.2 Schizophrenia and cognitive endophenotypes

The heterogeneous presentation of SZ and the symptomology overlap with other neuropsychiatric disorders makes it difficult to diagnosis, especially because diagnosis is based solely on clinical presentation. Intermediate quantifiable phenotypes are known as endophenotypes. They are heritable and can even precede clinical presentation of a condition. It was hypothesised that genetic variation associated with SZ risk may subtly impact the function of the encoded protein, and that these effects may be more readily observed in endophenotypic traits that have simpler genetics than SZ (Braff et al., 2007). Use of an endophenotype is appropriate in psychiatry when (1) it is heritable, (2) deficits in the trait are associated with the disorder, (3) it co-segregates with the disorder, (4) it is stable and (5) is present at higher rate in unaffected relatives than in the general population (Gould and Gottesman, 2006). Endophenotypes can be used to understand how SZ risk variations functionally contribute to the condition (Walters and Owen, 2007). Due to the quantitative nature of endophenotypes individuals can be ranked, this provide additional statistical power to detect genetic contributions to the endophenotype and therefore disease (Glahn et al., 2014).
Neurocognitive measures and neurophysiologic measures that exhibit large variations between patients with SZ and controls have been used as endophenotypes in SZ research (Gur et al., 2007). A number of the established neurocognitive measures relate to executive function. Executive functions are mental processes that control inhibition, working memory and cognitive flexibility which are crucial to efficient reasoning, problem-solving and planning (Diamond, 2013). The endophenotypes used in this thesis can be broken broadly into five domains of cognition (1) IQ, (2) episodic memory, (3) working memory, (4) attention and (5) social cognition. Working memory and attention are both considered to be executive functions.

(1) IQ or general cognitive ability is previously described in Section 1.3. The Wechsler Adult Intelligence Scale (WAIS) is the most commonly used scale to measure IQ. Version 3 (WAIS III) was developed in 1997 and includes measures of performance IQ, verbal IQ, full scale IQ as well as secondary tests measuring verbal comprehension, processing speed, and perceptual organisation (Wechsler, 1997a). Four subtests of the WAIS III battery have been included in this thesis; full scale IQ, performance IQ, verbal IQ and the Wechsler Test of Adult Reading (WTAR) (Wechsler, 2001) which together, take approximately 30 minutes to complete.

(2) Episodic memory is the collection of past personal experiences that occurred at particular place or time. It is a type of declarative or explicit memory which refers to recalling facts, data or events. This thesis includes measures of immediate and delayed episodic memory function. Four subtests in total have been included: Logical Memory I and II from WAIS III (Wechsler, 1997a) and Faces I and II from the Wechsler Memory Scale (WMS) III (Wechsler, 1997b) which measure visual memory. Individuals with SZ have pronounced deficits in episodic memory function (Dickinson et al., 2004).

(3) Working memory is the cognitive system that is responsible for the temporary holding of information available for processing. There are two types of working memory:
auditory (storing what you hear) and visual-spatial (storing what you see). Working memory aids in making connections between different perceptual inputs and is therefore critical for other cognitive processes such as reasoning and problem solving (Baddeley, 2010, Diamond, 2013). Working memory function is impaired in SZ patients (Frydecka et al., 2014) which may be caused by abnormal activity in the dorsolateral prefrontal cortex (DLPFC) (Van Snellenberg et al., 2016). Two measures of working memory have been included in this thesis: the spatial working memory task from the Cambridge Neuropsychological Test Automated Battery (CANTAB) (Cambridge Cognition, 2019) the Letter-Number Sequencing (LNS) task from WMS III (Wechsler, 1997b).

(4) Attention refers to how we actively process specific information in our environment. Attentional control depends on experience and on the associated stimuli, with less experience generally requiring more focus on a stimulus. Patients with SZ have difficulty focusing attention on salient cues and overcoming distracting stimuli (Braff, 1993). The Continuous Performance Task (CPT) (Cornblatt et al., 1988) and the Sustained Attention to Response Task (SART) (Robertson et al., 1994) are two measures of attention included in this thesis.

(5) Social cognition as mentioned earlier relates to the ability to process information carried by social stimuli and use this information to generate socially acceptable responses. Theory of mind is an important social cognitive tool which encompasses the ability to attribute emotions, desires and beliefs and to understand that others beliefs and thoughts may differ from your own (Brüne, 2005). Reading the Mind in the Eyes task was used in this thesis to evaluate theory of mind. Participants in this task attempt to infer complex mental states of people from pictures of their eyes (Baron-Cohen, 2001). The Hinting Task, was also used to examine theory of mind and is based on examining individuals’ ability to infer intentions behind direct speech (Corcoran, 1995). Attributional bias was also used as a measure of social cognition, it refers to how individuals infer the cause of negative and positive events. The Internal, Personal and Situational Attribution Questionnaire (IPSAQ) was used to measure attributional bias (Kinderman and Bentall, 1997). Paranoid ideation
Chapter 1

(Kinderman and Bentall, 1997) and persecutory delusions (Martin and Penn, 2002) in SZ have been associated primarily with external-personal attributions for negative events (patients with SZ are more likely to attribute negative events to the actions of others).

1.3.3 Genetics of cognition

Heritability estimates of general cognition are age-dependent. Heritability of intelligence increases linearly from ~ 20% in infancy to 40% in adolescence and 60% in adulthood (Haworth et al., 2010). Heritability of intelligence can reach ~ 80% in later adulthood, but it once again declines to 60% after age 80 (Lee et al., 2010). Genetic correlations between different measures of cognition, such as spatial cognition and vocabulary, are high which suggests that the same genes are influencing the heritability of these tests. Assortative mating is the phenomenon whereby individuals with similar phenotypes mate more frequently than would be expected. Assortative mating for intelligence is far greater than for height, weight or personality (Plomin and Deary, 2014).

A recent large GWAS (n=300,486), found 148 genome-wide significant independent loci associated with general cognitive ability (g) (Davies et al., 2018). The SNP-based heritability of g, in this study was 0.25. Many of the variants identified have been previously linked to neurodegenerative disorders such as Alzheimer’s and Parkinson’s, physical traits such as height and BMI and psychiatric conditions such as SZ, ASD and bipolar disorder. General cognitive function is genetically correlated to education (discussed in Section 1.4). A novel approach of multi-trait analysis of genome-wide association studies (MTAG) has been used to combine results from large GWAS of education and intelligence, increasing statistical power. This approach found 187 independent loci associated with intelligence, implicating 538 genes (Hill et al., 2019). This analysis implicated genes involved in myelination, neurogenesis, the regulation of the nervous system and synaptic genes in intelligence (Ganna et al., 2016).

Educational attainment (EA) refers to the number of years in education that an individual has completed. EA positively correlates with IQ (Piffer, 2019) and other cognitive traits including verbal-numeric reasoning, memory and reaction time (Davies et al., 2016). The
large sample size for which EA data are available and easily accessible in many
populations makes it a suitable proxy phenotype for GWAS of cognitive traits. The
largest GWAS to date analysed data from approximately 1.1 million individuals,
identifying 1,271 independent genome-wide significant SNPs (Lee et al., 2018). These
SNPs implicate genes that regulate brain development processes and neuronal
communication.

1.3.4 Genetic overlap between schizophrenia and cognition

An inverse relationship between cognitive ability and SZ risk has been repeatedly
observed at the genetic level, suggesting high IQ has a protective effect against SZ
(Smeland et al., 2017, Lencz et al., 2014). Somewhat counterintuitively genetic studies
have revealed a positive relationship between EA an SZ risk (Hill et al., 2019, Okbay et
al., 2016). A pleiotropic meta-analysis analysis of cognition, EA and SZ, found that there
are subsets of SNPs that are associated with greater risk for SZ, lower cognitive ability
and lower EA. Genes affected by these “concordant” SNPs were enriched in early
neurodevelopmental pathways. There was also a subset of “discordant” SNPs were
greater SZ risk was associated with higher EA, these SNPs are within genes that are
enriched for adulthood synaptic pruning pathways (Lam et al., 2019). It is possible that
efficient synaptic pruning mechanisms are critical to academic performance however
enhanced synaptic pruning may contribute to SZ pathophysiology as discussed
previously.

1.4 Centrosomes

The observation of the centrosome was first published in 1887, completely independently
but almost simultaneously by Theodor Boveri and Edouard van Beneden, both of whom
were studying the mitotic divisions of fertilised eggs from Parascaris equorum, a
roundworm parasite. In these early publications, the centrosome was recognised as a
permanent cell organelle that was capable of self-replication and was passed onto
daughter cells. Boveri also commented that the centrosome was the dynamic centre of
the cell, responsible for its organisation particularly during cell division (Van Beneden,
1887, Boveri 1887, reviewed by Scheer, 2014). Boveri observed a denser granule within
the centre of the centrosome, for which he went on to coin the term ‘centriole’ (Boveri 1888, Boveri, 1900). The centrosome is now known to act as the major microtubule organizing centre in animal cells. In quiescent cells, the mother centriole can also seed the growth of a primary cilium, a ‘sensory antenna’ for the cell, which is discussed in detail in Section 1.5. Centrosome and ciliary dysfunction have now been implicated in a number of conditions, including cancer, where an increase in centrosome number has been shown to promote tumorigenesis (Levine et al., 2017, Basto et al., 2008) and the ciliopathies, developmental and degenerative disorders caused by mutations that affect ciliary structure and function (Hildebrandt et al., 2011). This has led to a renewed interest in centrosome biology in recent years.

1.4.1 Centrosome structure

1.4.1.1 Centrioles

The core components of the centrosome are the centrioles. The configuration of these centrioles is dependent on cell cycle stage, which is discussed in detail in Section 1.4.3. The mitotic centrosome consists of two orthogonally arranged centrioles, a daughter and a mother, which are both surrounded by pericentriolar material (PCM). The organelle is not membrane bound, therefore the PCM defines its boundaries. Each centriole consists of nine triplet microtubules arranged into a cylinder (Figure 1.1). In mammals, each triplet microtubule is composed of a complete microtubule (A-tubule) and two partial microtubules (B- and C-tubules). Projections from each A-tubule link it to the neighbouring C-tubule (Li et al., 2012). The microtubules confer a polarity to the centriole structure: the plus ends, from which the microtubules extend, are positioned at the distal end of the centriole (Winey and O’Toole, 2014). Centriolar microtubules are very stable; they are resistant to cold treatment as well as to microtubule-destabilising drugs and have a very slow turn-over rate (Kochanski and Borisy, 1990). Centriolar microtubules are extensively modified, with post-translational modifications present on tubulin subunits include; glutamylation, detyrosination and acetylation. Glutamylation of actin has been shown to be important for centriole stability (Bobinnec et al., 1998). Tubulin acetylation is a marker of stable MTs and its removal leads to destabilisation of
axonemal MTs in the primary cilia (Ran et al., 2015); further studies are required to dissect its importance at the centriole.

![Figure 1.1: Centrosome and centriole structure](image)

**Figure 1.1: Centrosome and centriole structure**

A. Schematic of a typical centrosome during G1. Each of the centrioles is composed of nine MTs. The daughter centriole lacks maturity markers, distal and subdistal appendages. Both parental centrioles are embedded in the PCM and are connected by flexible fibrous linkers. B. Cross-sectional view of an immature mammalian centriole (procentriole), highlighting the cartwheel structure (red) contained within the lumen of the centriolar MTs (green). The cartwheel structure is absent from mature mammalian centrioles.

In mammalian cells, within the lumen created by the microtubule scaffold of immature centrioles, a cartwheel structure can be visualised (Figure 1.1 B, Guichard et al., 2018). The cartwheel consists of a central hub, extending radially from which are nine spokes, which contact the A-microtubule of each triplet microtubule. The cartwheel appears at the initial centriole assembly process; multiple cartwheels are usually stacked within the centriole, the exact number varying between organisms (Hirono, 2014). SAS-6 is the major component of the hub and central part of the nine spokes and is important for cartwheel stabilisation and establishing the 9-fold symmetry of the centriole (Nakazawa et al., 2007). In line with this role as a scaffold for centriole assembly, SAS-6 has been shown to be essential for centriole biogenesis and duplication (Strnad et al., 2007). Mutations in the gene that encodes SAS-6 cause primary microcephaly (Khan et al., 2014). In mammals, the cartwheel structure has only been observed in immature centrioles.
1.4.1.2 Pericentriolar material and satellites

The PCM is an electron-dense assembly of proteins that surrounds the centrioles, a platform for protein complexes that regulate centrosome function. The PCM is important for organelle trafficking and protein degradation and the nucleation and anchoring of microtubules (MT). Some key PCM proteins include, CDK5RAP2, CEP152, CEP192, CPAP, and γ-tubulin (Woodruff et al., 2014). Many of these PCM proteins contain a high degree of coiled-coil domains, which facilitate protein interactions. Ablation of PCM proteins has been shown to cause MT instability, reduced PCM, centrosome separation and centriole duplication defects (Woodruff et al., 2014). γ-tubulin is a key PCM component that is not incorporated into microtubules but instead forms ring complexes termed γ-tubulin ring complexes (γ-TuRCs) (Zheng et al., 1995). In animal cells, the core component of γ-TuRCs is the γ-Tubulin Small Complex (γ-TuSC), which is composed of single molecules of GCP2 and GCP3 and two molecules of γ-Tubulin (Murphy et al., 1998). Three additional proteins- GCP4, 5, and 6- can complex with the γ-TuSC to form a γ-TuRC. NEDD1, while not a part of the γ-TuRC complex, plays a crucial role in tethering the complex to the centrosome (Haren et al., 2006). γ-TuRCs have been demonstrated to cap the minus ends of PCM derived microtubules and nucleate microtubule assembly (Moritz et al., 2000). They are therefore crucial to the MTOC function of the centrosome, as discussed in Section 1.4.2.1.

During interphase, CEP152, CEP192, pericentrin and CdkRap2 have been visualised as concentric rings with varying sizes around the centrioles (Fu and Glover, 2012, Mennella et al., 2012, Lawo et al., 2012, Sonnen et al., 2012). During mitosis, the PCM lacks a uniform organization. As the cell cycle progresses, centrosome maturation occurs and there is an expansion of the PCM in preparation for spindle pole formation. Centrosome maturation is dependent on PLK1, which phosphorylates a number of PCM components (Lee and Rhee, 2011). This phosphorylation promotes increased recruitment of the PCM components, pericentrin, CdkRap2 and γ-Tubulin, and the formation of phase-separated condensates that have enhanced capacity for microtubule nucleation (Woodruff et al., 2017). Depletion of PLK1 leads to an inhibition of centrosome maturation and defects in the assembly and maintenance of bi-polar spindles (Santamaria et al., 2007, Peters et al., 2006, Lane and Nigg, 1996). Centrioles also play a crucial role in organising the PCM
and when they are disassembled by loading cells with a monoclonal antibody that specifically reacts with polyglutamylated tubulin, a scattering of PCM occurs (Bobinnec et al., 1998).

Centriolar satellites are small spherical granules that cluster around the centrosome and basal body of ciliated cells. Their size, composition and structure are variable and they are not observed during mitosis. The satellites are MT-anchored structures that play an important role in transporting proteins to the centrosome to regulate its function (Kubo et al., 1999). Centriolar satellites proteins are essential for ciliogenesis, MT organisation, spindle pole maintenance, centrosome duplication, dendritic patterning and neuronal development (Tollenare et al., 2015). The best characterised component of centriolar satellites is PCM1. This protein is predicted to contain eight coiled-coil domains, which mediate many of its protein-protein interactions. PCM1 is thought to function as a scaffold for the recruitment of other satellite components and is required for the assembly of centriolar satellites and ciliogenesis (Wang et al., 2016). Knockdown of PCM1 leads to reduced centrosomal concentrations of pericentrin, the centriolar component centrin and the subdistal appendage protein ninein, and defects in MT anchorage at the centrosome and in MT network establishment (Dammermann and Merdes, 2002).

1.4.1.3 G1-G2 tether

In addition to undergoing centriole maturation as the cell cycle progresses, centrioles also lose their proteinaceous tether or linker, which physically connects the centrioles during interphase (Bornens et al., 1987). Proteins that have been localised to this linker include rootletin, CEP68, LRRC45 and centlein (Fang et al., 2014, He et al., 2013, Graser et al., 2007b). C-NAP1 localises to the proximal ends of centrioles (Fry et al., 1998), and recruits rootletin (Bahe et al., 2005). Nek2 phosphorylation of C-Nap1, rootletin, Cep68, LRRC45 and centlein initiates their displacement from the centrosome, resulting in centriole separation. Depletion of any of these proteins results in a loss of centrosome cohesion (Flanagan et al., 2017, Man et al., 2015, Fang et al., 2014, Hardy et al., 2014, He et al., 2013, Bahe et al., 2005).
1.4.1.4 Centriole appendages

The mother centriole is the older, more mature centriole, and can be recognised by the presence of distal and subdistal appendages (Vorobjev and Chentsov, 1982). The distal appendages (DA) are pin-wheel like structures protruding from the distal end of the mother centriole. DAs promote the docking of the centriole to the cell membrane, which is essential for ciliogenesis (Tanou et al., 2013). Proteins that localise to the backbone of the pinwheel blades include Cep164, Cep83, SCLT1 and Cep89 (Yang et al., 2018a), all of which are essential for primary ciliogenesis initiation (Yang et al., 2018a, Graser et al., 2007a, Sillibourne et al., 2013, Tanou et al., 2013). Recent studies have also identified a matrix between the DA blades, to which FBF1 and intraflagellar transport (IFT) component IFT88 have been localised. Ablation of FBF1 leads to a reduction in ciliary capacity, as trafficking to the ciliary membrane is impaired suggesting that the matrix surrounding the DAs may play a role in ciliary gating (Yang et al., 2018a). Subdistal appendages (SDAs), have a rounded head and conical sometimes striated stem, with their base attaching to 2-3 MT triplets. Their number, thickness and distribution are highly variable and they can be spaced along the length of the mother centriole. They are sites of MT nucleation and in line with this, γ-tubulin has been localised to their heads and they can also anchor MTs to the centrosome (Delgehyr et al., 2005). Other SDA proteins include ninein, ε-tubulin and Cep170 (Guarguaglini et al., 2005, Chang et al., 2003). Ninein is the major contributor to the MT anchorage function of the centrosome (Abal et al., 2002). It has been shown to influence neurogenesis and stem cell fate and mutations that affect ninein function result in Seckel syndrome, characteristic features of which are microcephalic primordial dwarfism, cognitive deficits and increased sensitivity to DNA damaging agents (Dauber et al., 2012, Wang et al., 2009a, Ohama et al., 2009). ε-tubulin has been shown to be important for basal body morphogenesis in Chlamydomonas and for the anchorage and assembly of centriolar MTs in Paramecium (Ross et al., 2013, Dupuis-Williams et al., 2002). Depletion or overexpression of Cep170 results perturbs MT cytoskeletal structure and cell shape (Guarguaglini et al., 2005). Depletion of Cep170 also increases cillum length and the proportion of ciliated cells (Zhang et al., 2019), suggestive of a role in ciliary disassembly. Odf2 initiates the assembly of SDA and depletion of the protein disrupts the localisation of other SDA components, so that Odf2 also plays a crucial role in ciliogenesis (Huang et al., 2017, Ishikawa et al., 2005).
1.4.2 Centrosome functions

The centrosome is highly dynamic and serves as a platform for a number of key functions. The principal roles associated with centrosome include its function as the microtubule organising centre (MTOC) and its ability to function as a basal body to seed the growth of a primary cilium. Additional roles for the centrosome are being identified as the number of proteins that have been localised to the centrosome increases and their identities are unmasked. Recent research suggests that the centrosome is important for proteasome functioning (Vora and Phillips, 2016) and is an important signalling hub (Mullee and Morrison, 2016, Arquint et al., 2014).

1.4.2.1 Microtubule-organising centre (MTOC)

The centrosome acts as the principal MTOC of the animal cell and has the capability to nucleate, anchor and release MTs. Microtubules play critical roles in cell division, intracellular transport, cell shape, polarity and motility (Wiese and Zheng, 2006). The PCM is particularly important for the nucleation and anchoring of MTs. As mentioned above, γ-TuRCs have a key role to play in the nucleation of MTs. Purified γ-TuRCs have an enhanced ability to nucleate microtubules compared to purified γ-TuSCs (Oegema et al., 1999). Multiple studies have highlighted the essential roles of γ-TuRC proteins in human cells. When these components are disrupted, cells have defective mitotic spindle and microtubule arrays (Choi et al., 2010, Bahtz et al., 2012, Scheidecker et al., 2015). However, these studies also demonstrated that some MTs can be nucleated in the absence of γ-TuRCs, suggesting that γ-tubulin independent mechanisms of MT nucleation exist (Roostalu et al., 2015, Sampaio et al., 2001). Mutations in the γ-TuRC components GCP4 and GCP6 have been shown to cause microcephaly (Cota et al., 2017, Scheidecker et al., 2015, Martin et al., 2014) and mutations to γ-tubulin causes complex cortical dysplasia with other brain malformations-4 (Poirier et al., 2013). Centrioles have also been shown to be essential for timely mitotic spindle assembly and their absence leads to chromosomal instability (Bazzi and Anderson, 2014, Sir et al., 2013).

In interphase cells, MTs are anchored to the centrosome. The minus ends interact with components of subdistal appendages, including Ninein and ODF2 (Ishikawa et al., 2005,
Delgehyr et al., 2005). The components of the PCM have also been shown to play a role in MT anchoring (Dammermann and Merdes, 2002), as has FSD1, a recently-described centriolar component (Tu et al., 2018). Disruption of MT anchoring proteins causes defects in ciliogenesis (Tu et al., 2018, Wang et al., 2016, Ishikawa et al., 2005), suggesting that the MTs that are anchored to the centrosome are important for ciliogenesis. Some cells grow submerged cilia that are predominately confined within the ciliary pocket (discussed in section 1.5.1). Double mutants of C-NAP1 and the genes that encode the MT-anchoring protein OFD2 or SDA component CEP128, cause cilia to surface completely in these cells (Mazo et al., 2016), suggesting that MT anchoring contributes to the spatial control of ciliogenesis.

Release of microtubules from the centrosome is also important, particularly for migrating cells, where an over-expression of ninein has been shown to significantly impair migration (Abal et al., 2002). Centrosome-derived MTs are also rapidly released from differentiating neurons and contribute to axonal outgrowth (Yu et al., 1993, Ahmad et al., 1999). Taken together, these studies highlight the importance of the centrosome as the MTOC, in ciliogenesis, migration, differentiation and in ensuring that cell division is of high fidelity in animal cells.

1.4.2.2 Centrosomes as basal bodies

In cells that are not actively cycling, either due to nutrient deprivation or differentiation, the mother centriole can seed the growth of a primary cilium (Tucker et al., 1979), a sensory organelle which will be discussed in more detail in Section 1.5 (Singla and Reiter, 2006). Many centrosome proteins are crucial for ciliogenesis, including those of the distal appendages which, as discussed earlier, are essential for docking the centrosome to the plasma membrane (Tanos et al., 2013). In addition to DA proteins, disruption to a further 33 mother centriole proteins resulted in impaired ciliary function (Reiter and Leroux, 2018), highlighting the crucial role the basal body plays in ciliogenesis (discussed further in Section 1.5.3). Pericentrin is also required for ciliogenesis (Jurczyk et al., 2004) and multiple satellite components have also been shown to influence ciliogenesis, including PCM1 (Wang et al., 2016) and OFD1 (Tang et al., 2013). Many satellite components also localise to the centrosome and it remains unclear whether their function at the satellites
is essential for ciliogenesis. Centrosomes also serve as basal bodies for motile cilia and flagella, which function to facilitate cell movement and to direct fluid flow.

1.4.2.3 Centrosomes as signalling hubs

The centrosome has been hypothesised to be a scaffold that mediates degradation via the ubiquitin-proteasome system. Proteasomal complexes have been localised to the centrosome (Wigley et al., 1999), and purified centrosomal fractions are capable of degrading ubiquitinated substrates, similar to active proteasomes (Fabunmi et al., 2000). In addition to this, large protein aggregates that associate with the centrosome have been observed in cells that have been treated with proteasome inhibitors, called aggresomes (Johnston et al., 1998). Centrosome-associated degradation has been shown to have a role in determining neuronal morphology (Puram et al., 2013, 2010, Kim et al., 2009). Centrosome-associated degradation has also been associated with degradation of cell-cycle regulators (Máthé et al., 2004, Raff et al., 2002).

Several important cell cycle regulators have been localised to the centrosome, including Cyclins E and A, which are crucial for centriole duplication and the replication of DNA. These cyclins physically interact with and activate cyclin dependent kinases (CDKs) 1 and 2 (Pascreau et al., 2010, Matsumoto and Maller, 2004). CDK1 and CDK2 have a critical role in normal cell-cycle progression (Hanse et al., 2009, Santamaria et al., 2007). The cyclin B-CDK1 complex is the key initiator of mitosis and has been shown to be activated during prophase at centrosomes (Jackman et al., 2003). Other regulators of mitosis, including PLK1, Aurora A and CDC25C, have all also been shown to localise to the centrosome (Dutertre et al., 2004, Golsteyn et al., 1995, Debec and Montmory, 1992). The centrosome may accelerate entry into mitosis by increasing the local concentrations of these regulatory factors.

The centrosome has also been postulated to function in the DNA damage response (DDR). Several regulators of DDR have been localised to the centrosome including BRCA1 and 2, p53, CHK1 and 2, ATM and ATR (Chouinard et al., 2013, Tarapore et al., 2012, Löffler et al., 2007, Nakanishi et al., 2007, Tritarelli et al., 2004, Krämer et al., 2004).
p53 is a tumour suppressor that controls cell-cycle progression, DNA repair and apoptosis. Depletion of p53 leads to centrosome amplification (CA), an increase in the number of centrosomes (Fukasawa et al., 1996), and the protein has been shown to regulate the appropriate timing of centrosome separation (Nam and Deursen, 2014). Upon DNA damage, p53 localises to the centrosome, suggesting that this translocation may be important for initiating an appropriate DDR (Tritarellie et al., 2004). The product of the tumour suppressor gene BRCA1 plays a crucial role in maintaining genome stability. BRCA1 can exist in an ubiquitin E3 ligase complex, activity of which modulates centrosomal MT nucleation and centrosome number (Sankaran et al., 2005, Starita et al., 2004). Additional information on regulators of the DDR at the centrosome is reviewed by Mullee and Morrison, 2016. Several centrosomal proteins have been linked with the DDR. Centrin 2, a centrosomal protein, is required for efficient ciliogenesis (Prosser and Morrison, 2015). Centrin 2 also interacts with xeroderma pigmentosum group C protein, a core component of the nucleotide excision repair (NER) machinery (Araki et al., 2001). Knockdown of Centrin 2 in chicken cells has been shown to delay NER (Dantas et al., 2011). Similarly, mutations in \( PCNT \), which encodes the PCM component pericentrin, cause Seckel syndrome, which is characterised by a gross reduction in brain and body size, due to an impairment in ATR-dependent DNA damage signaling (Griffith et al., 2007). In addition to this, centrosome splitting, where the daughter and mother centriole lose cohesion, as well as CA, have both been observed in response to DNA damage (Inanç et al., 2010, Bourke et al., 2007, Dodson et al., 2004). The cellular function of CA is unclear at this point.

1.4.3 Centrosome cycle

The centrosome, like chromosomal DNA, is duplicated once per cell cycle. This duplication is tightly controlled and linked to cell cycle progression (Figure 1.2). When there is a disruption to this homeostasis through defective cell divisions, centrosome fragmentation or centriole overduplication, CA may occur (Duensing et al., 2007, Nigg, 2006, Salisbury et al., 2004). CA has been shown to increase cell invasion (Armandis et al., 2018) and chromosomal instability through the formation of multipolar spindles (Ganem et al., 2009, Cimini, 2008). CA has been observed frequently in many different types of cancer (Chan, 2011) and has even been shown to be sufficient to drive
tumorigenesis (Levine et al., 2017, Basto et al., 2008). This highlights the importance of maintaining a high fidelity centriole duplication mechanism, which is discussed below.

Figure 1.2: The centrosome cycle
Schematic illustration of the major events during centrosome cycle at the indicated phases of the cell cycle. At the M/G1 transition, the mother and daughter centrioles are licensed (formation of ‘G1-G2 tether’) for procentriole nucleation during S-phase (centriole disengagement). During S-phase, two new procentrioles are assembled on cartwheel structures at the proximal ends of the parental centrioles (procentriole biogenesis). The perpendicularly arranged procentrioles are attached to their respective parental centrioles by a flexible ‘S-M linker’, and they continue to elongate throughout S and G2 phases (procentriole elongation). At this stage, the parental centrioles are still connected by the intercentriolar linker. During G2, the two centrosomes recruit additional PCM and the new mother centriole acquires distal and subdistal appendages (centrosome maturation). Prior to mitosis, the fibrous G1-G2 tether is degraded (centrosome separation) and the two centrosomes migrate to form the bipolar spindle during mitosis.
1.4.3.1 Centriole disengagement

At the end of mitosis, individual centrosomes are positioned near the nuclear envelope. The mother and daughter centrioles need to physically disengage in order to allow centriole duplication in preparation for the next round of mitosis (Tsou and Stearns, 2006). A cohesion complex is localised to engaged centrioles (S-M linker) and must be cleaved by separase in co-operation with PLK1 to trigger centriole disengagement (Schöckel et al., 2011, Tsou et al., 2009). It has been shown that separase must also target pericentrin for successful centriole disengagement. Pericentrin functions to protect engaged centrioles from premature separation (Lee and Rhe, 2012, Matsuo et al., 2012). Astrin and AKI1 have also been shown to be important inhibitors of premature separation (Nakamura et al., 2009, Thein et al., 2007), which is crucial in preventing inappropriate reduplication of centrioles (Loncarek et al., 2008). Following centriole disengagement, the centriolar linker tethers the two centrioles together until entry into mitosis.

1.4.3.2 Procentriole formation

The disengagement of mother and daughter centrioles allows for the duplication of new centrioles which begins at the initiation of S phase. Procentrioles are nucleated at the base of existing centrioles, where they remain engaged until the completion of mitosis and centriole separation occurs in G1. PLK4 is considered to be the master regulator of centriole duplication. Over-expression of Plk4 causes an increase in the number of centrioles (Martins et al., 2007, Habedanck et al., 2005), while its depletion prevents centriole duplication (Bettencourt-Dias et al., 2005), both of which result in mitotic errors. The cartwheel appears at the initiation of procentriole formation, and serves as a foundation for centriolar MTs. PLK4 phosphorylates STIL, which enables STIL to interact with SAS-6, triggering centriolar loading of SAS-6 (Kratz et al., 2015, Ohta et al., 2014). As discussed in Section 1.4.1.1, SAS-6 is the main component of the central hub and spokes of the cartwheel and therefore its recruitment is crucial to procentriole biogenesis. The nine microtubule triplets then form a cylinder around the cartwheel. Similar to the nucleation of MTs at the centrosome, the formation of A-tubules in the centriolar triplet microtubules is dependent on γ-tubulin, while C-tubule formation
requires δ-tubulin and B and C-tubule formation requires ε-tubulin (Winey and O’Toole, 2014).

The centrioles are then elongated from the proximal to distal end until they reach their final length in G2. CPAP is a key regulator of centriole elongation that has been shown to stabilize the cartwheel and recruit MTs to facilitate elongation. Overexpression of CPAP results in elongated centrioles (Kohlmaier et al., 2009, Schmidt et al., 2009, Tang et al., 2009). CPAP complex formation with SAS-6 and STIL is important for its role in elongation (Tang et al., 2011, Cottee et al., 2013). OFD1 and POC5 are also important for the regulation of centriole elongation (Singala et al., 2010, Azimzadeh et al., 2009). The capping protein CP110 and its binding partner Cep97 play an important role in determining the final length of centrioles and depletion of CP110 or Cep97 results in increased centriolar lengths (Spektor et al., 2007).

1.4.3.3 Centrosome maturation

Centrosome maturation occurs during the G2/M transition of the cell cycle and is marked by an increase in PCM volume and the addition of sub-distal and distal appendages to the mother centrioles (Fujita et al., 2015). These changes are accompanied by an increase in microtubule nucleation. PLK1 plays a crucial role in PCM expansion, facilitating the efficient recruitment of pericentrin, NEDD1, γ-Tubulin, CEP215 and HSP70 to the PCM (Lee and Rhee, 2011, Fang et al., 2019). Depletion of PLK1 disrupts the recruitment of many of these key PCM MT nucleating factors (Haren et al., 2009). PLK1-dependent modification of daughter centrioles is essential for their ability to recruit PCM proteins (Wang et al., 2011) and constant PLK1 activity is required during mitosis to maintain the mitotic centrosomal structure (Mahen et al., 2011). Over-expression of some PCM components, for example pericentrin, is sufficient to drive PCM expansion, suggesting that accumulation of PCM is possible without PLK1 (Lawo et al., 2012). However, PLK1 is thought to drive rapid and spontaneous PCM assembly (Wueseke et al., 2016). Over-expression of PLK1 has been shown to accelerate centrosome maturation (Kong et al., 2014).
Although the mechanisms behind SDA and DA assembly have yet to be fully elucidated, a hierarchical structure for distal appendage formation has been developed (Tanos et al., 2013). CEP83 is recruited to DAs early on and facilitates the binding of SCTL1 and CEP89. Subsequent recruitment of CEP164 and FBF1 is dependent on SCTL1 (Figure 1.3 A). A hierarchical structure for subdistal appendage assembly has also been put forward (Huang et al., 2017) whereby OFD2 initiates SDA assembly, recruiting TCHP, CCDC68 and CCDC120. Ninein can then be recruited by TCHP and CCDC120, and CEP170 can be recruited by ninein, TCHP, CCDC120 and CCDC68 (Figure 1.3 B). In addition to PLK1’s role in PCM expansion, it has also been shown to be important in appendage assembly.

Figure 1.3: Hierarchal assembly of distal and subdistal appendages
A. Schematic model of distal appendage assembly, adapted from (Tanos et al., 2013) B. Schematic model of subdistal appendage assembly, adapted from (Huang et al., 2017).

1.4.3.4 Centrosome separation

Centrosome separation occurs at the G2/M transition and is dependent on NEK2 dependent degradation of the proteinaceous linker, as well as force generated by Eg5 at the spindle poles. Rootletin, CEP68, LRRC45 and centletin have all been shown to be part of the proteinaceous linker that connects centrosomes during interphase (Fang et al., 2014, He et al., 2013, Graser et al., 2007, Bahe et al., 2005). C-NAP1 localises to the proximal ends of centrioles and is important for docking rootletin (Fry et al., 1998). NEK2 has been shown to phosphorylate C-NAP1, rootletin, Cep68, LRRC45 and centletin. NEK2-dependent phosphorylation can dissociate C-NAP1 and rootletin from the linker, which most likely precipitates centrosome separation at mitosis (Hardy et al.,
Regulation of NEK2 is therefore very important for appropriate linker dissolution. PLK1 phosphorylates Mst2, which in combination with hSav1, regulates NEK2 location and activity (Mardin et al., 2011, 2010). In addition to this, pericentrin anchors NEK2 to the centrosome and dampens its activity (Matsuo et al., 2010). Once the linker has been broken down, the centrosomes begin to move in opposite directions in order to form a bipolar spindle for mitosis. The kinesin Eg5 generates much of the force that is necessary to separate the centrosome, in a PLK1- and microtubule-dependent manner (Mardin et al., 2012, Smith et al., 2011).

1.5 Primary Cilia

The primary cilium is a non-motile structure that extends from the cell surface. The majority of human cells have been shown to be have ciliary capacity. A single primary cilium extends from the cell during cell cycle arrest (Mirvis et al., 2018). This antenna-like structure is capable of sensing both mechanical and chemical stimuli in the extracellular space. The primary cilium has been shown to co-ordinate a number of signalling pathways crucial for normal development and the maintenance of cellular homeostasis (Wheway et al., 2018). The structural integrity of the organelle is crucial to its functioning and genetic disruption of a large number of ciliary genes leads to a group of disorders known collectively as ciliopathies (Hildebrandt et al., 2011).

1.5.1 Primary cillum structure

The axoneme is a cylindrical array of nine microtubule doublets, which extend from the triplet MTs within the basal body (Figure 1.4).
Figure 1.4: Structure of the primary cilium
Schematic showing the structure of a basal body within the centrosome, the transition zone, the ciliary pocket, and the ciliary axoneme. The distal and subdistal appendages become transition fibers and basal feet, respectively. The ciliary membrane is continuous with the plasma membrane, but it is functionally distinct. Y-links and the TFs make up the transition zone that controls protein and lipid movement into and out of the primary cilium (see text for further details).

Each MT doublet consists of a complete A-tubule and partial B-tubule. The plus end of the microtubule is oriented to the tip of the primary cilium (Mirvis et al., 2018). Within each tubule, αβ heterodimers have been shown to interact vertically, creating a B-type lattice. In primary cilia, the lumen formed by the microtubule doublets is hollow. IFT motor kinesin-17 and α-tubulin have been observed within the lumen. IFT is the bi-directional motor dependent cargo transport system used to assemble primary cilia and is also important in maintaining their function. IFT requires anterograde (IFT-B) and retrograde complexes (IFT-A) (Ishikawa and Marshall, 2017). This pool of IFT components and actin may be a result of the depolymerisation of axoneme MTs during disassembly or a reserve which is ready for assembly and transport to the ciliary tip (Luo et al., 2017). The axoneme is extremely rigid and resistant to bending. This is partially due to the hollow tubular structure but also thought to play a role are αβ tubulin isoforms, post-translational modifications of tubulin and microtubule associated proteins (MAPs).
α and β isoforms of tubulin differ mainly by length, amino acid composition and affinity for post-translational modification (Janke, 2014). Loss of α-tubulin isotype tba-6 in sensory cilia from *C. elegans* resulted in abnormal axoneme ultrastructure and impairments in cargo transport, highlighting the importance of specific isotypes (Silva et al., 2017). Furthermore, motor proteins have varying affinities for different tubulin isoforms and therefore isoform composition can be a mechanism to regulate motor proteins (Girotra et al., 2017). Like centriolar MTs, axonemal MTs are targets for a range of PTMs. Acetylation is a marker of stable MTs, and appears to promote ciliogenesis, but is not essential. However, it is important for the mechanosensory function of some primary cilia (Aguilar et al., 2014, Shida et al., 2010). MT deacetylation has been shown to be important in primary cilium disassembly (Ran et al., 2015, Pugacheva et al., 2007). Glycylation is a positive regulator of ciliogenesis (Rocha et al., 2014, Wloga et al., 2009), while glutamylation is thought to negatively regulate ciliary extension (Kubo et al., 2015). Some MAPs have been shown to influence stability and dynamics of axoneme MTs. Depletion of one such protein, MAP4, leads to elongated cilia (Ghossoub et al., 2013), while depletion of another, SAXO1, promotes axoneme elongation by stabilising MTs (Dacheux et al., 2015).

The axoneme is ensheathed by a ciliary membrane, which is contiguous with the plasma membrane, but has a distinct composition. Arl13b is membrane-associated G-protein that localises to the ciliary membrane and functions in transport and signalling (Larkins et al., 2011). Arl13b has been shown to anchor the smoothened receptor (SMO), an important component in Hh signalling, and the Somatostatin receptor-3 (SSTR3). These ciliary membrane proteins are targeted to the primary cilium via Arl3b interaction with axonemal tubulin (Revenkova et al., 2018). There are two functional barriers that divide the plasma and ciliary membrane - the ciliary necklace and the ciliary pocket. The ciliary necklace is a unique ciliary membrane structure that contains multiple rows of intramembranous particles, generally arranged in scalloped rows (Gilula and Satir, 1972). The second barrier is the ciliary pocket, an invagination that forms a double membrane encircling the base of the cilium. The base of this pocket interacts with transitional fibers that extend from the basal body. The pocket is characterised by the presence of budding clathrin-coated pits, in line with its role in endocytosis (Molla-Herman et al., 2010) and similar to what has been demonstrated for the flagellar pocket (Gadelha et al., 2009).
Actin has also been shown to dock to the ciliary pocket (Molla-Herman et al., 2010). The formation of the pocket is most likely due to the fusion of a vesicle that contained the cilium with the plasma membrane (Sorokin, 1962), discussed in more detail in Section 1.5.3.

The base of the cilium consists of the basal body. The distal appendages mature into transition fibers (TFs) when the mother centriole is converted to a basal body. TFs are a 9-bladed propeller structure and their anchoring point at the membrane defines the border between the ciliary and cell membrane. Similar to DAs, Cep164, Cep83, Cep89, SCLT1 and FBF1 have been shown to localise to TFs, and Cep83 is important in the targeting of the other proteins (Tanos et al., 2013). The Cep164 interactors Cby and TTBK2 have been shown to localise to TFs and they also have a role in regulating ciliogenesis. TFs play a role in membrane docking but also are required to initiate elongation of the axoneme (Čajánek and Nigg, 2014, Burke et al., 2014). It is unknown whether TFs act as a diffusion barrier as their ablation completely prevents ciliogenesis, making it difficult to determine their role as a functional barrier. Subdistal appendages become known as basal feet, conical structures that extend laterally and interact with cytoskeletal MTs (Albrecht-Buehler and Bushnell, 1980) increasing the structural integrity of the primary cilia. As mentioned previously, the subdistal appendages play an important role in anchoring centrosome derived MTs and ciliogenesis. Primary cilia also have numerous striated rootlets, collections of filamentous structures that radiate from the proximal end of the basal body. They have been found to be orientated in all directions appearing as arched, looped, and branched around the cytoplasm (Hagiwara et al., 2002). Rootletin is an important component of these rootlets. They are thought to play a role in transport to the primary cilia, and also in anchoring the primary cilium (Yang et al., 2005, Farris et al., 1998, Kobayashi and Hirokawa, 1988).

The transition zone (TZ) is just distal to the basal body and compartmentalises the axoneme from the cytoplasm. The TZ is characterised by the presence of Y-linked fibers that connect the ciliary membrane to the axoneme and the ciliary necklace. The main components of the TZ are MKS, NPHP and Cep290 complexes listed in Table 1.1, which consist of soluble and membrane associated proteins (Gonçalves and Pelletier, 2017).
Studies in human cells have shown that Cep290 is the most proximal characterised protein and is localised near the basal body. The MKS complex components MKS1, TCTN2 and TMEM67 are distal to this and occupy a wider area. The transmembrane proteins TCNT2 and TMEM67 are the most peripheral (Yang et al., 2015). Mouse knockout studies have shown that depletion of key components of the MKS, specifically the Cep290 and NPHP modules, affects ciliogenesis in a tissue-specific manner (Garcia-Gonzalo et al., 2011, Cui et al., 2011, Rachel et al., 2015, Ronquillo et al., 2016). Work in C. elegans has shown that Cep290 is required for the formation of Y-linkers, and Cep290 has also been localised to the linkers by electron microscopy in Chlamydomonas, suggesting it may be a structural component of these linkers (Rachel et al., 2015, Craige et al., 2010). It has been shown that mutating two proteins within the same TZ complex (Table 1.1), has no synergistic effect (Garcia-Gonzalo et al., 2011), although mutating members of different complexes greatly exacerbates observed ciliogenesis defects. These data suggest that there is a degree of redundancy within complexes, and that members of different complexes must co-operate to enable assembly of the TZ and the cilium (Yee et al., 2015).

**Table 1.1: Members of MKS, NPHP and CEP290 complexes**

<table>
<thead>
<tr>
<th>Complex</th>
<th>Protein</th>
<th>Sub-cellular localisation</th>
<th>Associated disease(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MKS</td>
<td>MKS1</td>
<td>Centrosome, TZ</td>
<td>MKS, JBTS, BBS</td>
</tr>
<tr>
<td></td>
<td>B9D1</td>
<td>BB, TZ, axoneme</td>
<td>MKS, JBTS, BBS</td>
</tr>
<tr>
<td></td>
<td>B9D2</td>
<td>BB, TZ, nucleus</td>
<td>MKS</td>
</tr>
<tr>
<td></td>
<td>TCTN1</td>
<td>BB, TZ</td>
<td>JBTS</td>
</tr>
<tr>
<td></td>
<td>TCTN2</td>
<td>BB, TZ, axoneme</td>
<td>MKS, JBTS</td>
</tr>
<tr>
<td></td>
<td>TCTN3</td>
<td>BB, TZ, axoneme</td>
<td>JBTS, OFD4</td>
</tr>
<tr>
<td></td>
<td>CC2D2A</td>
<td>Centrosome, TZ</td>
<td>MKS, JBTS, COACH</td>
</tr>
<tr>
<td></td>
<td>TMEM17</td>
<td>TZ</td>
<td>OFD6 (Li et al., 2016)</td>
</tr>
<tr>
<td></td>
<td>TMEM67</td>
<td>BB, TZ, axoneme</td>
<td>MKS, JBTS, COACH, NPHP, BBS</td>
</tr>
<tr>
<td></td>
<td>TMEM107</td>
<td>TZ</td>
<td>MKS, JBTS, OFD</td>
</tr>
<tr>
<td></td>
<td>TMEM216</td>
<td>BB, TZ, axoneme, Golgi</td>
<td>MKS, JBTS</td>
</tr>
<tr>
<td></td>
<td>TMEM231</td>
<td>TZ</td>
<td>MKS, JBTS</td>
</tr>
<tr>
<td></td>
<td>TMEM237</td>
<td>TZ</td>
<td>JBTS</td>
</tr>
<tr>
<td>NPHP</td>
<td>NPHP1</td>
<td>TZ, cell junctions</td>
<td>JBTS, NPHP, SLS</td>
</tr>
<tr>
<td></td>
<td>NPHP4</td>
<td>Centrosome, TZ, cilium, cell junctions</td>
<td>NPHP, SLS</td>
</tr>
<tr>
<td>CEP290</td>
<td>CEP290</td>
<td>Centriolar satellites, centrosome, TZ</td>
<td>MKS, JBTS, LCA, BBS, SLS</td>
</tr>
<tr>
<td></td>
<td>NPHP5</td>
<td>Centrosome, cilium, TZ</td>
<td>SLS</td>
</tr>
</tbody>
</table>

TZ; transition zone, BB; basal body, MKS; Meckel syndrome, JBTS; Joubert syndrome, BBS; Bardet-Biedl syndrome, NPHP; Nephronophthisis, OFD4; Orofaciodigital syndrome IV, OFD6; Orofaciodigital syndrome VI, COACH, COACH syndrome, SLS; Senior-Loken syndrome. Adapted from Gonçalves and Pelletier, 2017
1.5.2 Functions of the primary cilia

The primary cilium was first observed in the 1960s but was thought to be a vestigial organelle, retained even though it had lost its function in motility. However, it is now known that the primary cilium acts as the cell’s antenna, sensing external chemical and mechanical cues. A wide variety of signalling pathways have been linked to the primary cilium including Hh, WNT, Notch, Hippo, PDGF, mTOR and TGF-β (Wheway et al., 2018, Pala et al., 2017). The best characterised is the role of the primary cilium in Hh signaling. The sonic hedgehog (SHh) ligand binds to the transmembrane receptor PTCH1, which is localised to the ciliary membrane. In its inactive state, PTCH1 remains bound at the ciliary membrane and represses and excludes the smoothened receptor (SMO) from the cilium. In addition to this, the zinc-finger containing Gli transcription factors are sequestered and repressed by Suppressor of Fused (SUFU) at the tip of the primary cilium. Upon binding of a SHh ligand to the PTCH1 receptor, it becomes activated and no longer suppresses SMO. SMO translocates to the primary cilium, while PTCH1 is now excluded from the ciliary membrane. SMO then represses SUFU, allowing Gli transcription factors to become post-translationally modified, generating an activator form (GliA). This modified Gli is then transported out of the cilium via IFT motor proteins. It then enters the nucleus to activate the expression of downstream targets.

Primary cilia were first shown to play a role in the Hh signalling pathway through the study of IFT mutant mice, which displayed SHh phenotypes. Disruption to the Hh pathway leads to a characteristic set of phenotypes including midline defects (Chiang et al., 1996), neural tube defects (Echelard et al., 1993), polydactyly (Hui and Joyner, 1993), lung hypoplasia (Warburton et al., 2000), and coloboma (Schimmenti et al., 2003). These phenotypes are seen in many ciliopathies; Meckel-Gruber syndrome is one such ciliopathy and patients diagnosed with the condition often present with all of these symptoms (Hartill et al., 2017). The Hh signalling pathway has also been shown to have a role in cancer; the presence of primary cilia on granule neuron precursors was shown to have both an activating and inactivating role in tumorigenesis depending on the underline oncogenic event (Han et al., 2009, Wong et al., 2009). In addition to this, loss of primary cilia on tumour cells has been observed in a number of cancers (Kim et al., 2011, Moser et al., 2009).
Figure 1.5: Hedgehog signalling at the primary cilium

A. In the absence of Hedgehog (Hh) ligand, Patched (PTCH) inhibits Smoothened (SMO) localisation to the primary cilium. Nuclear translocation of Glioma (Gli) protein is also inhibited by Suppressor of fused (SUFU) to repress transcription of Hh target genes. B. Binding of Hh to PTCH releases the inhibition of SMO. SMO translocates to the primary cilium and relieves SUFU inhibitory signal on Gli. Gli proteins then translocate to the nucleus to activate Hh target genes.

The best characterised mechanosensory function of primary cilia is in renal epithelial cells and vascular endothelial cells, where primary cilia sense the fluid flow of urine and blood flow and pressure respectively. The primary cilia bend in response to the sheer force leading to an influx of extracellular calcium (Praetorius and Spring, 2001). This process is dependent on polycystin-1 (PC-1) and PC-2 proteins, which complex with one another, forming a mechanosensory complex. The bending of the primary cilium causes opening of the PC-2 calcium channel and there is an influx of external Ca$^{2+}$, leading to an increase in intracellular calcium levels (Nauli et al., 2003). Calcium acts as an important second messenger. Increases in its concentration in vascular endothelial cells lead to the release of nitric oxide gas, which diffuses to surrounding smooth muscle cells causing vasodilation (Nauli et al., 2008). Mutations that affect the function of PC-1 or PC-2 can lead to the development of a ciliopathy known as Autosomal Dominant Polycystic Kidney Disease (APKD) (The International Polycystic Kidney Disease Consortium, 1995, Mochizuki et al., 1996). The characteristic feature is numerous fluid filled renal cysts that eventually cause kidney failure, although patients also suffer from
hypertension. Importantly, in the absence of flow, PC-1 is cleaved and becomes a transcriptional activator. Expression of a non-cleavable form of PC-1 also leads to renal cyst formation (Yu et al., 2007).

Results from proteomic and genetic screens of ciliated cells suggest that between 1,200 and 1,800 genes are required for normal primary cilium structure and function in mammals (Gherman et al., 2006; Wheway et al., 2015). The importance of the primary cilium function is highlighted by the disorders that are caused by ciliary dysfunction collectively known as ciliopathies. I have previously mentioned polycystic kidney disease and Meckel-Gruber syndrome. Others include Joubert syndrome, Senior-Løken syndrome, Bardet-Biedl syndrome and Alström syndrome. The conditions differ greatly in severity and often mutations in many different genes can result in the same syndrome. In addition to this, some genes have been associated with a number of different syndromes (Helou et al., 2007, Delous et al., 2007). Symptoms include polydactyly, deafness, neural tube and midline defects, situs defects, kidney fibrosis, obesity, neurodevelopmental defects, cystic kidney disease and retinal degeneration. The severity of the conditions is thought to be related to the genetic locus that is mutated, the effect of the mutation on the produced protein and the presence of additional mutations to modifier genes (Hildebrandt et al., 2011).

1.5.3 Primary cilia assembly and disassembly

There are two distinct ciliogenesis pathways, described as intracellular and extracellular. The defining feature is whether growth of the primary cilium initiates intracellularly and is then docked to the membrane or whether the primary cilium is formed after mother centriole/basal body docking to the plasma membrane. The extracellular pathway has been observed in epithelial cell of lungs and kidneys, while the intracellular pathway occurs in retinal epithelial cells and fibroblasts (Sorokin et al., 1968, 1962). The intracellular pathway is better characterised and will be focused on here (Figure 1.6). In order for ciliogenesis to occur, the mother centriole must be mature and have acquired DAs and SDAs, as discussed previously. Upon serum starvation, cytoplasmic vesicles accumulate around the DAs. These vesicles are derived from the Golgi and recycling endosomes. Their trafficking and docking relies on an intact MT and actin network. The
kinesin motor protein KIFC1 is recruited to the Golgi upon serum starvation and its depletion leads to an accumulation of ciliary proteins at the Golgi due to impaired transport (Lee et al., 2017b). Depletion of the motor protein myosin Va disrupts the centrosomal actin network and prevents vesicle docking to the DAs (Wu et al., 2018a). After docking, the membrane shaping protein EHD1 is recruited to these vesicles and initiates their fusion into a large ciliary vesicle (CV) (Lu et al., 2015b). The CV caps the entire distal end of the basal body. Recruitment of the small GTPase Rab8a promotes docking of the vesicle through its interaction with DA proteins CEP164 and chibby (Burke et al., 2014, Schmidt et al., 2012), but is also crucial for the extension of the CV (Lu et al., 2015b). Rab8a recruitment has been shown to be dependent on CEP290 (Kim et al., 2008a).
Coincident with ciliary membrane assembly is the migration of the centrosome to the cell surface. This migration is mainly dependent on mechanical forces generated from MT rearrangement and also partly by actin. There is a dramatic increase in MT nucleation and stabilisation at the centrosome during centrosome migration. These MTs bundle together and orient towards the apical surface, pushing the centrosome towards the surface (Pitaval et al., 2017). In *C. elegans*, the centriole has been shown to move along MTs in a Dynein-1 dependent manner (Li et al., 2017). The basal body then docks to the membrane via DAs (Sánchez and Dynlacht, 2016).

The CP110-Cep97 cap negatively regulates ciliogenesis, and needs to be removed from the mother centriole during ciliogenesis (Spektor et al., 2007). Recruitment of TTBK2 by Cep164 to DAs is thought to play a role in this cap removal (Čajánek and Nigg, 2014). In addition to this the suppression of the kinase Aurora A is crucial to ciliogenesis, as the kinase promotes the disassembly of primary cilia (Inaba et al., 2016, Inoko et al., 2012). Once ciliation is initiated, the microtubules within the axoneme and the ciliary membrane elongate. This process relies heavily on tubulin and IFT. Large quantities of tubulin enter the cilium through diffusion and also IFT (Harris et al., 2018). IFT molecules concentrate around the DAs (Yang et al., 2018). IFT complexes are formed with the co-operation of the BBsome, a stable protein complex that plays a role in trafficking membrane proteins to the cilium (Wei et al., 2012). Perturbation of IFT components or cytoplasmic tubulin levels results in defects in axoneme elongation and cilium length (Sharma et al., 2011, Besschetnova et al., 2010, Marshall et al., 2005). Axoneme elongation is coupled with extension of the ciliary membrane. Studies have shown that over-expression of Rab8a or depletion of IFT genes can lead to the extension of the ciliary membrane beyond the length of the axoneme (Absalon et al., 2008, Nachury et al., 2007). Abnormal extension of the ciliary membrane has also been shown to promote further axoneme growth, resulting in longer cilia (Lu et al., 2015a, Nachury et al., 2007). These studies suggest the existence of a mechanism by which the axoneme length influences ciliary membrane extension, possibly through regulation of Rab8a.

Disassembly of the primary cilia requires the destabilisation and depolymerisation of axonemal MTs. Aurora A has been shown to play a key role in destabilising axonemal
MTs (Pugacheva et al., 2007). Upon receiving cell-cycle re-entry cues, Aurora A is activated and the kinase phosphorylates and activates HDAC6. HDAC6 is a histone deacetylase that deacetylates and destabilises axoneme MTs (Ran et al., 2015, Plotnikova et al., 2012). A multitude of signalling pathways converge on Aurora A to influence cilium disassembly (Wang and Dynlacht, 2018b). Kif2A and Kif24 are two kinesins that have the ability to depolymerize MTs and have been linked to primary cilium disassembly. Kif2A activation in quiescent cells promotes MT depolymerisation and cilium disassembly (Miyamoto et al., 2015), while Kif24 interacts with CP110 and its depletion promotes aberrant cilium assembly (Kobayashi et al., 2011). Cilium disassembly also requires ciliary membrane remodelling. During cilium reabsorption the ciliary pocket has been shown to undergo active remodelling and also enhanced endocytosis is observed during this period. Perturbation to endocytosis at the ciliary pocket has been shown to block cilium reabsorption, highlighting its key role in cilium disassembly (Saito et al., 2017). The release of ciliary vesicles also contribute to cilium disassembly. These vesicles are released by ectocytosis, and are generated by cilium decapitation via intra-ciliary actin polymerization (Nager et al., 2017, Phua et al., 2017). Ciliary vesicles have been shown to contain IFT-B, removal of which limits cilium regrowth (Phua et al., 2017). Perturbation of ciliary decapitation can cause defective ciliary disassembly, highlighting its crucial role in this process (Nager et al., 2017, Phua et al., 2017).

1.6 The role of the centrosome/cilia in schizophrenia and cognition

The molecular mechanisms by which associated genes contribute to SZ risk and cognitive function are not well understood. It is pertinent to ask if an overlapping phenotypic characteristic of SZ and cognitive deficits could be due to genetic variability in a shared pathobiology, like centrosomal and ciliary dysfunction. A disproportionate number of brain disorders are associated with centrosomal genes, suggesting that the organelle underlies normal brain and cognitive development. Strikingly, mutations in 77 of 87 genes that encode ciliary and centrosomal proteins associated with human ciliopathies result in neurodevelopmental or cognitive deficits (Guo et al., 2015). Variable neuropsychiatric phenotypes have been reported in patients with ciliopathies (Forsythe and Beales, 2013). A study of 41 candidate risk genes for SZ and other neuropsychiatric
disorders found that for 20 of these, knockdown of gene expression resulted in a ciliary phenotype (Marley et al., 2012). Recently, a defect in primary cilia formation has been reported in cells derived from patients with SZ and bipolar disorder (Muñoz-Estrada et al., 2018). In addition to this, centrosome and ciliary directed processes have been implicated in SZ pathogenesis. As described previously, the primary cilium plays an important role in Hh signalling, a pathway-wide association study found Hh signalling to be one of the five top pathways associated with SZ (Liu et al., 2017). The Hh pathway plays an important role in brain development and the most characterized ligand, SHh has been shown to regulate axonal guidance (Hong et al., 2010). The primary cilium also plays an important role in WNT signalling. Dysregulated WNT gene expression and plasma protein levels have been reported in patients with SZ (Hoseth et al., 2018). Gene expression alterations in the WNT signalling pathway have also been observed in human induced pluripotent stem cells (hiPSC) and forebrain patterned neural progenitor cells derived from patients with SZ (Topol et al., 2015, Brennand et al., 2011). There is evidence for altered dopamine signalling in SZ (reviewed by Howes et al., 2015). Dopamine receptors have been shown to localise to the primary cilia and this localisation has been shown to depend on components of IFT and the BBS complex (Leaf and Zastrow, 2015, Domire et al., 2011, Marley and Zastrow, 2010). Furthermore, dopamine signalling has been shown to influence the length of primary cilia in striatal neurons (Miyoshi et al., 2014).

Disrupted in schizophrenia 1 (Disc1) was identified as a SZ risk gene when one-third of a Scottish family that harboured a balanced translocation that disrupted DISC1 suffered from major mental illness (Millar et al., 2000, St Clair et al., 1990). DISC1 has been shown to localise to the centrosome, nucleus, mitochondria and the basal body of ciliated cells. It has been co-localised with γ-tubulin and has been shown to anchor dynein motor proteins at the centrosome (Kamiya et al., 2005). Ablation of the protein results in defective ciliogenesis and a decrease of dopamine receptors on the cell surface (Marley and Zastrow, 2010). DISC1 is involved in neurite outgrowth and development of the cerebral cortex (Kamiya et al., 2005). DISC1 has been shown to complex with PCM1 and BBS4. Variants in PCM1 were also initially thought to contribute to SZ risk (Datta et al., 2008, Gurling et al., 2006), although these associations were not replicated in Japanese
populations (Hashimoto et al., 2011), or in more recent, much larger datasets of European ancestry (Pardiñas et al., 2018).

1.7 Project aims

This thesis investigates the contribution of GWAS-identified variants in centrosomal genes to SZ and cognitive traits. I first hypothesised that GWAS-identified risk genes for SZ, which have centrosomal functions, are associated with cognitive function. I aimed to identify genes at SZ risk loci that encode proteins with centrosomal localisation and if variants in these candidate genes were associated with cognition. Risk variants in candidate risk genes with centrosomal localisation were analysed using a dataset of Irish psychosis cases and controls (n=1,235), to test for effects on IQ, working memory, episodic memory and attention.

Following this analysis, I focused on a specific centrosomal protein, SDCCAG8, based on its association with social cognition in the Irish dataset, and its association with ciliopathies where patients present with intellectual disability. I aimed to identify novel roles for SDCCAG8 at the centrosome and in primary cilia which may contribute to SZ pathophysiology and cognition. I used reverse genetic analyses of SDCCAG8 to explore potential mechanisms by which genes that function at the centrosome could contribute to SZ and cognitive function.

Given that ciliation capacity is reduced in cells that lack SDCCAG8 and that the residual cilia display defective cilium-dependent signaling, I hypothesised that gene expression would be dysregulated in SDCCAG8-deficient cells. I aimed to identify the genes and pathways that are dysregulated as a consequence of SDCCAG8 deficiency and investigate whether the pathways affected by SDCCAG8 deficiency are enriched for genes associated with SZ and cognition.
2 Materials and Methods

2.1 Materials

2.1.1 Chemicals and solutions

In the course of this study, unless otherwise stated, all of the analytical grade chemicals, including organic solvents, acids and alcohols, were obtained from Sigma Aldrich (Arklow, Ireland), Fisher Scientific (Loughborough, UK), VWR (Dublin, Ireland), Invivogen (Toulouse, France), Lennox Supplies Ltd (Dublin, Ireland), Melford Laboratories Ltd. (Suffolk, UK) and Merck (Darmstadt, Germany). Western blotting detection system was obtained from GE Healthcare Life Sciences (Buckinghamshire, UK). Acrylamide (37:5:1) protein gel solutions were purchased from Severn Biotech Ltd. (Worcestershire, UK). All buffers and hydrophilic reagents were prepared by dissolution in double distilled water (ddH₂O) or Milli-Q purified water collected from the ELGA Purelab flex 3 (ELGA LabWater, Veolia Water Systems, Kildare, Ireland) and were autoclaved or filtered, where appropriate. All chemical reagents and buffers used in this study are listed in Table 2.1.

<table>
<thead>
<tr>
<th>Solution/Reagent</th>
<th>Composition</th>
<th>Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocking solution for immunofluorescence (IF) microscopy</td>
<td>1% BSA in 1x Phosphate buffered saline (1x PBS)</td>
<td>Blocking cells and diluting antibodies during immunofluorescence staining</td>
</tr>
<tr>
<td>Blocking solution for western blot (WB)</td>
<td>5% dried semi-skimmed milk in 1x PBS</td>
<td>Minimising non-specific binding of antibodies to nitrocellulose membrane during immunoblot</td>
</tr>
<tr>
<td>Crystal violet staining buffer</td>
<td>0.5% crystal violet, 6% glutaraldehyde in 1x PBS</td>
<td>Staining and visualising colonies formed from clonogenic survival assay</td>
</tr>
<tr>
<td>Cytoskeleton buffer (CB)</td>
<td>137mM NaCl, 5mM KCl, 1.1mM Na₂HPO₄, 0.4mM KH₂PO₄, 2mM MgCl₂, 2mM</td>
<td>Fixing and staining IF slides to visualise</td>
</tr>
<tr>
<td><strong>DNA loading buffer (6x)</strong></td>
<td>EGTA, 5mM PIPES, 5.5mM glucose, pH 6.1</td>
<td>microtubules, filter sterilised</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>----------------------------------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td><strong>IF fixation buffer</strong></td>
<td>4% Paraformaldehyde (PFA) in 1x PBS or methanol containing 5mM EGTA, pH 8.0 at -20°C</td>
<td>Fixing cells prior to IF microscopy staining</td>
</tr>
<tr>
<td><strong>Luria-Bertani (LB) medium</strong></td>
<td>1% Tryptone, 0.5% Yeast Extract, 1% NaCl, pH 7.0</td>
<td>Growth of bacterial (E. coli) cultures</td>
</tr>
<tr>
<td><strong>Mounting medium</strong></td>
<td>0.3% N-propyl gallate, 80% Glycerol</td>
<td>Mounting slides after IF staining</td>
</tr>
<tr>
<td><strong>1x Phosphate buffered saline (PBS)</strong></td>
<td>137mM NaCl, 2.7mM KCl, 1.4mM NaH₂PO₄, 4.3mM Na₂HPO₄, pH 7.4</td>
<td>Washes, blocking and antibody dilution during IF</td>
</tr>
<tr>
<td><strong>1x PBST</strong></td>
<td>1x PBS, 0.25% Tween-20 and 1% BSA</td>
<td>Washing cells, preparing blocking and antibody solution during FACS</td>
</tr>
<tr>
<td><strong>Permeabilisation buffer</strong></td>
<td>0.15% Triton X-100 in 1x PBS or 1x CB</td>
<td>Permeabilise cells after fixation with PFA</td>
</tr>
<tr>
<td><strong>Ponceau S</strong></td>
<td>0.5% w/v Ponceau S, 5% acetic acid</td>
<td>Staining nitrocellulose membrane during immunoblotting</td>
</tr>
<tr>
<td><strong>RIPA Buffer</strong></td>
<td>20mM Tris-HCL (pH 7.5), 150mM NaCl, 1mM EDTA, 1mM EGTA, 1% Triton-X, 0.1% SDS, 0.1% Sodium deoxycholate, 1X cOmplete™ Protease Inhibitor Cocktail (Roche)</td>
<td>Cellular protein extraction for immunoblotting</td>
</tr>
<tr>
<td><strong>SDS-PAGE running buffer</strong></td>
<td>24mM Tris, 192mM glycine, 0.1% SDS</td>
<td>Running SDS-PAGE gels</td>
</tr>
<tr>
<td><strong>6x Sample buffer</strong></td>
<td>440mM Tris, 12% SDS, 30% Glycerol, 0.1% Bromophenol blue, 30% β-Mercaptoethanol</td>
<td>Binding protein for SDS-PAGE gels</td>
</tr>
<tr>
<td><strong>Semi-Dry transfer buffer</strong></td>
<td>1x Tris-Glycine, 20%; Methanol, 0.0375% SDS</td>
<td>Semi-dry transfer of proteins to nitrocellulose</td>
</tr>
<tr>
<td>1 x TAE</td>
<td>40mM Tris, 20mM Acetic acid, 1mM EDTA</td>
<td>Agarose DNA gels</td>
</tr>
<tr>
<td>1x Tail buffer</td>
<td>50mM Tris-HCl (pH 8.8), 100mM EDTA (pH 8.0), 100mM NaCl, 1% SDS</td>
<td>Lysis of mammalian cells to extract genomic DNA</td>
</tr>
<tr>
<td>1x TBST</td>
<td>137mM NaCl, 2.7mM KCl, 19mM Tris with 0.025% Tween20</td>
<td>Washing and blocking during Western blot</td>
</tr>
</tbody>
</table>

### 2.1.2 Molecular Biology Reagents

Biological reagents used for DNA restriction digestion and subcloning, such as restriction enzymes, DNA ligase, DNA polymerase (Klenow Fragment I), T4 Polynucleotide Kinase and calf alkaline phosphatase were obtained from New England Biolabs (NEB, Hitchin, UK). Shrimp Alkaline Phosphatase was obtained from United States Biochemical (USB, Santa Clara, CA, USA). DNA size markers used were Quick-load 100bp DNA ladder (NEB) or 1Kb Plus DNA ladder (Thermofisher Scientific, Loughborough, UK). KOD polymerase was obtained from Novagen (Merck Millipore). A list of cloning and expression vectors used is presented in Table 2.2.

The Colour protein standard, broad range marker (NEB) was used as a protein size marker.
A list of commercially available molecular biology kits used in this study is presented in Table 2.3

**Table 2.2: Plasmids used in the study**

<table>
<thead>
<tr>
<th>Plasmid name</th>
<th>Use Description</th>
<th>Source/References</th>
</tr>
</thead>
<tbody>
<tr>
<td>pGEM-T-Easy</td>
<td>General subcloning</td>
<td>Promega (Madison, WI, USA)</td>
</tr>
<tr>
<td>pEGFP-C1/N1</td>
<td>Expression of various genes in mammalian cells</td>
<td>Clontech (Mountain View, CA, USA)</td>
</tr>
<tr>
<td>pX330</td>
<td>Cas9 nuclease and guide RNA for expression of CRISPR system</td>
<td>Cong et al. 2013 (Obtained from Addgene, Cambridge, MA, USA)</td>
</tr>
<tr>
<td>pLoxNeo</td>
<td>Co-transfected with CRISPR system as a selection marker</td>
<td>Arakawa et al. 2001</td>
</tr>
<tr>
<td>pcDNA3.1(+)</td>
<td>Expression of $SDCCAG8$ in mammalian cells</td>
<td>Thermofisher Scientific</td>
</tr>
</tbody>
</table>

**Table 2.3: Kits used in this study.**

<table>
<thead>
<tr>
<th>Name</th>
<th>Use Description</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>GenElute plasmid miniprep kit</td>
<td>Small scale plasmid DNA extraction</td>
<td>Sigma-Aldrich</td>
</tr>
<tr>
<td>High Capacity RNA-to-cDNA kit</td>
<td>cDNA synthesis</td>
<td>Applied Biotecnologies (Life Technologies)</td>
</tr>
<tr>
<td>NucleoBond Xtra midi (EF)</td>
<td>Large scale (endotoxin-free) plasmid DNA extraction</td>
<td>Macherey-Nagel (Düren, Germany)</td>
</tr>
<tr>
<td>QIAquick gel extraction kit</td>
<td>Extraction and purification of DNA fragments from agarose gel</td>
<td>Qiagen (Crawley, UK)</td>
</tr>
<tr>
<td>QIAquick purification kit</td>
<td>Purification of DNA fragments</td>
<td>Qiagen</td>
</tr>
<tr>
<td>RNeasy Mini Kit</td>
<td>Extraction and purification of RNA from cells</td>
<td>Qiagen</td>
</tr>
<tr>
<td>TruSeq® Stranded mRNA Library Prep</td>
<td>Preparation of library preps for RNA-Seq</td>
<td>Illumina</td>
</tr>
</tbody>
</table>
2.1.3 Antibodies

Primary and secondary antibodies were used for detecting protein of interest during western blotting (WB) and immunofluorescence microscopy (IF). Freeze-dried conjugated affinity purified antibodies were re-suspended as recommended by suppliers in 1x PBS or Milli-Q H2O. Detailed lists of primary and secondary antibodies used throughout this study are shown in Table 2.4 and 2.5.

Table 2.4: List of primary antibodies used for this study

<table>
<thead>
<tr>
<th>Antigen</th>
<th>Clone/ Catalog number</th>
<th>Host species</th>
<th>Dilution (WB)</th>
<th>Dilution (IF)</th>
<th>Manufacturer</th>
</tr>
</thead>
<tbody>
<tr>
<td>α-Tubulin</td>
<td>B512</td>
<td>Mouse monoclonal</td>
<td>1:10000</td>
<td>1:2000</td>
<td>Sigma-Aldrich</td>
</tr>
<tr>
<td>Acetylated tubulin</td>
<td>T6793</td>
<td>Mouse monoclonal</td>
<td>1:2000</td>
<td></td>
<td>Sigma-Aldrich</td>
</tr>
<tr>
<td>Arl13b</td>
<td>17711-1-AP</td>
<td>Rabbit polyclonal</td>
<td>1:2000</td>
<td></td>
<td>Proteintech (Manchester, UK)</td>
</tr>
<tr>
<td>Centrin</td>
<td>20H5</td>
<td>Mouse monoclonal</td>
<td>1:1000</td>
<td></td>
<td>EMD Millipore (Darmstadt, Germany)</td>
</tr>
<tr>
<td>Centrobin</td>
<td>6D4 F4</td>
<td>Mouse monoclonal</td>
<td>1:300</td>
<td></td>
<td>(Ogungbenro et al., 2018)</td>
</tr>
<tr>
<td>Cep97</td>
<td>22050-1-AP</td>
<td>Rabbit monoclonal</td>
<td>1:2000</td>
<td></td>
<td>Proteintech</td>
</tr>
<tr>
<td>Cep135</td>
<td>1457 748</td>
<td>Rabbit polyclonal</td>
<td>1:2000</td>
<td></td>
<td>Dr. Alex Bird (Bird and Hyman 2008)</td>
</tr>
<tr>
<td>Antigen</td>
<td>Label</td>
<td>Host species</td>
<td>Dilution (IB)</td>
<td>Dilution (IF)</td>
<td>Manufacturer</td>
</tr>
<tr>
<td>----------</td>
<td>-------------</td>
<td>--------------</td>
<td>---------------</td>
<td>---------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>Cep164</td>
<td>IF3 G10</td>
<td>Mouse monoclonal</td>
<td>1:100000</td>
<td>(Daly et al. 2016)</td>
<td></td>
</tr>
<tr>
<td>Cep290</td>
<td>IC3 G10</td>
<td>Mouse monoclonal</td>
<td>1:1000</td>
<td>Dr. David Gaboriau</td>
<td></td>
</tr>
<tr>
<td>C-Nap1</td>
<td>6F2 C8</td>
<td>Mouse monoclonal</td>
<td>1:5</td>
<td>(Flanagan et al. 2017)</td>
<td></td>
</tr>
<tr>
<td>CP110</td>
<td>12780-1-AP</td>
<td>Rabbit polyclonal</td>
<td>1:2000</td>
<td>Proteintech</td>
<td></td>
</tr>
<tr>
<td>Gamma-tubulin</td>
<td>SC-7396</td>
<td>Goat polyclonal</td>
<td>1:1000</td>
<td>Santa-Cruz (Dallas, TX, USA)</td>
<td></td>
</tr>
<tr>
<td>Pericentrin</td>
<td>Ab28144</td>
<td>Mouse monoclonal</td>
<td>1:500</td>
<td>Abcam</td>
<td></td>
</tr>
<tr>
<td>SMO</td>
<td>Ab113438</td>
<td>Rabbit polyclonal</td>
<td>1:300</td>
<td>Abcam</td>
<td></td>
</tr>
<tr>
<td>SDCCAG8</td>
<td>13471-1-AP</td>
<td>Rabbit polyclonal</td>
<td>1:1000</td>
<td>Proteintech</td>
<td></td>
</tr>
<tr>
<td>SDCCAG8</td>
<td>66284-1-lg</td>
<td>Mouse monoclonal</td>
<td>1:2000</td>
<td>Proteintech</td>
<td></td>
</tr>
<tr>
<td>TTBK2</td>
<td>15072-1-AP</td>
<td>Rabbit polyclonal</td>
<td>1:500</td>
<td>Proteintech</td>
<td></td>
</tr>
</tbody>
</table>

Table 2.5: List of secondary antibodies used for this study
<table>
<thead>
<tr>
<th>Antibody Type</th>
<th>Fluorophore</th>
<th>Species</th>
<th>Dilution</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mouse IgG</td>
<td>Alexa 488</td>
<td>Donkey</td>
<td>1:1000</td>
<td>Jackson ImmunoResearch (Suffolk, UK)</td>
</tr>
<tr>
<td>Mouse IgG</td>
<td>Alexa 594</td>
<td>Donkey</td>
<td>1:1000</td>
<td>Jackson ImmunoResearch</td>
</tr>
<tr>
<td>Mouse IgG</td>
<td>HRP</td>
<td>Goat</td>
<td>1:10000</td>
<td>Jackson ImmunoResearch</td>
</tr>
<tr>
<td>Rabbit IgG</td>
<td>HRP</td>
<td>Goat</td>
<td>1:10000</td>
<td>Jackson ImmunoResearch</td>
</tr>
<tr>
<td>Rabbit IgG</td>
<td>Alexa 488</td>
<td>Donkey</td>
<td>1:1000</td>
<td>Jackson ImmunoResearch</td>
</tr>
<tr>
<td>Rabbit IgG</td>
<td>Alexa 594</td>
<td>Donkey</td>
<td>1:1000</td>
<td>Jackson ImmunoResearch</td>
</tr>
<tr>
<td>Goat IgG</td>
<td>Cy5</td>
<td>Donkey</td>
<td>1:1000</td>
<td>Bethyl Laboratory Inc. (Texas, USA)</td>
</tr>
</tbody>
</table>

2.2 Biological materials and tissue culture reagents

2.2.1 Bacterial strains

Top 10 *Escherichia coli* (F− mcrA Δ (mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 nupG recA1 araD139 Δ (ara-leu) 7697 galE15 galK16 rpsL (StrR) endA1 λ−) were obtained from Invitrogen and used for all cloning throughout this study. Ampicillin or kanamycin antibiotics, obtained from Sigma-Aldrich, were used for clone selection at final concentrations of 50µg/ml or 30µl, respectively.
2.2.2 Tissue culture reagents

All sterile plastic ware used for tissue culture was obtained from Sarstedt (Numbrecht, Germany), Corning (Corning, NY, USA), Fisher and Sigma-Aldrich. Tissue culture reagents such as trypsin, foetal bovine serum (FBS), penicillin and streptomycin (pen/strep), dimethyl sulfoxide (DMSO) and the culture media used in this study, Dulbecco’s modified eagle medium with Ham’s F12 nutrient mix (DMEM-F12), were purchased from Sigma-Aldrich. Opti-MEM serum-free medium, lipofectamine 2000 and oligofectamine were purchased from Invitrogen, Life Technologies.

2.2.3 Cell lines and culture conditions

This study was carried out both in non-transformed hTERT-RPE1 which shows a relatively stable diploid karyotype, and SH-SY5Y, a neuroblastoma cell, originally derived from a metastatic bone tumour biopsy. Further information is displayed in Table 2.6.

**Table 2.6: Cell lines used in this study**

<table>
<thead>
<tr>
<th>Cell line</th>
<th>Brief description</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>hTERT-RPE1</td>
<td>Human retina epithelial cell line, non-transformed,</td>
<td>CRL-4000, ATCC (Middlesex, UK)</td>
</tr>
<tr>
<td></td>
<td>telomerase immortalised</td>
<td></td>
</tr>
<tr>
<td>SH-SY5Y</td>
<td>Human neuronal-like cell line, transformed.</td>
<td>CRL-2266, ATCC</td>
</tr>
</tbody>
</table>

All cells were cultured in a 95% air, 5% CO2 incubator at 37°C and their culture media were 1% pen-strep. hTERT-RPE1 media was supplemented with 10% FBS while SH-SY5Y media was supplemented with 15% FBS.
2.3 Identifying SZ risk genes with centrosomal function

A list of genes with centrosomal functions was compiled based on results from the MiCroKiTS 4.0 database (http://microkit.biocuckoo.org/index.php), which has collected all proteins identified to be localised in several sub-cellar regions, including the centrosome (Huang et al., 2015). To this list I added results from Centrosome:db, which contains a set of human genes encoding proteins that are localised to the centrosome, either as centrosome constituents or as centrosome visitors (Alves-Cruziero et al., 2014). These genes were assembled into a final list of unique genes with centrosomal localisation (Supplementary Table 3.1), to which I refer as “centrosomal genes” in this Thesis. The centrosomal gene list (N = 806) was cross-referenced with genes located in 108 chromosomal regions associated with SZ risk (N = 350) (Schizophrenia Working Group of the Psychiatric Genomics, 2014) to identify candidate genes for this study.

2.4 Neuropsychological testing of SZ risk SNPs within centrosomal genes

Colleagues in the Cognitive Genetics & Cognitive Therapy (Coggene) Group, now in National University Ireland, Galway and previously in Trinity College Dublin, collected and prepared all genotype and phenotype information described below prior to my arrival to the group. This dataset has been described in previous publications (Whitton et al., 2016), and is described briefly below.

2.4.1 Study participants and neuropsychological tests

This study used an Irish dataset of broad psychosis cases (N = 890) and controls (N = 330) who had completed tests in five main areas of cognition and for whom GWAS data was available. Cases were clinically stable patients with a DSM-IV (Diagnostic and Statistical Manual of Mental Disorders, 4th Edition) diagnosis of SZ (N = 559), schizoaffective disorder (N = 111), bipolar disorder (N = 148), major depressive disorder with psychotic features (n = 38), or psychosis not otherwise specified (N = 34) recruited from five sites across Ireland (Table 2.7). Inclusion criteria required that participants were clinically stable at the time of neuropsychological testing, were aged 18 to 65 years, had no history
of comorbid psychiatric disorder, no substance abuse in the preceding 6 months, no prior head injury with loss of consciousness, and no history of seizures. Diagnosis was confirmed by trained psychiatrists using the Structured Clinical Interview for DSM-IV-TR (First et al., 2002). Healthy control participants were recruited via online and poster advertising. They were aged 18 to 65 years, with no history of substance abuse in the preceding 6 months, no prior head injury with loss of consciousness, no history of seizures, and no personal history of psychosis or in their first-degree relatives. All assessments were conducted in accordance with the relevant ethics committees’ approval from each participating site. All participants had four grandparents born in Ireland and provided written informed consent.

Table 2.7: Participant demographics – neuropsychological test sample

<table>
<thead>
<tr>
<th>Variable</th>
<th>Psychosis broad</th>
<th>Psychosis narrow</th>
<th>Healthy controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>890</td>
<td>670</td>
<td>330</td>
</tr>
<tr>
<td>Schizophrenia (N)</td>
<td>559</td>
<td>559</td>
<td>NA</td>
</tr>
<tr>
<td>Schizoaffective disorder</td>
<td>111</td>
<td>111</td>
<td>NA</td>
</tr>
<tr>
<td>Bipolar disorder</td>
<td>148</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Major depressive disorder</td>
<td>34</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Psychosis not otherwise specified</td>
<td>38</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Male: Female ratio</td>
<td>1.94:1.0</td>
<td>2.4:1.0</td>
<td>1.0:1.2</td>
</tr>
<tr>
<td>Age, mean (S.D.)</td>
<td>42.93 (12.42)</td>
<td>42.28 (12.48)</td>
<td>35.81 (12.66)</td>
</tr>
<tr>
<td>Age at onset, mean (S.D.)</td>
<td>23.78 (8.28)</td>
<td>23.23 (7.68)</td>
<td>NA</td>
</tr>
<tr>
<td>Cognitive full scale IQ, mean (S.D.)</td>
<td>92.04 (19.21)</td>
<td>90.29 (18.08)</td>
<td>119.81 (15.57)</td>
</tr>
</tbody>
</table>

Participants completed a neuropsychological assessment battery designed to target the cognitive deficits observed in SZ. This battery included general cognitive function, episodic memory, working memory and attentional control. General cognitive functioning (IQ) was measured using selected subtests (Vocabulary, Similarities, Block Design, and Matrix Reasoning) from the Wechsler Adult Intelligence Scale (WAIS; Wechsler, 1997a) which assessed verbal, performance and full-scale IQ as well as the Wechsler Test of Adult Reading (WTAR). Episodic memory was assessed using the
logical memory (LM) 1 & 2 tasks as well as the “Faces 1 & 2” task subtests from the Wechsler Memory Scale (WMS) III (Wechsler, 1997b) as well as the Paired-Associates Learning task (PAL). Working memory was assessed using the spatial working memory task (SWM) from the Cambridge Neuropsychological Test Automated Battery (CANTAB) (Robbins et al., 1994) and letter-number sequencing from the WMS III (Wechsler, 1997b). Attentional control was assessed using the Continuous Performance Task (CPT)–Identical Pairs version, and the Sustained Attention to Response Task (Cornblatt et al., 1988, Robertson, 1994). Social cognition was measured by the Reading the Mind in the Eyes task ((Baron-Cohen, 2001), the hinting task (Corcoran, 1995) and the Internal, Personal and Situational Attribution Questionnaire (Kinderman and Bentall, 1997).

2.4.2 Genotyping and statistical analysis

Genotyping was conducted on DNA extracted from blood or saliva from patient and control participants. SNP data was obtained from two different sites; a recent GWAS using the Affymetrix SNP Array 6.0 platform, conducted as part of the Wellcome Trust Case Control Consortium 2 (Irish Schizophrenia Genomics Consortium and the Wellcome Trust Case Control Consortium 2, 2012) and a collaborative GWAS with Cardiff University using an Illumina HumanCoreExome (+custom) SNP array. Full GWAS data on these samples included recently completed imputation using 1000 Genomes Phase I data and IMPUTE2 (Howie et al., 2009) to give ~10 million SNPs genome-wide per sample. Only samples that had passed the quality control filtering were imputed using the 1000 Genomes reference panel. Imputed data were converted to best guess genotypes using PLINK (Purcell et al., 2007). I then used PLINK to extract data on the index SNP at each gene from GWAS data and combine it in SPSS (Statistical Package for the Social Sciences) SPSS 21, IBM) with data from an extensive neuropsychological battery. I carried out statistical analysis in SPSS using a linear based model of regression. Due to differences in cognitive scores between male and female participants and between older and younger individuals, gender and the age at the time of assessment were used as co-variate measures to determine a general genotype effect across all cases and controls. Cognitive test scores for each genotype group were compared using a three group analysis, i.e. homozygous non-risk versus heterozygous versus homozygous risk genotype groups. R² is a measure of how close the data fits to
the regression line. $R^2$ change values were used in this analysis, which identifies changes in the original $R^2$ based on the linear contribution of variables added into the regression model.

2.4.3 Replication

The following work was completed by collaborator Prof James Walters and his group at Cardiff University. 772 individuals with SZ or schizoaffective disorder—depressive subtype were recruited from community mental health teams in Wales and England. The Schedule for Clinical Assessment in Neuropsychiatry (SCAN) interview was conducted followed by case note review and consensus diagnosis according to DSM-IV criteria. Individuals were assessed using the Measurement and Treatment Research to Improve Cognition in Schizophrenia (MATRICS) Consensus Cognitive Battery (MCCB). The MCCB is a 10 test cognitive battery consisting of seven cognitive domains: speed of processing, attention and vigilance, working memory, verbal learning, visual learning, reasoning and problem solving and social cognition (Nuechterlein et al., 2008). These domains were chosen as primary cognitive outcomes given that the MATRICS tests were selected on the basis of the domains they measured, and the fact that this approach minimized measurement error as some domains are assessed by more than one test. All interviewers were trained in the use of the SCAN interview and MCCB. The UK Multicentre Research Ethics Committee (MREC) approved the study and all participants provided valid informed consent. These samples were genotyped at the Broad Institute, Massachusetts on the Illumina HumanOmniExpressExome-8v1. QC procedures followed those of the PGC (Schizophrenia Working Group of the Psychiatric Genomics, 2014). Imputation was carried out by the PGC using IMPUTE2 and for this analysis they converted imputed to best guess genotypes using PLINK (Purcell et al., 2007; Howie et al., 2009).
2.5 Enrichment Analysis

2.5.1 GWAS Data

Initial analysis were bases on summary statistics from the Psychiatric Genomics Consortium (PGC) 2 SZ GWAS (Schizophrenia Working Group of the Psychiatric Genomics, 2014) which was obtained from the PGC website (http://www.med.unc.edu/pgc/results-and-downloads, SCZ2 Download full SNP results). This study included data on 36,989 cases and 113,075 controls. The most recent SZ GWAS (Pardiñas et al., 2018) was also analysed as the data became available from the Walters group data repository on the MRC Centre for Neuropsychiatric Genetics and Genomics website (http://walters.psycm.cf.ac.uk, CLOZUK+PGC2 meta-analysis summary statistic). This study expands on the previous GWAS with a final sample size of 40,675 cases and 64,643 controls. Summary statistics from the most recent EA GWAS (Lee et al., 2018) were obtained from the Social Science Genetic Association Consortium (SSDAG) website (http://ssgac.org/Data.php, Summary data file: GWAS_EA_excl23andMe.txt - Educational attainment (EA) meta-analysis of all discovery cohorts except 23andMe). This study had a sample of approximately 1.1 million individuals. Summary statistics from the most recent IQ GWAS (Savage et al., 2018) were obtained from the Center for Neurogenomics and Cognitive Research website (https://ctg.cnrc.nl/software/summary_statistics, SavageJansen_IntMeta_sumstats.zip). This study included data from 269,867 individuals.

2.5.2 Gene-set Analysis

A gene-set analysis (GSA) is a statistical method for simultaneously analysing multiple genetic markers in order to determine their joint effect. I performed GSA using MAGMA (de Leeuw et al., 2015) (http://ctg.cnrc.nl/software/magma) and summary statistics from various GWAS (Section 2.5.1). An analysis involved three steps. First, in the annotation step, I mapped SNPs with available GWAS results on to genes (GRCh37/hg19 start-stop coordinates +/-20kb). Second, in the gene analysis step, I computed gene P values for
each GWAS dataset. This gene analysis is based on a multiple linear principal
components regression model that accounts for linkage disequilibrium (LD) between
SNPs. The European panel of the 1000 Genomes data was used as a reference panel for
LD. Third, a competitive GSA based on the gene P values, also using a regression
structure, was used to test if the genes in a gene-set were more strongly associated with
either phenotype than other genes in the genome. The MHC region is strongly associated
in the SZ GWAS data. This region contains high LD and the association signal has been
attributed to just a small number of independent variants (Sekar et al., 2016). However,
MAGMA still identifies a very large number of associated genes despite factoring in the
LD information. Of 278 genes that map to chromosome 6 (25-35Mb), 130 genes were
associated with SZ in my MAGMA analysis. To avoid the excessive number of
associated genes biasing the MAGMA GSA, I excluded all genes within the MHC region
from my GSA of SZ. MAGMA was chosen because it corrects for LD, gene size and
gene density (potential confounders) and has significantly more power than other GSA
tools (de Leeuw et al., 2016).

Sets of ‘brain-expressed’ genes (N=14,243) and ‘brain-elevated’ genes (N=1,424) were
sourced from the Human Protein Atlas (https://www.proteinatlas.org/humanproteome/brain) and used as covariates in the GSA. Brain-elevated genes are those that show an elevated expression in brain compared to other tissue types.

2.5.3 Analysis of gene-sets using rare variant data

A list of genes harbouring de novo mutations (DNMs) identified in patients with SZ,
autism spectrum disorder (ASD), intellectual disability (ID) and in unaffected siblings
and controls was sourced from Genovese et al., 2016. The variants were each assigned a
class (protein altering (Prot) which included loss-of-function (LoF) and missense varianrs
and silent). I performed enrichment analysis of these gene lists with my gene-sets using
denovolyzeR (Ware et al., 2015) an open-source software package. Briefly, denovolyzeR
uses a statistical framework developed by Samocha et al., (Samocha et al., 2014) were
the mutability of each gene is determined on a gene by gene basis. The consequence of
every possible DNM is computed, probabilities of DNMs occurring are generated by gene
and further separated by variant class. Using this information, it is possible to calculate the number of expected DNMs for a defined gene-set and compare this to the mutations that were observed in a particular study, using a Poisson framework. The software therefore can compute if there is a significant enrichment of DNMs observed in a genome wide study at a gene set of interest (Ware et al., 2015). Bonferroni multiple test correction was performed for the tests of DNMs Phenotypes (SZ, ID and ASD,) and variant classes (LoF and missense) within each of these phenotypes (N=6 in total).

2.6 Cell biology methods

2.6.1 Cell maintenance and proliferation analysis

All cells were maintained at exponential growth phase; cells were subcultured when they attained confluency of between 80-90%. hTERT-RPE1 and Sh-SY5Y have doubling time of approximately 24 h. To passage cells, cells were trypsinised with 0.05% trypsin and 0.02% EDTA in 1x PBS at 37°C for 5 min. When cells were detached from the plate, trypsin was inactivated by the addition of fresh growth media before cells were subcultured into plates containing fresh growth media where they were maintained at confluency between 20-80%. Mycoplasma testing was performed every 3 months.

For freezing and long term storage, adherent cells were first trypsinised and centrifuged at 250 g. Supernatant was decanted and the pellet0 resuspended in a mixture of 70% FBS, 20% growth media and 10% DMSO and aliquoted into cryotubes. All cells were placed in a Mr. Frosty Freezing Container (Thermo Scientific, Wilmington, DE, USA) at room temperature for overnight storage at -80°C before transfer into liquid nitrogen for long-term storage. To wake up cells, the contents of a frozen vial were warmed in a 37°C water bath and transferred into 5 ml of fresh growth media, cells were spun down and then transferred to a dish containing fresh growth media.

For proliferation analysis, hTERT-RPE1 cells were cultured in triplicate, starting from equal cell densities of 8x10⁴ cells in a 6-well dish. Cells were then counted every 24 h for
96 h. When cells approached confluency (4x10^5 cells/well), the cultures were diluted and the dilution factor was taken into account when calculating and plotting the cell numbers.

2.6.2 Stable cell line generation

To generate stably transfected cell lines, cells were plated to be 70% confluent on the following day in antibiotic-free growth media. For generation of SDCCAG8 null cells using CRISPR/Cas9 technology, 2 µg of pX330-Ex2 or pX330-Ex3 and 0.7 µg of pLoxNeo plasmids were complexed with Lipofectamine 2000 (Invitrogen) in Opti-MEM (Gibco) for 20 min. For other stable cell lines, linearised plasmid was complexed with Lipofectamine 2000. The mixture was added to cells and incubated at 37°C for 4–6 h. After 24 h of recovery, cells were trypsinised and serial dilutions were performed into media containing G418 (Invivogen) at a final concentration of 1mg/ml and were incubated for at least 3 days, after which cells were incubated in antibiotic free growth media at 37°C until single colonies appeared, after 10-14 days. Single colonies were picked and expanded using 3 mm Scienceware cloning discs (Sigma-Aldrich) into 48-well plates (Sarstedt). Cells were further expanded and screened by immunoblotting and immunofluorescence microscopy and DNA sequencing (Source Bioscience, Waterford, Ireland).

2.6.3 Transient transfections

For transient plasmid DNA expression, cells were plated to attain confluency of 70-80% on the day of transfection. The next day, cells were transfected with 1:2 DNA to lipofectamine 2000 ratio complexed in Opti-MEM serum-free medium for 20 min. The mixture was added to cells and incubated for 4 h before replacing with fresh growth media. The cells were then incubated for the indicated time at 37°C before fixation for immunofluorescence (section 2.6.6.2) or immunoblot (section 2.7.4) to analyse protein expression levels.

2.5.5 Serum starvation

To induce primary cilia formation in hTERT-RPE1 or SH-SY5Y, cells were plated to be 70-80% confluent the next day. The following day, cells were washed twice with 1x PBS
and once with serum free DMEM F-12 before incubation in DMEM-F12 supplemented with 0.1% FBS and 1% pen-strep for up to 48 h. For microscopy analysis of modified tubulin, cells were incubated on ice for 30 min to depolymerise the microtubule cytoskeleton before fixation for microscopy (section 2.6.6.1)

2.6.6 Microscopy methods

2.6.6.1 Cell fixation

Coverslips were sterilised by exposure to ultraviolet radiation under the laminar flow. Cells were grown on sterile coverslips to attain 30-50% confluency before immunofluorescence (IF) microscopy. Cell medium was aspirated and cells were either fixed by methanol or paraformaldehyde treatment dependent on the nature of the antibody. For methanol fixation, cells were fixed and permeabilised in 95% methanol containing 5 mM EGTA at -20°C, for 10 min. Cells were then washed three times in 1x PBS before proceeding to staining or IF microscopy (section 2.6.6.1). For paraformaldehyde fixation, cells were fixed using 4% paraformaldehyde for 10 min at room temperature and were permeabilised with permeabilisation buffer (see Table 2.1) for 2 min. Cells were then washed thrice in 1x PBS before staining for IF microscopy.

2.6.6.2 Immunofluorescence microscopy

Prior to antibody staining, cells were incubated in IF blocking buffer (Table 2.1) for 45 min at room temperature. Cells were incubated with appropriate primary antibodies (see Table 2.4) diluted in IF blocking buffer for 1 h at room temperature, followed by three 5 min washes with 1x PBS. This was followed by incubation with appropriate fluorophore-labelled secondary antibodies (see Table 2.5), diluted in IF blocking buffer in a dark humid chamber at room temperature for 45 min. The slides or coverslips were washed three times for 5 min in 1x PBS and mounted in 80% v/v glycerol containing 3% N-propyl-gallate in 1x PBS. DNA was stained with Hoechst 33258 (Sigma Aldrich) and coverslips were sealed with nail varnish and stored at 4°C in the dark. Cell images were captured using an Olympus IX81 microscope (Hamamatsu C4742-80-12AG camera) with a 100× oil objective, numeric aperture (NA) 1.35 and a Z-step of 0.4 μm, using
Volocity software v6.2.1 (Improvision Perkin-Elmer, Coventry, UK). Images are presented as maximum intensity projections of Z-stacks after deconvolution.

2.7 Protein methods

2.7.1 Protein extraction and preparation

To isolate cellular proteins for analysis, cells were harvested, washed in 1x PBS and pelleted at 250 g for 5 min. Cell pellets were resuspended in the appropriate volume (20-50 µl) of cell lysis buffer (Table 2.1) with 1% eComplete™ Protease Inhibitor Cocktail (Roche) for 20 - 40 min on ice with occasional vortexing. Samples were then centrifuged for 20 min at top speed at 4°C and the supernatant transferred to a fresh 1.5 ml eppendorf tube. Total protein concentration of the supernatant was determined by Bradford assay, which is a dependent on binding of Coomassie Brilliant Blue G-250 dye to amino acids in an acidic medium that results in stoichiometric change in dye wavelength. To perform this assay, 1 µl of a protein sample was diluted in 1 ml of 1:1 MilliQ H2O: Bradford solution (Sigma-Aldrich) and incubated at room temperature for 5 min. The absorbance was measured at 595nm using a BioPhotometer (Eppendorf) or ThermoScientific Nanodrop 2000c Spectrophotometer (Nanodrop Products, Wilmington, DE, USA) and the protein concentration was interpolated based on BSA standard curve (y=0.0595x-0.0154 or y=0.0641x+0.0216) for the respective spectrophotometers. For the BSA standard curve, absorbance was plotted against increasing concentrations of the BSA protein.

2.7.2 SDS-PAGE

SDS-PAGE gel provides pore sizes through which protein samples are separated based on their molecular weights and the desired gel percentage (composition in Table 2.8) is dependent on the size of the protein of interest. To set a desired SDS-PAGE gel percentage (usually 8-15%), glass plate and gel combs were cleaned with distilled water followed by 100% ethanol which was wiped off with lint-free paper towels. The gel apparatus was set up and resolving gel was poured and gently overlaid with 100% isopropanol. After at least 20 min of allowing resolving gel to set, the isopropanol was
removed, washed and blotted off with whatman papers. The stacking gel was then poured, a clean gel comb was inserted and allowed to set for at least 20 min at room temperature. After polymerisation, gel was submerged in ddH2O and the gel comb was gently removed, water was discarded and using syringe, the un-polymerised gel and water were removed and the gel wells were straightened and gently rinsed with 1x TG running buffer. The gel apparatus chamber was filled with 1x TG running buffer and protein samples were loaded.

For loading on SDS-PAGE gel, appropriate protein concentration (20-80 µg of whole cell lysate) was transferred into a fresh tube and 5x protein sample buffer (Table 2.1) was added to the samples, prior to boiling at 95°C for 5 min. The extract was either stored at -80°C or directly loaded on SDS-PAGE gel alongside a prestained broad range protein marker to estimate protein molecular size. Electrophoresis was performed at a constant current of 30-70 mA for 90-120 min in 1x TG running buffer using PerfectBlue Vertical Electrophoresis systems (VWR, Leuven, Belgium).

Table 2.8: SDS-PAGE gel composition used for protein detection

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Resolving gel</th>
<th>Stacking gel</th>
</tr>
</thead>
<tbody>
<tr>
<td>30% Acrylamide mix (37:5:1)</td>
<td>8-15%</td>
<td>4.5%</td>
</tr>
<tr>
<td>Tris-HCl, pH8.8</td>
<td>275mM</td>
<td>-</td>
</tr>
<tr>
<td>Tris-HCl, pH6.8</td>
<td>-</td>
<td>125mM</td>
</tr>
<tr>
<td>Sodium dodecyl sulphate (SDS)</td>
<td>0.1%</td>
<td>0.1%</td>
</tr>
<tr>
<td>Ammonium persulfate (APS)</td>
<td>0.1%</td>
<td>0.1%</td>
</tr>
<tr>
<td>Tetraacetylethlenediamine (TEMED)</td>
<td>0.05%</td>
<td>0.1%</td>
</tr>
</tbody>
</table>

2.7.3 Semi-dry transfer

Following protein separation on SDS-PAGE gel, the gel was rinsed with water and equilibrated in 1x TG transfer buffer (see table 2.1) for 5-10 min. A pre-soaked nitrocellulose (0.45 µm pore size) membrane (GE Healthcare) was placed on 2 whatman papers (Hartenstein, Würzburg, Germany) pre-soaked in 1x TG transfer buffer. The equilibrated gel was gently placed on the membrane, then, another 2 pre-soaked whatman
sheets was placed on the gel and air bubbles carefully rolled out. The membrane/gel sandwich was assembled on the appropriate transfer apparatus in a way that the membrane was closest to the anode and the gel to the cathode. Proteins were transferred from the SDS-PAGE gel to the nitrocellulose membrane using Hoefer TE 77 semi-dry transfer unit (GE Healthcare) at 1 mA/cm² for 2 h.

2.7.4 Immunoblotting technique

To visualise and assess transfer efficiency, membranes were washed with ddH2O and then stained with Ponceau S solution (Table 2.1) for 5 min at room temperature. This was followed by several washes with ddH2O until distinct protein bands were visible and a photograph of the stained membrane was taken as a reference. Blots were then incubated in WB blocking solution (Table 2.1) for 45 min at room temperature. This was followed by membrane incubation with the appropriate primary antibody (Table 2.4) dilution in 1x TBST for 1.5 h at room temperature or overnight at 4°C. To remove unbound primary antibody, membranes were washed for three times with 1x TBST and incubated in HRP-conjugated secondary antibody (Table 2.5) in 1x TBST for 45 min at room temperature. membranes were then washed thrice with 1x TBST. All of the above washes and incubations were carried out with constant agitation on a rocking platform. To facilitate protein detection, ECL (GE Healthcare) was used with strict adherence to the manufacturer’s instructions. Briefly, membranes were covered with a 1:1 mixture of the ECL solutions and incubated for 5 min followed by autoradiograph film exposure and development on a CP 1000, AGFA film processor (Superior Radiographics Ltd, Madison, WI, USA).

2.8 Nucleic acid methods

2.8.1 RNA isolation and cDNA synthesis

For total RNA extraction RNeasy (Qiagen) was used according to manufacturer’s instructions. Cells were lysed directly from plates using buffer RLT and the optional on-column DNA digestion was performed. RNA was eluted from the column in Nuclease free water (Thermo Scientific) and RNA concentration was determined on
ThermoScientific Nanodrop 2000c Spectrophotometer at 260 nm and samples stored at -80°C. Using 2 µg extracted RNA, cDNA was synthesised using High-Capacity RNA-to-cDNA kit (Applied Biosystems). The reactions were set up according to the manufacturer’s instructions and the cDNA generated was used as a template in PCR reactions (section 2.8.4) for the amplification of specific cDNAs.

2.8.2 Genomic DNA extraction

Cells were harvested from a confluent 6 well dish into a 1.5 ml microfuge tube and incubated overnight at 37°C in a 500 µl Tail buffer (Table 2.1) containing 0.5 mg/ml Proteinase K (Sigma-Aldrich). On the following day, the cell lysate was subjected vigorous mixing on a thermomixer at 1400 rpm for 5 min after which 240 µl of 5 M NaCl was added to the tubes. Cell lysates were mixed for an additional 5 min before being centrifuged at 4°C, top speed for 10 min to remove precipitated proteins. The supernatants were transferred into new microfuge tubes and DNA was precipitated with 750 µl of isopropanol which was added and mixed by inverting the tubes. This was followed by centrifugation at top speed, after which the supernatant was removed and was followed by a wash in 500 µl of 70% ethanol. The pellet was air dried for approximately 10 min at room temperature and the DNA was resuspended and solubilised in 50 µl of Milli-Q H2O at 37°C for at least 1 h. Genomic DNA was used for PCR amplification of specific sequences (section 2.8.4).

2.8.3 Plasmid DNA prep

Mini and midi plasmid DNA were prepared using the GenElute™ or Nucleobond Xtra Midi (Endotoxin-free) kits, respectively. In these procedures, plasmids were isolated according to the manufacturer’s instructions. For miniprep, a single bacterial colony was used to inoculate 5 ml of LB broth with appropriate antibiotic (see section 2.2.1) and incubated with shaking overnight at 37°C. 4 ml of the overnight culture was used for miniprep DNA extraction while 100 ml overnight culture was used for midiprep extractions. The resulting plasmids were resuspended in 50 µl and 160 µl Milli-Q water, respectively.
2.8.4 Polymerase Chain Reaction (PCR)

PCR was performed to amplify specific DNA sequence for sub-cloning into linearised vectors that was subsequently used for functional analysis. Usually, KOD Hot Start DNA Polymerase from Novagen (Darmstadt, Germany) was the DNA polymerase used. The annealing temperature and magnesium concentration were optimised for each reaction and a representative PCR reaction is shown in Table 2.9.

Table 2.9: PCR conditions for KOD Hot Start

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Final concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCR Buffer</td>
<td>1X</td>
</tr>
<tr>
<td>Primers</td>
<td>200 µM</td>
</tr>
<tr>
<td>dNTPs</td>
<td>200 nM</td>
</tr>
<tr>
<td>Mg(^{2+})</td>
<td>1.5 mM</td>
</tr>
<tr>
<td>DMSO</td>
<td>1-5%</td>
</tr>
<tr>
<td>Plasmid/cDNA/ genomic DNA (Template)</td>
<td>10 ng /100 ng</td>
</tr>
<tr>
<td>Enzyme</td>
<td>0.02 U/µl</td>
</tr>
</tbody>
</table>

All PCR reactions were performed on an Eppendorf Mastercycler Nexus Gradient (Hamburg, Germany) and typical program cycling conditions are listed in Table 2.10.

Table 2.10: PCR programme cycling conditions.

<table>
<thead>
<tr>
<th>PCR step</th>
<th>Temperature</th>
<th>Time</th>
<th>Number of cycles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial denaturation</td>
<td>95°C</td>
<td>2 min</td>
<td>1 cycle</td>
</tr>
<tr>
<td>Denaturation</td>
<td>95°C</td>
<td>20 s</td>
<td></td>
</tr>
<tr>
<td>Annealing</td>
<td>50-65°C</td>
<td>10 s</td>
<td>30 cycles</td>
</tr>
<tr>
<td>Extension</td>
<td>70°C</td>
<td>25 s/Kb</td>
<td></td>
</tr>
<tr>
<td>Final extension</td>
<td>72°C</td>
<td>5 min</td>
<td>1 cycle</td>
</tr>
</tbody>
</table>
Chapter 2

A list of primers used in this study are detailed in Table 2.11:

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Use</th>
<th>Sequence (5’→3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDCCAG8 Fwd</td>
<td>Cloning full length human SDCCAG8</td>
<td>taagcagaattctATGGCGAAGTCCCCGGAGAACTCT</td>
</tr>
<tr>
<td>SDCCAG8 Rev</td>
<td></td>
<td>tgcttagattcctGCAATCAGATTGTGCGCATGCTGGG</td>
</tr>
<tr>
<td>MAD1L1 Fwd</td>
<td>Cloning full length human MAD1L1</td>
<td>taagcagaattctATGGAAGACCTGCGGGGAACACCC</td>
</tr>
<tr>
<td>MAD1L1 Rev</td>
<td></td>
<td>tgcttagattcctCGCCACGGTCTGGCGGCTGGAG</td>
</tr>
<tr>
<td>PRKD1 Fwd</td>
<td>Cloning full length human PRKD1</td>
<td>taagcagaattctATGAGCGCCCCTCCGTCCTGCGGG</td>
</tr>
<tr>
<td>PRKD1 Rev</td>
<td></td>
<td>tgcttagattcctGAGGATGCTGACACGCTCACC</td>
</tr>
<tr>
<td>GIGYF2 Fwd</td>
<td>Cloning full length GIGYF2</td>
<td>taagcactegagtATGGCGAGCGAAACGCAGACACTG</td>
</tr>
<tr>
<td>GIGYF2 Rev</td>
<td></td>
<td>tgcttagattcctGTAGTCATCCAAACGTCTCGATTTCC</td>
</tr>
<tr>
<td>MAPK3 Fwd</td>
<td>Cloning full length MAPK3</td>
<td>taagcagaattctATGGCGCCGCGCGGCGCGCGCTAGGGG</td>
</tr>
<tr>
<td>MAPK3 Rev</td>
<td></td>
<td>tgcttagattcctGGGGGCTCCAGCAGCCTCGGGCTG</td>
</tr>
<tr>
<td>SDCCAG8 Exon 3 Fwd</td>
<td>CRISPR gRNA targeting SDCCAG8</td>
<td>caceGTGGCGCAACAAAGCAGA</td>
</tr>
<tr>
<td>SDCCAG8 Exon 3 Rev</td>
<td></td>
<td>aaacTATCTGCTGTGGCGCAAC</td>
</tr>
<tr>
<td>SDCCAG8 Clone Fwd</td>
<td>SDCCAG8 CRISPR clone screening</td>
<td>AGTAAGGTGCGTGGTTGGTGCAACC</td>
</tr>
<tr>
<td>SDCCAG8 Clone Rev</td>
<td></td>
<td>TATGCCATGCTTTAAAAATTTTTTCTGC</td>
</tr>
</tbody>
</table>

Lower cases indicate restriction enzyme sequences and bases required for restriction enzyme cleavage close to DNA fragment ends. The capital letters are regions complementary to the region of interest to be amplified.

2.8.5 Restriction digestion of DNA

Restriction digestion patterns were predicted for DNA sequences using the NEBCutter bioinformatics tool or SnapGene software (Table 2.12). NEB supplied all restriction endonucleases used. DNA digestion reactions were set up according to the manufacturer's protocol supplied and incubated at the recommended optimal temperature for the restriction enzyme for minimum of 2 h. Where necessary, restriction enzymes were heat inactivated at manufacturer’s recommended temperature and the digestions were analysed by agarose gel electrophoresis (section 2.8.9).
2.8.6 Modification of DNA ends

2.8.6.1 Phosphorylation of DNA ends

After PCR reaction, DNA inserts needed for ligation were subjected to 5’ DNA end phosphorylation and removal of 3’ phosphates by T4 Polynucleotide Kinase (T4 PNK) from NEB. The reaction was carried out as per the manufacturer’s instructions, briefly, 1x T4 DNA ligase buffer containing 1 mM ATP and 1 U T4 PNK per µg DNA were added to each purified PCR reaction and incubated at 37°C for 30 min. Enzyme inactivation was achieved by incubating the reaction mixture at 65°C for 20 min.

2.8.6.2 Dephosphorylation of DNA ends

To minimise re-circularisation of linear DNA needed for ligation, the linearised DNA vector was treated with recombinant shrimp alkaline phosphatase (rSAP) from USB Affymetrix to dephosphorylate the 5’ ends of linear DNA. The enzyme was used according to the manufacturer’s protocol. Briefly, after heat inactivation and ice-cooling or spin column clean-up of restriction enzyme reaction, 1 U SAP per µg DNA was added to the digestion reaction and incubated at 37°C for 30 min. This was followed by rSAP enzyme denaturation at 65°C for 10 min.

2.8.7 Ligation

The prepared plasmid and insert DNA were verified by agarose gel electrophoresis. Ligations were carried out using T4 DNA ligase in 1X T4 DNA ligase buffer at 4°C overnight or at room temperature for 3 hours prior to transformation of competent E. coli cells. An excess of insert DNA was generally used (1:3 to 1:10, as calculated from NanoDrop 2000c spectrophotometer analysis).
2.8.7 Assembly of CRISPR plasmids

To build CRISPR targeting vectors used for generating knockout clones, oligonucleotides were annealed at 95°C for 5 min and then cooled slowly on the thermoblock. The oligos were then phosphorylated with T4 Polynucleotide Kinase (section 2.8.6.1). The pX330 (Addgene) plasmid was digested with BbsI restriction endonuclease and dephosphorylated with rSAP (section 2.8.6.2). Annealed and phosphorylated oligonucleotides were then ligated to the digested pX330 plasmid and successful cloning was verified by commercial DNA sequencing.

2.8.8 Analysis of cloned DNA and sequencing

Usually, 300 ng of DNA from mini or midi prep was analysed by restriction digestion (section 2.8.5) and commercial DNA sequencing (Source Bioscience). Plasmid DNA sequencing was performed to ensure nucleic acid fidelity and in-frame insertion of DNA fragments following cloning. Typically, 100 ng of mini or miniprep and 10 pM sequencing primers were sent to Source Bioscience for Sanger sequencing and the output data were analysed on SnapGene (Table 2.12) to construct correct vector maps with the analysed sequences.

2.8.9 Agarose gel electrophoresis

Agarose gel is the standard method used to separate, identify, and in some cases, to purify DNA fragments. Usually, a 0.5-1.2% agarose gel was prepared using electrophoresis grade agarose (Sigma-Aldrich) in 1X TAE buffer (Table 2.1) containing 0.5 µg/ml ethidium bromide. The Hoefer HE33 or HE99x Horizontal Submarine Unit (ThermoFisher Scientific) were used according to the manufacturer’s instructions for a maximum of 2 h. After completion of electrophoresis, separated DNA molecules were viewed on the gel using a Multi-Image Light Cabinet (ChemiImager 5500, Alphalmager, ProteinSimple, San Jose, CA, USA) and images were taken with a digital camera. For DNA extraction, bands of interest were excised from the agarose gel with a scalpel blade under UV light and the DNA was purified as described in Section 2.8.10.1
2.8.10 DNA purification methods

2.8.10.1 Gel extraction and Column clean-up

DNA was purified from agarose gel slices (section 2.8.9) using the Qiagen QIAquick Gel Extraction Kit according to the manufacturer’s instructions. Typically, a gel slice was placed in a 2 ml microfuge tube and bound DNA was eluted off the column in 20-40 µl of warm Milli-Q H2O. Column clean-up was performed using QIAquick PCR purification kit according to the manufacturer’s protocol and bound DNA elution was in 20-40 µl of warm MilliQ H2O.

2.8.10.2 Ethanol precipitation

For purifying linearised plasmid DNA before transfections, one-tenth volume of 3 M sodium acetate, pH 5.2 and 1 volume of 100% isopropanol were added to the DNA solution. The sample was vortexed and the DNA was pelleted at 4°C, top speed for 10 min. The supernatant was removed and the pellet was washed with 500 µl of 70% ethanol. The sample was then centrifuged at 4°C top speed for 5 min, the supernatant was also discarded and the pellet was air-dried at room temperature for 10 min. The pellet was resuspended in an appropriate volume of MilliQ H2O and stored at -20°C.

2.8.11 Preparation and transformation of chemically competent *E.coli*

Chemically competent Top10 *E. coli* competent cells were prepared by inoculating an overnight mini culture into 500 ml of LB broth in a sterile 2 L flask which was incubated at 37°C with shaking until an OD₆₀₀ between 0.4 and 0.6 was reached, usually about 2 h. The bacterial cells were chilled on ice before centrifugation at 6000 g for 10 min at 4°C. The pellet was resuspended in 250 ml of ice-cold 0.1 M CaCl₂ and incubated on ice for 30 min. The cells were then spun down again under the same conditions stated above and the pellet was resuspended in 50 ml of cold 0.1 M CaCl₂ supplemented with 15% glycerol before 50 µl aliquots were prepared in the cold room and flash frozen on dry ice. The competent cell aliquots were stored at -80°C.
To transform competent cells, an aliquot of *E. coli* was thawed on ice and was transformed by heat shock as described in Sambrook et al. 1989. In general, 5-10 µl of a ligation reaction or 0.5 µl of plasmid DNA was added and gently mixed by tapping with the thawed E. coli under sterile conditions. The cells were placed on ice for 20 min, heat shocked for 90 s at 42°C, followed by 2 min incubation on ice. After incubation, 900 µl Luria Broth (LB) was added to cells and were then incubated shaking (300-350 rpm) at 37°C for 1 h. Transformed cells were plated on LB plates containing the appropriate antibiotic and incubated overnight at 37°C. Transformed cells were resistant to the antibiotic and cells yielded colonies. These colonies were subsequently used to prepare broth cultures for DNA mini or midi-preparations after screening.

2.9 Migration analysis

Cells were serum starved for 24 h, trypsinised cells were suspended in 0.1% FBS medium and 3.5*10-5 cells were added to the upper chamber of transwell inserts (6.5 mm Transwell® with 8.0 µm pore polyester (PET) insert, Corning® Costar®). Medium with 10% serum was added to the lower chamber. After 16 and 24 h incubations for SH-SY5Y cells, non-migrated cells on the upper surface of the membrane were scrapped off, cells on the lower surface were stained with 0.5% crystal violet and photographed. The crystal violet was then solubilized using 10% acetic acid, absorbance was read at 595 nm. Percentage of migrated cells was calculated relative to a positive control of 3.5*10-5 cells plated the bottom of the 24-well dish. The experiment was repeated twice.

2.10 Differentiation of SH-SY5Y cells

On day one 1.5*10^5 cells in 10% FBS media were plated in 6-well dishes that had been coated with 4µg/cm² poly-lysine and 1µg/cm² lamin. Day 2 media was replaced with media containing 2.5% FBS and 10µM retinoic acid. Day 4 media was replaced with media containing 1% FBS and 10µM retinoic acid. Day 6 media was replaced with media containing 0.1% FBS and 10µM retinoic acid. On day 8 cells were fixed or harvested.
2.9 RNA-Seq

I completed RNA extraction, constructed cDNA libraries and performed differential expression analysis for all RNA-Seq experiments. The Genomics Core Technology Unit at Queen’s University Belfast, aligned all reads to the genome, aggregated reads by gene and assessed the quality of the data generated.

2.9.1 Cell culture conditions and RNA extraction

Cells were plated to be 80% confluent the next day, the following day either total RNA was extracted (section 2.8.1) or cells were serum starved (section 2.6.5) to induce primary cilium formation, in which RNA was extracted 48 h post serum starvation. The yield and quality of the isolated RNA were assessed using a NanoDrop8000 Spectrophotometer (Thermo Scientific) and Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA), respectively.

2.9.2 cDNA library construction and sequencing

For stranded RNA-Seq, cDNA libraries were prepared with a TruSeq stranded mRNA library prep Kit (cat# RS-122-2101, Illumina, San Diego, CA, USA) as per the manufacturer’s protocol. The quality of cDNA generated was assessed using the Agilent 2100 Bioanalyzer (Agilent Technologies). The resulting libraries were sequenced on the NextSeq 550 (Illumina) using a paired-end run (2 × 75 bases). A minimum of 20 M reads were generated per sample.

2.9.3 Mapping and counting

Raw fastq files were aligned to human genome build GRCh37 using Spliced Transcripts Alignment to a Reference (STAR) to produce bam files. STAR is a high precision mapping strategy that allows for accurate alignment of high throughput RNA-Seq data (Dobin et al., 2013). Reads were then aggregated by gene using HTSeq-counts in order to produce raw counts per gene per sample. Only reads mapping unambiguously to a single gene are counted (Anders et al., 2015)
2.9.4 Differential expression analysis

HTSeq-count files were used for differential expression (DE) analysis. The DE analysis was performed by the R package DESeq2. SDCCAG8 KO cells were compared to wildtype cells under baseline and serum starved conditions in SH-SY5Y and RPE1 cells. Briefly DESeq2 perform internal normalisation where a geometric mean is calculated for each gene across all samples inputted. The median of all these ratios becomes the size factor for that sample and counts are then divided by this size factor. This effectively corrects for sequencing depth discrepancies and RNA composition bias. DESeq2 uses shrinkage estimators for dispersion and fold change, which leads to improved gene ranking and a highly sensitive and precise methodology that controls effectively for false positives (Love et al., 2014). All genes with an absolute log2 fold change of greater than one and a Benjamini-Hochberg adjusted p-value less than 0.05 were considered differentially expressed genes.

2.9.5 Pathway Analysis

Ingenuity pathway analysis (IPA) was used to identify gene ontologies, pathways, and regulatory networks to which DE genes belong to, as well as upstream regulators. IPA has a comprehensive, manually curated Ingenuity Knowledge Base which is used in algorithms to test for enrichment of genes in particular pathways using Fisher’s exact test to calculate enrichment P-Values. A Z-score is also calculated which is a statistical measure of how well the observed gene expression data matches the expected relationship direction. ConsensusPathDB-human (CPDB, http://cpdb.molgen.mpg.de/), an open source online overrepresentation gene-sets analysis tool was also used.

2.10 Computer programmes and Bioinformatics tools

A number of softwares and online bioinformatics tools were routinely used in this study. A list of such is shown in Table 2.12.
<table>
<thead>
<tr>
<th><strong>Tool</strong></th>
<th><strong>Source</strong></th>
<th><strong>Use</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Adobe illustrator CS6</td>
<td>Adobe Systems</td>
<td>Processing images and Figures</td>
</tr>
<tr>
<td>BLAST</td>
<td><a href="https://www.ncbi.nlm.nih.gov/BLAST">https://www.ncbi.nlm.nih.gov/BLAST</a></td>
<td>Comparing DNA or protein sequences to database for analysis</td>
</tr>
<tr>
<td>ConsensusPathDB-human</td>
<td><a href="http://cpdb.molgen.mpg.de/">http://cpdb.molgen.mpg.de/</a></td>
<td>Overrepresentation analysis of gene-sets</td>
</tr>
<tr>
<td>ClustalW</td>
<td><a href="https://www.ebi.ac.uk/Tools/msa">https://www.ebi.ac.uk/Tools/msa</a> /</td>
<td>Protein and DNA sequence alignment</td>
</tr>
<tr>
<td>FASTQC</td>
<td><a href="https://www.bioinformatics.babraham.ac.uk/projects/fastqc/">https://www.bioinformatics.babraham.ac.uk/projects/fastqc/</a></td>
<td>Used to run quality metrics for raw fastq files</td>
</tr>
<tr>
<td>GraphPad Prism 5.0c</td>
<td>Graphpad, La Jolla, CA, USA</td>
<td>Statistical analysis of acquired data</td>
</tr>
<tr>
<td>HTSeq-counts</td>
<td><a href="https://htseq.readthedocs.io/en/release_0.9.1/counting.html">https://htseq.readthedocs.io/en/release_0.9.1/counting.html</a></td>
<td>Generating raw counts per gene per RNA-Seq sample</td>
</tr>
<tr>
<td>ImageJ</td>
<td>imagej.nih.gov/ij/</td>
<td>Processing and quantifying image signals</td>
</tr>
<tr>
<td>Ingenuity Pathway Analysis</td>
<td>(IPA, QIAGEN Redwood City, <a href="http://www.qiagen.com/ingenuity">www.qiagen.com/ingenuity</a>)</td>
<td>Pathway analysis on differentially expressed genes from RNA-Seq analysis</td>
</tr>
<tr>
<td>LocusZoom</td>
<td><a href="http://locuszoom.org/">http://locuszoom.org/</a></td>
<td>Preparing regional plots for GWAS data</td>
</tr>
<tr>
<td>MAGMA</td>
<td><a href="https://ctg.cncr.nl/software/mag">https://ctg.cncr.nl/software/mag</a> ma</td>
<td>Gene-set analysis of GWAS data</td>
</tr>
<tr>
<td>Tool Name</td>
<td>URL/Link</td>
<td>Description</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>NEBcutter</td>
<td><a href="http://www.labtools.us/nebcutter-v2-0/">http://www.labtools.us/nebcutter-v2-0/</a></td>
<td>Identifying restriction enzyme sites on DNA sequence</td>
</tr>
<tr>
<td>NEBcloner</td>
<td><a href="https://nebcloner.neb.com/#!/redigest">https://nebcloner.neb.com/#!/redigest</a></td>
<td>Determining buffer compatibility for restriction endonuclease</td>
</tr>
<tr>
<td>Nucleic Acid Sequence Massager</td>
<td><a href="http://www.attotron.com/cybertoany/analysis/seqMassager.htm">http://www.attotron.com/cybertoany/analysis/seqMassager.htm</a></td>
<td>To convert DNA sequence into its reverse, complement, or reverse complement sequence</td>
</tr>
<tr>
<td>OligoCalc</td>
<td><a href="http://biotools.nubic.northwestern.edu/OligoCalc.html">http://biotools.nubic.northwestern.edu/OligoCalc.html</a></td>
<td>Calculating primer properties such as Tm, self-complementarity etc.</td>
</tr>
<tr>
<td>PLINK</td>
<td><a href="http://zzz.bwh.harvard.edu/plink/">http://zzz.bwh.harvard.edu/plink/</a></td>
<td>Extracting GWAS data on SZ risk SNPs</td>
</tr>
<tr>
<td>R Studio</td>
<td><a href="https://www.rstudio.com/">https://www.rstudio.com/</a></td>
<td>Differential expression analysis of RNA-Seq data</td>
</tr>
<tr>
<td>SnapGene</td>
<td><a href="http://www.snapgene.com/">http://www.snapgene.com/</a></td>
<td>Drawing DNA plasmid maps and identifying restriction sites on DNA sequence</td>
</tr>
<tr>
<td>Statistical Package for the Social Sciences</td>
<td>IBM (Armonk, New York, United States)</td>
<td>Statistical analysis of cognitive scores</td>
</tr>
<tr>
<td>STAR</td>
<td><a href="https://github.com/alexdobin/STAR">https://github.com/alexdobin/STAR</a></td>
<td>Aligning raw fastq files to the human genome</td>
</tr>
<tr>
<td>Translation tool</td>
<td>Expasy.org</td>
<td>To translate nucleotide sequence to protein sequence</td>
</tr>
<tr>
<td>Volocity v6.21</td>
<td>Improvision/Perkin-Elmer</td>
<td>Acquisition and analysis of microscopy images</td>
</tr>
</tbody>
</table>
3 Investigation of centrosomal genes for contribution to SZ and cognition

3.1 Introduction

Embryonic development of the central nervous system requires the proliferation of neural progenitor cells (NPC) to be tightly regulated, resulting in an organ with the right size and shape. Centrosomes control the internal spatial organization of somatic cells, and are key regulators of cell division, cell polarity and migration. Upon exiting the cell cycle, most cell types convert their mother centriole into a basal body, which seeds the growth of a primary cilium, which plays an important role in sensing and signal transduction at the cell surface. As one of the most structurally complex organs in our body, the brain is particularly susceptible to centrosome dysfunction (Saade et al., 2018). Mutations in several centrosomal components can result in autosomal recessive primary microcephaly (MCPH), a disorder characterised by a small brain at birth. Strikingly, mutations in 77 of 87 genes that encode ciliary and centrosomal proteins associated with human ciliopathies result in neurodevelopmental or cognitive deficits (Guo et al., 2015).

As discussed in Section 1.6, several centrosomal genes have been associated with SZ risk including; DISC1, PCM1 and AHI1. Additionally knockdown of gene expression for 20 candidate genes for neuropsychiatric disorders, resulted in a ciliary phenotype (Marley et al., 2012). These studies were prior to the advent of large GWASs for neuropsychiatric disorders. To my knowledge the contribution of centrosomal genes to neuropsychiatric conditions and cognition has not been assessed in light if this new data.

A meta-analysis of over 50 GWAS studies performed by the Psychiatric Genomics Consortium (PGC) has culminated in the identification of 108 risk loci for SZ that contain some 350 genes (Schizophrenia Working Group of the Psychiatric Genomics, 2014). This study asks: do any genes at new SZ risk loci, identified through GWASs regulate centrosomal functioning? For those loci containing those genes, I next ask whether they independently associate with cognitive functions? The overall hypothesis is that GWAS-
identified risk genes for SZ, which regulate centrosomal functioning, are associated with deficits in cognitive function.

3.2 Identification of SZ risk genes with centrosomal function

When I started my PhD in 2015, the PGC had identified 108 loci associated with SZ in the largest SZ GWAS at the time. These regions contained 350 genes (Schizophrenia Working Group of the Psychiatric Genomics, 2014). This gene list was cross-referenced with my centrosomal gene list (Supplementary Table 3.1), which was generated by combining data from MiCroKiTS 4.0 and Centrosome:db. This identified 21 genes with centrosomal functions which are located in SZ risk regions (Figure 3.1).

![Figure 3.1: Identification of SZ risk genes with centrosomal function.](image)

Venn diagram showing overlap of genes with centrosomal function and the genes present in regions associated with SZ risk. Twenty-one genes are common to both gene-sets. Gene symbols for each gene-set are listed in Supplementary Table 3.1.

Some of the centrosomal genes are located in SZ risk loci that are gene-rich. Therefore, I assessed whether SZ risk could be confidentially attributed to the identified centrosomal genes. Following a review of the GWAS regional plots containing the 21 genes, the candidate gene list was reduced to five genes (MPHOSPH9, GIGYF2, PRKD1, MAD1L1, and SDCCAG8), where each gene is located directly under the index risk SNP (i.e. the most associated SNP). Two of these genes (MAD1L1 and PRKD1) were the only protein-coding genes within their SZ risk locus; the regional plots for these genes are shown in Figures 3.2 and 3.3, respectively. For the remaining three genes, the positional information and the LD patterns from the regional plots strongly suggested that the
centrosomal genes were the likely SZ risk gene, which can be seen in Figures 3.4, 3.5 and 3.6. Based on the analysis of results from a meta-analysis of expression quantitative trait loci (eQTLs) in the brain (Kim et al., 2014), I also included MAPK3 in this analysis, as it was clear that index SNP rs12691307 was significantly associated with the expression of this gene. The regional plot for rs12691307 is depicted in Figure 3.7.
Regional plots were generated with LocusZoom (http://locuszoom.org/) using SZ summary statistics. The x-axis is chromosomal position, the y-axis is significance of association with SZ represented as $-\log_{10}(P)$. Index SNP rs58120505 is indicated with a diamond. Colouring is based on degree of LD with the index SNP as represented by $r^2$. Legend for $r^2$ is given in the upper left corner. The blue line denotes regional recombination rates derived from HapMap. Lower panel displays genes in the region, vertical blue lines indicate exons and arrowheads indicate the direction of transcription.
Regional plots were generated with LocusZoom (http://locuszoom.org/) using SZ summary statistics. The x-axis is chromosomal position, the y-axis is significance of association with SZ represented as -log10 (P). Index SNP rs2068012 is indicated with a diamond. Colouring is based on degree of LD with the index SNP as represented by $r^2$. Legend for $r^2$ is given in the upper left corner. The blue line denotes regional recombination rates derived from HapMap. Lower panel displays genes in the region, vertical blue lines indicate exons and arrowheads indicate the direction of transcription.

**Figure 3.3 PRKD1 Regional Plot**
Regional plots were generated with LocusZoom (http://locuszoom.org/) using SZ summary statistics. The x-axis is chromosomal position, the y-axis is significance of association with SZ represented as -log10 (P). Index SNP rs2068012 is indicated with a diamond. Colouring is based on degree of LD with the index SNP as represented by $r^2$. Legend for $r^2$ is given in the upper left corner. The blue line denotes regional recombination rates derived from HapMap. Lower panel displays genes in the region, vertical blue lines indicate exons and arrowheads indicate the direction of transcription.
Figure 3.4 *MPHOSPH9* Regional Plot
Regional plots were generated with LocusZoom (http://locuszoom.org/) using SZ summary statistics. The x-axis is chromosomal position, the y-axis is significance of association with SZ represented as -log10 (P). Index SNP rs2851447 is indicated with a diamond. Colouring is based on degree of LD with the index SNP as represented by $r^2$. Legend for $r^2$ is given in the upper left corner. The blue line denotes regional recombination rates derived from HapMap. Lower panel displays genes in the region, vertical blue lines indicate exons and arrowheads indicate the direction of transcription.
Figure 3.5 *GIGYF2* Regional Plot
Regional plots were generated with LocusZoom (http://locuszoom.org/) using SZ summary statistics. The x-axis is chromosomal position, the y-axis is significance of association with SZ represented as -log10 (P). Index SNP rs6704768 is indicated with a diamond. Colouring is based on degree of LD with the index SNP as represented by $r^2$. Legend for $r^2$ is given in the upper left corner. The blue line denotes regional recombination rates derived from HapMap. Lower panel displays genes in the region, vertical blue lines indicate exons and arrowheads indicate the direction of transcription.
Regional plots were generated with LocusZoom (http://locuszoom.org/) using SZ summary statistics. The x-axis is chromosomal position, the y-axis is significance of association with SZ represented as $-\log_{10}(P)$. Index SNP rs77149735 is indicated with a red diamond, rs10803138 a blue diamond and rs14403 a green diamond. LD to each index SNP is denoted by the intensity of the same colour. The blue line denotes regional recombination rates derived from HapMap. Lower panel displays genes in the region, vertical blue lines indicate exons and arrowheads indicate the direction of transcription.

Figure 3.6 SDCCAG8 Regional Plot
Regional plots were generated with LocusZoom (http://locuszoom.org/) using SZ summary statistics. The x-axis is chromosomal position, the y-axis is significance of association with SZ represented as $-\log_{10}(P)$. Index SNP rs77149735 is indicated with a red diamond, rs10803138 a blue diamond and rs14403 a green diamond. LD to each index SNP is denoted by the intensity of the same colour. The blue line denotes regional recombination rates derived from HapMap. Lower panel displays genes in the region, vertical blue lines indicate exons and arrowheads indicate the direction of transcription.
Figure 3.7: MAPK3 Regional Plot
Regional plots were generated with LocusZoom (http://locuszoom.org/) using SZ summary statistics. The x-axis is chromosomal position, the y-axis is significance of association with SZ represented as -log10 (P). Index SNP rs1291307 is indicated with a diamond. Colouring is based on degree of LD with the index SNP as represented by $r^2$. Legend for $r^2$ is given in the upper left corner. The blue line denotes regional recombination rates derived from HapMap. Lower panel displays genes in the region, vertical blue lines indicate exons and arrowheads indicate the direction of transcription.
3.3 Prioritised genes

The final six genes (with index SNPs) for analysis are listed here, including brief details of function and descriptions of any cognitive disorders caused by these genes:

1. **MPHOSPH9** (M-Phase Phosphoprotein 9, rs2851447) is a member of a family of proteins phosphorylated in the M (mitotic) phase. MPHOSPH9 has been shown to localise to the distal and proximal end of centrioles. MPHOSPH9 appears to dissociate from the distal end of the mother centriole in cells which have ciliated, which is suggestive of a role in ciliogenesis (Jakobsen et al., 2011). An SNP within the gene has been identified as a Multiple Sclerosis risk locus (International Multiple Sclerosis Genetics Consortium, 2010).

2. **MAD1L1** (Mitotic Arrest Deficient 1 Like 1, rs10650434) is a checkpoint gene, and its dysfunction is associated with chromosomal instability (Tsukasaki et al., 2001). **MAD1L1** was shown to be affected by a singleton microdeletion as well as a singleton microduplication in early-onset bipolar disorder (BD) patients (Priebe, 2012). **MAD1L1** SNPs were the second and third most significant results in the meta-analysis of GWAS of BD (Cichon et al., 2011). A more recent BD GWAS in the Japanese population also detected a genome wide significant SNP within the **MAD1L1** locus (Ikeda et al., 2017).

3. **SDCCAG8** (Serologically Defined Colon Cancer Antigen 8, rs77149735, rs10803138, and rs14403) regulates the centrosomal accumulation of PCM and neuronal migration in the developing cortex (Insolera et al., 2014). In this case, three SNPs were analysed as they are all LD independent and each one is significantly associated with SZ. Mutations in **SDCCAG8** (BBS16) is estimated to account for 1% of Bardet-Biedl syndrome (BBS) cases. BBS is an autosomal recessive ciliopathy that presents with retinal dystrophy, obesity, polydactyly, cognitive impairment, urogenital anomalies and renal abnormalities as primary clinical features (Castro-Sánchez et al., 2015). Elements of obsessive compulsive behaviour can be seen in some patients. Other patients have more severe behavioural phenotypes and develop autistic spectrum disorder or psychosis (Forsythe and Beales, 2013). **SDCCAG8** is also a causative gene in Senior-Løken
Chapter 3

syndrome, an autosomal recessive disease the main features of which are nephronophthisis and Leber congenital amaurosis. In rare cases, the disorder presents with intellectual disability and microcephaly (Otto et al., 2010).

4.  **GIGYF2** (GRB10 Interacting GYF Protein 2, rs6704768) appears to have a role in Insulin-like Growth factor (IGF) pathway signalling (Giovannone et al, 2009). Three *de novo* truncating mutations were identified in a meta-analysis of autism patients, with no *de novo* mutations reported in 1,786 unaffected siblings. Additional *de novo* mutations in other components of the IGF signalling pathway were also identified, implicating this pathway in the development of ASD (Krumm et al., 2015). The gene has previously been implicated in Parkinson’s disease with cognitive impairment (Ruiz-Martinez et al., 2015) but these results remain inconclusive.

5.  **PRKD1** (Protein Kinase D1, rs2068012) is a serine/threonine kinase that regulates a variety of cellular functions including membrane receptor signalling, transport at the Golgi, protection from oxidative stress at the mitochondria, gene transcription, and regulation of cell shape, motility, and adhesion (Eiseler et al. 2009). The gene has been associated with atypical Rett syndrome. PRKD1 is thought to contain a FOXG1-regulatory element, deletion of which could be responsible for a subset of atypical Rett syndrome cases (Allou et al., 2012).

6.  **MAPK3** (Mitogen-Activated Protein Kinase 3, rs12691307) is a serine/threonine kinase, an essential component of the MAP kinase signal transduction pathway that regulates various cellular processes such as proliferation, differentiation, and cell cycle progression in response to a variety of extracellular signals. MAPK3, also known as ERK1, is reported to be released under neurodegenerative conditions (Klafki et al., 2009). MAPK3 has also been suggested to have a regulatory role in striatum-dependent behavioural plasticity and drug addiction (Mazzucchelli et al., 2002). Rare variants in MAPK3 have been detected in individuals with ASD (Schaaf et al., 2011). 16p11.2 microduplication is one of the CNVs most strongly associated with SZ and autism, spanning multiple genes including MAPK3. MAPK3 has been put forward as a potential driver gene within this CNV (Blizinsky et al., 2016, Park et al., 2017), although other
3.4 Neuropsychological analysis

Each of the nine index SZ risk SNPs were analysed against four areas of cognition (IQ, working memory, episodic memory and social cognition) in the broad sample of all psychosis patients and controls. Table 3.1 shows all nominally significant results. One result survives Bonferroni correction for the four cognitive domains and the nine variants tested (adjusted p-value threshold = 0.0014). For this strongest association, rs14403 at *SDCCAG8*, the SZ risk allele C is associated with an increased likelihood of attributing negative events to external factors relative to positive events (p=0.001) with an R² value of 0.021, which indicates that 2.1% of the variance in Externalisation Bias (EB) is explained by rs14403 genotype. The variant had another significant result within the social cognition domain on the Reading the Mind in the Eyes Task (p=0.015) and in the domain of working memory on the Letter Number Sequencing task (p=0.019). Another SNP within the same gene, rs5782266, also had a significant result for EB (p=0.012).

For four of the remaining five genes there was at least one nominally significant result. *MAD1L1* had significant results in four different measures of episodic memory, the most significant of which was on the Logical Memory delayed condition task (p=0.007) where the SZ risk allele, an insertion of bases CT, was associated with poorer memory performance. *MPHOSPH9* also had significant results for episodic memory on Logical Memory I&II (p=0.025 and p=0.044), as well as EB (p=0.006). *PRKD1* had significant results for performance IQ (p=0.026) and on the Reading the Mind in the Eyes task (p=0.049). *MAPK3* had a single significant result in the domain of social cognition, on the Hinting task (p=0.031).
Table 3.1: Regression analysis of neuropsychological variable test scores in all broad psychosis cases and controls

<table>
<thead>
<tr>
<th>Gene/Variant (SZ risk allele)</th>
<th>Cognitive Domain</th>
<th>Neuropsychological Variable</th>
<th>Genotype (sample N)</th>
<th>Mean Score (SD)</th>
<th>p-value (r² change)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SDCCAG8/rs14403 (C)</strong></td>
<td>Social cognition</td>
<td>Reading the Mind in the Eyes test</td>
<td>TT (27)</td>
<td>20.48 (6.85)</td>
<td>0.015 (0.009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TC (148)</td>
<td>22.11 (5.90)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CC (318)</td>
<td>23.37 (5.89)</td>
<td></td>
</tr>
<tr>
<td><strong>SDCCAG8/rs14403 (C)</strong></td>
<td>Social cognition</td>
<td>Externalisation Bias score</td>
<td>TT (19)</td>
<td>0.37 (3.47)</td>
<td>0.001 (0.021)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TC (148)</td>
<td>0.68 (3.76)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CC (318)</td>
<td>1.76 (3.96)</td>
<td></td>
</tr>
<tr>
<td><strong>SDCCAG8/rs14403 (C)</strong></td>
<td>Working memory</td>
<td>Letter Number Sequencing</td>
<td>TT (38)</td>
<td>7 (3.92)</td>
<td>0.019 (0.005)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TC (282)</td>
<td>8.48 (3.73)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CC (575)</td>
<td>8.86 (3.84)</td>
<td></td>
</tr>
<tr>
<td><strong>MADI1L1/rs10650434 (Ins CT)</strong></td>
<td>Episodic Memory</td>
<td>Logical Memory 1</td>
<td>AA (210)</td>
<td>8.21 (4.36)</td>
<td>0.039 (0.004)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>A_ACT (465)</td>
<td>7.92 (4.27)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ACT_ACT (260)</td>
<td>7.35 (4.02)</td>
<td></td>
</tr>
<tr>
<td><strong>MADI1L1/rs10650434 (Ins CT)</strong></td>
<td>Episodic Memory</td>
<td>Logical Memory 2</td>
<td>AA (207)</td>
<td>8.96 (4.19)</td>
<td>0.007 (0.007)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>A_ACT (462)</td>
<td>8.49 (4.05)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ACT_ACT (260)</td>
<td>7.94 (3.90)</td>
<td></td>
</tr>
<tr>
<td><strong>MADI1L1/rs10650434 (Ins CT)</strong></td>
<td>Episodic Memory</td>
<td>Faces Task 1</td>
<td>AA (141)</td>
<td>9.77 (3.06)</td>
<td>0.012 (0.009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>A_ACT (318)</td>
<td>9.46 (3.12)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ACT_ACT (176)</td>
<td>8.79 (2.57)</td>
<td></td>
</tr>
<tr>
<td><strong>MADI1L1/rs10650434 (Ins CT)</strong></td>
<td>Episodic Memory</td>
<td>Faces Task 2</td>
<td>AA (140)</td>
<td>10.24 (2.96)</td>
<td>0.032 (0.006)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>A_ACT (315)</td>
<td>9.90 (3.16)</td>
<td></td>
</tr>
<tr>
<td>Genes</td>
<td>Cognition</td>
<td>Test</td>
<td>Genotype</td>
<td>Mean</td>
<td>Standard Deviation</td>
</tr>
<tr>
<td>-------</td>
<td>-----------</td>
<td>------</td>
<td>----------</td>
<td>------</td>
<td>--------------------</td>
</tr>
<tr>
<td><strong>MPHOSPH9/rs2851447 (G)</strong></td>
<td>Episodic Memory</td>
<td>Logical Memory 1</td>
<td>CC (519)</td>
<td>8.01</td>
<td>(4.21)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GC (359)</td>
<td>7.77</td>
<td>(4.21)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GG (55)</td>
<td>6.45</td>
<td>(4.45)</td>
</tr>
<tr>
<td><strong>MPHOSPH9/rs2851447 (G)</strong></td>
<td>Episodic Memory</td>
<td>Logical Memory 2</td>
<td>CC (517)</td>
<td>8.64</td>
<td>(4.00)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GC (356)</td>
<td>8.28</td>
<td>(4.11)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GG (54)</td>
<td>7.57</td>
<td>(4.20)</td>
</tr>
<tr>
<td><strong>MPHOSPH9/rs2851447 (G)</strong></td>
<td>Social Cognition</td>
<td>Externalisation Bias</td>
<td>CC (284)</td>
<td>1.62</td>
<td>(3.86)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GC (180)</td>
<td>1.15</td>
<td>(3.99)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GG (24)</td>
<td>-0.5833</td>
<td>(3.60)</td>
</tr>
<tr>
<td><strong>PRKD1/rs2068012 (C)</strong></td>
<td>IQ</td>
<td>Performance IQ</td>
<td>TT (445)</td>
<td>99.49</td>
<td>(22.80)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CT (235)</td>
<td>94.94</td>
<td>(22.44)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CC (24)</td>
<td>96.79</td>
<td>(17.79)</td>
</tr>
<tr>
<td><strong>PRKD1/rs2068012 (C)</strong></td>
<td>Social Cognition</td>
<td>Reading the Mind in the Eyes</td>
<td>TT (331)</td>
<td>23.13</td>
<td>(5.91)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CT (162)</td>
<td>22.27</td>
<td>(5.86)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CC (18)</td>
<td>21.83</td>
<td>(7.89)</td>
</tr>
<tr>
<td><strong>MAPK3/rs12691307 (A)</strong></td>
<td>Social Cognition</td>
<td>Hinting Task</td>
<td>GG (158)</td>
<td>16.23</td>
<td>(2.66)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AG (242)</td>
<td>15.76</td>
<td>(3.06)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AA (99)</td>
<td>15.40</td>
<td>(3.60)</td>
</tr>
</tbody>
</table>
3.5 Replication

I sought independent replication of the associations between both rs10650434 (*MAD1L1*) and rs2851447 (*MPHOSPH9*) and episodic memory. The UK dataset I used for replication is based on the MATRICS battery and so composite scores were used for cognitive domains. Neither SNP showed an association with episodic memory in this dataset (Table 3.2). Unfortunately, I had no access to a replication dataset that had collected information on social cognition, which is the domain under which the most significant finding between rs14403 (*SDCCGA8*) and externalising bias was achieved.

| Table 3.2: UK Replication Data for Episodic Memory |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Gene            | Variant (risk allele) | Cognitive domain | Neuropsychological Variable | Beta (Standard Error) | P-value |
| MAD1L1          | rs10650434 (Ins CT) | Episodic Memory  | Z score HVLT Total          | 0.024 (0.084)          | 0.517 |
| MPHOSPH9        | rs2851447 (G)      | Episodic Memory  | Z score HVLT Total          | 0.012 (0.153)          | 0.742 |

3.6 Analysis of centrosomal genes in neurodevelopmental disorders and cognition

Work up until now was completed from 2015 till 2016. I revisited the link between the centrosome and neurodevelopmental disorders, as well as general cognition in 2018 when larger GWAS studies had become available and newer bioinformatic tools, such as MAGMA and denovolyzeR, had been developed to aid in the analysis of human genetic studies.
3.6.1 Analysis of centrosomal genes using common variant data

I used MAGMA (de Leeuw et al., 2015) for gene-set analysis (GSA) to simultaneously study multiple genetic markers in order to determine their joint effect and test if the genes in the centrosomal list were more strongly associated with SZ or EA or IQ than other genes in the genome. MAGMA uses summary statistics (SNP P values) from GWAS and a significant enrichment within the centrosomal list points to variation across those genes influencing SZ and/or EA, and/or IQ. This would provide further evidence that biological functions related to the centrosome are relevant to the phenotypes tested. My initial analysis used the largest SZ GWAS at the time (Schizophrenia Working Group of the Psychiatric Genomics, 2014) and found that the centrosomal gene-set was enriched for SZ risk genes (P=0.014). Subsequently, larger datasets became available and GSA on the more recent SZ GWAS (Pardiñas et al., 2018) shows that there is no significant enrichment for the centrosomal gene set. However the centrosomal gene set reveals enrichment for EA (Lee et al., 2018, P=0.044) and IQ genes (Savage et al., 2018, P=0.006), as shown in Figure 3.8.

Brain-expressed genes are a major contributor to SZ (Schizophrenia Working Group of the Psychiatric Genomics, 2014) and EA (Okbay et al., 2016). It is possible that the enrichment detected here could be due to the centrosome gene-set representing a set of ‘brain-expressed’ genes. However, the centrosome list enrichment was robust to the inclusion in the analyses of ‘brain-elevated’ (N=1,424) gene-sets as a covariate (EA: P=0.042 and IQ: P=0.0048). The associations with IQ were robust to the inclusion in the analyses of a ‘brain-expressed’ (N=14,243) gene-set as a covariate (IQ: P=0.019).

To examine if the enrichment I detect for IQ is a property of polygenic phenotypes in general, I obtained GWAS summary statistics for 6 phenotypes and I tested the centrosome gene-set for enrichment in each one. These were other brain-related disorders (AD and STR) and non-brain related diseases (CAD, CD, UC and T2D). No enrichment was detected for any of the 6 phenotypes (Figure 3.8 and Supplementary Table 3.2). These data provide some evidence that variation in centrosome genes contribute to cognitive ability.
When I investigated whether SNPs within my prioritised gene list were contributing to variance in EA and IQ, I found that several SNPs within the \textit{MAD1L1} gene are significantly influencing EA (rs13240401, rs11772232, rs56085180, rs1637770 and rs10264573). However, none of these SNPs are in LD with the original SZ risk SNP that was used for the neuropsychological analysis in this study. The SZ risk SNP within \textit{PRKD1} was also found to be influencing EA, the risk allele is associated with an increase in EA. Two SNPs within \textit{SDCCAG8} were also associated with EA (rs3897821 and rs2994326). The most significantly associated SNP (rs3897821, \( P = 3.58\times10^{-25} \)) is in moderate LD with one of the original SZ risk SNPs within the \textit{SDCCAG8} locus (rs10803138, \( r^2 = 0.33 \)). On The Genotype-Tissue Expression (GTEx) portal (https://gtexportal.org/home/), an ongoing effort to build a comprehensive public resource to study tissue-specific gene expression and regulation, the SZ risk SNP is a
significant eQTL that influences the expression of SDCCAG8 in three of the thirteen brain tissues tested and also in whole blood (Table 3.3). The allele that is associated with SZ risk is associated with lower expression of SDCCAG8. This suggests that one potential mechanism by which the SZ risk allele exerts its effect is by reducing expression of SDCCAG8.

A single SNP within PRKD1 (rs971681) was associated with IQ but this SNP was not in LD with the SZ risk SNP analysed in this study. This suggests that the functional mechanism by which genetic variation in PRKD1 contributes to these phenotypes may be completely independent.

Table 3.3: SZ risk allele at rs10803138 decreases SDCCAG8 expression

<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>SNP</th>
<th>Reference Allele</th>
<th>P-Value</th>
<th>NES</th>
<th>T-statistic</th>
<th>Tissue</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDCCAG8</td>
<td>rs10803138</td>
<td>A</td>
<td>0.083</td>
<td>-0.19</td>
<td>-1.8</td>
<td>Amygdala</td>
<td>88</td>
</tr>
<tr>
<td>SDCCAG8</td>
<td>rs10803138</td>
<td>A</td>
<td>0.001*</td>
<td>-0.27</td>
<td>-3.4</td>
<td>Anterior cingulate cortex (BA24)</td>
<td>109</td>
</tr>
<tr>
<td>SDCCAG8</td>
<td>rs10803138</td>
<td>A</td>
<td>0.0014*</td>
<td>-0.23</td>
<td>-3.3</td>
<td>Caudate (basal ganglia)</td>
<td>144</td>
</tr>
<tr>
<td>SDCCAG8</td>
<td>rs10803138</td>
<td>A</td>
<td>0.00033*</td>
<td>-0.25</td>
<td>-3.7</td>
<td>Cerebellar Hemisphere</td>
<td>125</td>
</tr>
<tr>
<td>SDCCAG8</td>
<td>rs10803138</td>
<td>A</td>
<td>0.14</td>
<td>-0.12</td>
<td>-1.5</td>
<td>Cerebellum</td>
<td>154</td>
</tr>
<tr>
<td>SDCCAG8</td>
<td>rs10803138</td>
<td>A</td>
<td>0.026</td>
<td>-0.21</td>
<td>-2.3</td>
<td>Cortex</td>
<td>136</td>
</tr>
<tr>
<td>SDCCAG8</td>
<td>rs10803138</td>
<td>A</td>
<td>0.8</td>
<td>-0.022</td>
<td>-0.26</td>
<td>Frontal Cortex (BA9)</td>
<td>118</td>
</tr>
<tr>
<td>SDCCAG8</td>
<td>rs10803138</td>
<td>A</td>
<td>0.051</td>
<td>-0.18</td>
<td>-2</td>
<td>Hippocampus</td>
<td>111</td>
</tr>
<tr>
<td>SDCCAG8</td>
<td>rs10803138</td>
<td>A</td>
<td>0.01</td>
<td>-0.18</td>
<td>-2.6</td>
<td>Hypothalamus</td>
<td>108</td>
</tr>
<tr>
<td>SDCCAG8</td>
<td>rs10803138</td>
<td>A</td>
<td>0.21</td>
<td>-0.093</td>
<td>-1.3</td>
<td>Nucleus accumbens (basal ganglia)</td>
<td>130</td>
</tr>
<tr>
<td>SDCCAG8</td>
<td>rs10803138</td>
<td>A</td>
<td>0.13</td>
<td>-0.14</td>
<td>-1.5</td>
<td>Putamen (basal ganglia)</td>
<td>111</td>
</tr>
<tr>
<td>SDCCAG8</td>
<td>rs10803138</td>
<td>A</td>
<td>0.24</td>
<td>-0.12</td>
<td>-1.2</td>
<td>Spinal cord (cervical c-1)</td>
<td>83</td>
</tr>
<tr>
<td>SDCCAG8</td>
<td>rs10803138</td>
<td>A</td>
<td>0.66</td>
<td>-0.049</td>
<td>-0.45</td>
<td>Substantia nigra</td>
<td>80</td>
</tr>
<tr>
<td>SDCCAG8</td>
<td>rs10803138</td>
<td>A</td>
<td>1.70E-07*</td>
<td>-0.24</td>
<td>-5.4</td>
<td>Whole Blood</td>
<td>367</td>
</tr>
</tbody>
</table>

Notes: Outputs from eQTL calculator on GTEx (https://gtexportal.org/home/testyourown), effect of the alternative allele on gene expression relative to the reference allele is measured. The alternative allele (G) is also the SZ risk allele and is associated with reduced expression of SDCCAG8. NES; Normalized effect size. The T-statistic measures the size of the difference relative to the variation in the sample data. N; the number of individuals for which gene expression and genotype information was available. * Survives Bonferroni correction.
3.6.2 Analysis of centrosomal genes using rare variant data

I used denovolyzeR to study genes harbouring DNMs identified in patients with SZ, ASD, intellectual disability (ID) and in unaffected siblings and controls (Genovese et al., 2016, Ware et al, 2015). DNMs are categorized as protein altering a combination of loss of function and missense variants or silent. The centrosome list was enriched for genes containing de novo mutations reported in ASD, (ASD_LoF, P=0.0179, ASD_prot, P= 0.00696 and ASD_mis, P=0.0388), ID (ID_prot, P=0.00774 and ID_mis, P=0.0267) and SCZ (SCZ_LoF, P=0.0191). These enrichments survive Bonferroni correction. Importantly, the centrosomal list was not enriched for genes harbouring de novo variants that were reported in the unaffected siblings or control data (Table 3.4).

Table 3.4: Enrichment analysis of the centrosomal list for genes harbouring de novo variants in SZ, ASD and ID

<table>
<thead>
<tr>
<th>Phnotype_Mutation</th>
<th>Observed Variants (N)</th>
<th>Expected Variants (N)</th>
<th>P-value</th>
<th>Enrichment Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCZ_prot</td>
<td>49</td>
<td>38.7</td>
<td>0.0612</td>
<td>1.27</td>
</tr>
<tr>
<td>SCZ_silent</td>
<td>9</td>
<td>14.1</td>
<td>0.941</td>
<td>0.638</td>
</tr>
<tr>
<td>ASD_prot</td>
<td>182</td>
<td>150.5</td>
<td>0.00696*</td>
<td>1.21</td>
</tr>
<tr>
<td>ASD_silent</td>
<td>58</td>
<td>54.9</td>
<td>0.353</td>
<td>1.06</td>
</tr>
<tr>
<td>ID_prot</td>
<td>15</td>
<td>7.3</td>
<td>0.00774*</td>
<td>2.07</td>
</tr>
<tr>
<td>ID_silent</td>
<td>1</td>
<td>2.6</td>
<td>0.929</td>
<td>0.378</td>
</tr>
<tr>
<td>unaff/con_prot</td>
<td>66</td>
<td>77.4</td>
<td>0.914</td>
<td>0.853</td>
</tr>
<tr>
<td>unaff/con_silent</td>
<td>26</td>
<td>28.2</td>
<td>0.686</td>
<td>0.922</td>
</tr>
</tbody>
</table>

*Survives Bonferroni correction

3.7 Cloning and overexpression of prioritised genes

To confirm centrosomal localisation of prioritised genes, I cloned MAD1L1, SDCCAG8, GIGYF2, MAPK3 and PRKD1 cDNA into cloning and mammalian expression vectors. MPHOSPH9 cloning was also attempted but it proved unsuccessful. First, I designed primers to the 5’ and 3’ ends of the prioritised genes, as displayed in Table 2.11. RT-PCR was successfully performed for five of the six prioritised genes, as confirmed by agarose gel electrophoresis. The amplicons were cloned into pGEM-T-Easy and then subcloned to a pEGFP-N1 backbone in the case of SDCCAG8, PRKD1, MAD1L1 and MAPK3 using EcoRI and BamHI to yield pEGFP-SDCCAG8, pEGFP-MAPK3, pEGFP-PRKD1, and
pEGFP-MAD1L1 plasmids, respectively. In the case of *GIGYF2* XhoI and BamHI were used, yielding pEGFP-GIGYF2. Figure 3.9 A is a representative plasmid map, in this case depicting pEGFP-SDCCAG8. Restriction enzyme digests of the plasmids generated were performed with EcoRI and BamHI, with the exception of pEGFP-GIGYF2 which was digested with XhoI and BamHI. Digestion of the plasmids released the expected ~2.1 kb, 2.1 kb, 1.1 kb, 3.9 kb and 2.7 kb inserts from the 4.7kb backbone (Figure 3.9 B). The plasmids were also sequenced to verify the cDNA sequence.

**Figure 3.9** Cloning full length cDNA of prioritised genes into pEGFP-N1
A. Schematic of pEGFP-N1-SDCCAG8 plasmid showing the CMV promoter, kanamycin resistance marker used for bacterial selection and the restriction sites used in cloning. B. Restriction digest revealed the following confirmatory band sizes along with the 4.7 kb backbone: 1. pEGFP-MAD1L1, 2.1 kb; 2. pEGFP-SDCCAG8, 2.1 kb; 3. pEGFP-MAPK3, 1.3 kb; 4. pEGFP-GIGYF2, 3.9 kb; 5. pEGFP-PRKD1, 2.7 kb.
To test the expression constructs, I transiently transfected them into wild-type hTERT-RPE1 cells (Figure 3.10). Using Centrin to mark the centrosome during immunofluorescence microscopy, I confirmed that the tagged SDCCAG8, MAD1L1, PRKD1, GIGYF2 and PRKD1 protein expressed from the pEGFP plasmids localised to the centrosome (Figure 3.10). MAD1L1, PRKD1 and GIGYF2 co-localise completely given that Centrin is a distal lumen marker suggesting that these proteins also localise to the distal lumen. SDCCAG8 and MAPK3 do not completely localise with Centrin but are in very close proximity suggesting a localisation to an alternative region of the centrosome, or the pericentriolar material.

Figure 3.10: Transient expression of full-length human MAD1L1, MAPK3, PRKD1, SDCCAG8 and GIGYF2 cDNA
Immunofluorescence microscopy analysis of transient expression of pEGFP-MAD1L1, pEGFP-MAPK3, pEGFP-PRKD1, pEGFP-SDCCAG8 and pEGFP-GIGYF2. Proteins expressed from all plasmids showed centrosomal localisation with centrin (red). Scale bars, 5 µm and inset scale bars, 1 µm.
3.8 Discussion

I developed a list of 806 genes that have been reported to localise to the centrosome. These genes were investigated for a role in SZ and cognition by first determining which of these genes are present at loci that are genome-wide significant for association with SZ based on the largest GWAS that was available at that time. The six candidate SZ risk genes that were prioritised for further analysis have previously been associated with neurological conditions, for example, MS (*MPHOSPH9*), Atypical Rett Syndrome (*PRKDL1*), BD (*MAD1L1*) and ASD (*GIGYF2* and *MAPK3*). Five (*SDCCAG8*, *MAD1L1*, *PRKDL1*, *MPHOSPH9* and *MAPK3*) of the six prioritised genes contained SNPs that were found to be associated with performance in tests from four of five cognitive domains; IQ, episodic memory, working memory and social cognition. For five of the six candidate genes, the SZ index SNP is located within the gene. It is difficult to discern the impact of these SNPs on the gene. For some of the SNPs, brain eQTL data identifies potentially functional impacts of SNPs on gene expression. Significant eQTLs for rs10650434 and *MAD1L1* (using proxy SNP rs12668848), rs2851447 and *MPHOSPH9* (using proxy SNP rs2851447) and rs12691307 and *MAPK3* are reported in the Kim et al. (Kim et al., 2014) brain eQTL data that included 424 brain samples. There was also a significant eQTL for rs10803138 and *SDCCAG8* reported on GTEx portal in whole blood (*P*=1.8*10^{-11}, *N*=338 samples) and also in Brain Cortex (*P*=0.0055, *N*=96 samples).

The most significant association in the neuropsychological analysis was between rs14403 (*SDCCAG8*) and EB (externalising bias), which is a measure of social cognition. The SZ risk allele C is associated with an increased likelihood of attributing negative events to external factors relative to positive events (*p*=0.001) and 2.1% of the variance in EB is explained by rs14403 genotype. My analyses combined psychosis cases and healthy participants together to maximize sample size and power. When, in *post hoc* analyses, these samples were considered separately, the association between the risk allele at rs14403 was not significant in the healthy sample. However, it remained significant in the patient only sample (*P*=0.006).

My attempts to replicate significant results were conservatively restricted to following up the associations between rs2851447 (*MPHOSPH9*) and rs10650434 (*MAD1L1*) and
Chapter 3

episodic memory. I did not uncover independent data to support my associations. There are a number of caveats to consider that could explain the non-replication. For the UK sample, comparable measures of memory were not available. The modest though significant amount of variance associated with these variants, which was estimated as being around 0.9 and 0.5% respectively, may also suggest that the samples included here, though large for these types of studies, lacked sufficient power to detect association.

By applying GSA to GWAS data using the centrosome list, I identified a specific contribution of centrosomal proteins to cognitive function. The initial analysis tested for enrichment in the PGC 2 SZ GWAS where there was an enrichment for centrosomal proteins which was robust to the inclusion of ‘brain-elevated’ and ‘brain expressed’ gene-lists as covariates. Unfortunately, this association was not repeated in a more recent and larger GWAS of SZ (Pardiñas et al., 2018). The centrosome list was associated with EA, although this association was not robust to the inclusion of ‘brain-elevated’ genes as a covariate. EA is generally accepted as a proxy for cognitive ability because of the large samples that can be gathered with greater ease (Deary et al., 2007). I tested the association of centrosomal proteins with IQ, which was the most significant finding and was robust to the inclusion of ‘brain-elevated’ and ‘brain expressed’ gene-lists as covariates. This suggests that centrosomal proteins contribute to cognitive function. The crucial role of the centrosome in brain development has been highlighted previously. Disorders including lissencephaly, microcephaly and disorders of growth where the brain is disproportionately affected, like microcephalic osteodysplastic primordial dwarfism type II, have all been linked to disruption of centrosomal proteins (Bettencourt-Dias et al., 2011). These results suggest that common variants in centrosomal genes present in the general population are influencing cognitive function and to my knowledge, this is the first study to demonstrate this. As new large GWAS for neurocognitive phenotypes and for SZ, are produced, it will be important to determine if genetic variation within biological processes regulated by the centrosome influence specific traits or instead exert an influence across multiple behavioural and neuropsychiatric phenotypes.

I also used denovolyzeR to analyse data from trios studies (Genovese et al., 2016) and to ascertain whether the centrosomal list was enriched for rare de novo mutations in SZ,
ASD and ID. There is an enrichment of protein altering variants in ASD and ID that survives multiple-test correction suggesting centrosomal proteins contribute to cognitive function and neurodevelopmental disorders. Importantly, there was no enrichment for DNMs in centrosomal genes of any mutation class in unaffected siblings or controls, which suggests mutations in centrosomal proteins have a specific contribution to the phenotypes tested here.

The localisation analysis of prioritised candidate genes was performed to ensure that the proteins were indeed localising to the centrosome, as the evidence available in some cases was limited. GIGYF2 was computationally predicted to localise to the centrosome based on orthologues, but has never been experientially validated until now where I see co-localisation with centrin in hTERT-RPE1 cells. In the case of PRKD1 (Papazyan et al., 2008) and MAPK3 (Willard and Crouch, 2001) previous localisation analysis was focused on cells undergoing mitosis and no data on centrosomal localisation was presented for interphase cells. MAD1L1 centrosomal localisation was first observed during metaphase (Dong-Yan et al., 1999) and subsequently during interphase (Wan et al., 2014). In the case of MPHOSPH9, localisation was observed during interphase (Jakobsen et al., 2011), which was also the case for SDCCAG8 (Otto et al., 2010). This analysis supports the localisation of all candidate proteins to the centrosome, as previously reported, with the exception of MPHOSPH9, which I was unable to clone into a GFP plasmid. The localisation analysis supports centrosomal functions for the proteins analysed. However, there are caveats to using GFP tagging. The transient transfection of these proteins into cells results in an overexpression of the protein which can lead to protein aggregation. In addition, the presence of the GFP tag can reduce the stability of the protein or cause mis-targeting of the protein (Weill et al., 2019).

In summary, I have identified that some single gene associations detected in SZ GWAS are genes with centrosomal functions. I have identified associations between SZ variants within these genes and cognition in the Irish dataset. I have confirmed that five of the six candidate genes localise to the centrosome. I have also built on these single gene associations, to show that centrosomal genes in general are enriched for common variants associated with EA and IQ, and also for rare variants that increase the risk of ID and
ASD. These findings are supported by existing literature which suggests that centrosomal function is important for brain development. Thus, this study provides evidence that the molecular mechanisms that underpin SZ and cognitive function include perturbations of the biological processes influenced by the centrosome in the brain.

3.9 Supplementary data

**Supplementary Table 3.1: Centrosomal gene list**

<table>
<thead>
<tr>
<th>Gene</th>
<th>Gene</th>
<th>Gene</th>
<th>Gene</th>
<th>Gene</th>
<th>Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAAS</td>
<td>CEP350</td>
<td>FSD1</td>
<td>MAP9</td>
<td>PLCG1</td>
<td>SNCG</td>
</tr>
<tr>
<td>AATF</td>
<td>CEP41</td>
<td>FTCD</td>
<td>MAPK1</td>
<td>PLD2</td>
<td>SOCS1</td>
</tr>
<tr>
<td>ACTR1A</td>
<td>CEP44</td>
<td>FYN</td>
<td>MAPK10</td>
<td>PLEKHA7</td>
<td>SORBS1</td>
</tr>
<tr>
<td>ACTR1B</td>
<td>CEP55</td>
<td>FZR1</td>
<td>MAPK14</td>
<td>PLEKHG6</td>
<td>SP1</td>
</tr>
<tr>
<td>ACTR2</td>
<td>CEP57</td>
<td>G6PD</td>
<td>MAPK3</td>
<td>PK1</td>
<td>SPAG5</td>
</tr>
<tr>
<td>AHI1</td>
<td>CEP57L1</td>
<td>GADD45A</td>
<td>MAPK7</td>
<td>PLK2</td>
<td>SPAST</td>
</tr>
<tr>
<td>AJUBA</td>
<td>CEP63</td>
<td>GDJ2</td>
<td>MAPK8</td>
<td>PLK3</td>
<td>SPATC1</td>
</tr>
<tr>
<td>AK5</td>
<td>CEP68</td>
<td>GEN1</td>
<td>MARCKS</td>
<td>POC1A</td>
<td>SPCE1</td>
</tr>
<tr>
<td>AKAP11</td>
<td>CEP70</td>
<td>GIGYF2</td>
<td>MAPRE1</td>
<td>PML</td>
<td>SPC25</td>
</tr>
<tr>
<td>AKAP9</td>
<td>CEP72</td>
<td>GIMAP5</td>
<td>MPT</td>
<td>PNP</td>
<td>SPD1</td>
</tr>
<tr>
<td>AKT1</td>
<td>CEP76</td>
<td>GIT1</td>
<td>POC1B</td>
<td>PP1A</td>
<td>SPG20</td>
</tr>
<tr>
<td>ALMS1</td>
<td>CEP78</td>
<td>GMNN</td>
<td>MARK4</td>
<td>POC5</td>
<td>SPOK1</td>
</tr>
<tr>
<td>ALS2</td>
<td>CEP83</td>
<td>GNAI1</td>
<td>MASTL</td>
<td>POC5</td>
<td>SPH2</td>
</tr>
<tr>
<td>ANAPC1</td>
<td>CEP85</td>
<td>GNAI2</td>
<td>MBD3</td>
<td>POLR3H</td>
<td>SPH2</td>
</tr>
<tr>
<td>ANAPC10</td>
<td>CEP85L</td>
<td>GNAI3</td>
<td>MCM2</td>
<td>PPARA</td>
<td>SNA1</td>
</tr>
<tr>
<td>ANKRD26</td>
<td>CEP89</td>
<td>GNAI1</td>
<td>MCM3</td>
<td>PP1A2</td>
<td>SSX2P</td>
</tr>
<tr>
<td>ANKRD32</td>
<td>CEP95</td>
<td>GNB2L1</td>
<td>MCM4</td>
<td>PP1CA</td>
<td>ST7L</td>
</tr>
<tr>
<td>ANXA11</td>
<td>CEP97</td>
<td>GOLGA2</td>
<td>MCM5</td>
<td>PP1CC</td>
<td>STAT3</td>
</tr>
<tr>
<td>APC</td>
<td>CETN1</td>
<td>GOLGA4</td>
<td>MCM6</td>
<td>PP1R1</td>
<td>STAT5A</td>
</tr>
<tr>
<td>APEXI</td>
<td>CETN2</td>
<td>GPS1</td>
<td>MCM7</td>
<td>PP1R12A</td>
<td>STIL</td>
</tr>
<tr>
<td>APOA1BP</td>
<td>CETN3</td>
<td>GPSM2</td>
<td>MCPH1</td>
<td>PP1R2</td>
<td>STK26</td>
</tr>
<tr>
<td>APP</td>
<td>CHEK1</td>
<td>GRB2</td>
<td>MCRS1</td>
<td>PP1R42</td>
<td>STK3</td>
</tr>
<tr>
<td>ARHGEF10</td>
<td>CHEK2</td>
<td>GRK5</td>
<td>MDH1</td>
<td>PP2R1A</td>
<td>STK38</td>
</tr>
<tr>
<td>ARHGEF12</td>
<td>CHMP3</td>
<td>GSG2</td>
<td>MDM2</td>
<td>PP2R5A</td>
<td>STK38L</td>
</tr>
<tr>
<td>ARHGEF2</td>
<td>CHUK</td>
<td>GSK3A</td>
<td>MEAF6</td>
<td>PP2R5C</td>
<td>STX1B</td>
</tr>
<tr>
<td>ARL2</td>
<td>CIB1</td>
<td>GSK3B</td>
<td>MELK</td>
<td>PP4C</td>
<td>STX8</td>
</tr>
<tr>
<td>ARL2BP</td>
<td>CKAP2</td>
<td>H2AFX</td>
<td>MIB1</td>
<td>PPP4R2</td>
<td>SUMO1</td>
</tr>
<tr>
<td>ARL3</td>
<td>CKAP2L</td>
<td>H2AFY</td>
<td>MID2</td>
<td>PR1C</td>
<td>SUN2</td>
</tr>
<tr>
<td>ARPC1B</td>
<td>CKAP5</td>
<td>H3F3B/H3F3A</td>
<td>MKKS</td>
<td>PRKAA1</td>
<td>SUN3</td>
</tr>
<tr>
<td>ARRB2</td>
<td>CLASP1</td>
<td>HAP1</td>
<td>MKN1</td>
<td>PRKAC1</td>
<td>SYK</td>
</tr>
<tr>
<td>ASPM</td>
<td>CLASP2</td>
<td>HARBI1</td>
<td>MKS1</td>
<td>PRKACB</td>
<td>SYTL4</td>
</tr>
<tr>
<td>ATM</td>
<td>CLEC11A</td>
<td>HAUS1</td>
<td>MMP14</td>
<td>PRKAR2A</td>
<td>TACC1</td>
</tr>
<tr>
<td>AUNIP</td>
<td>CLJC4</td>
<td>HAUS2</td>
<td>MOB1A</td>
<td>PRKAR2B</td>
<td>TACC2</td>
</tr>
<tr>
<td>AURKA</td>
<td>CLIP1</td>
<td>HAUS3</td>
<td>MOB1B</td>
<td>PRKCB</td>
<td>TACC3</td>
</tr>
<tr>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>AURKB</td>
<td>CNTLN</td>
<td>HAUS4</td>
<td>MOB4</td>
<td>PRKCD</td>
<td>TBCD</td>
</tr>
<tr>
<td>AURKC</td>
<td>CNTRL</td>
<td>HAUS5</td>
<td>MOS</td>
<td>PRKCE</td>
<td>TBC</td>
</tr>
<tr>
<td>AXIN2</td>
<td>CNTROB</td>
<td>HAUS6</td>
<td>MPHOSPH</td>
<td>PRKC1</td>
<td>TBCD1</td>
</tr>
<tr>
<td>AZIN1</td>
<td>COMT</td>
<td>HAU7</td>
<td>MTK1</td>
<td>PRKQ</td>
<td>TBCD</td>
</tr>
<tr>
<td>B9D1</td>
<td>CPS2</td>
<td>HAUS8</td>
<td>MTK2</td>
<td>PRKZ</td>
<td>TEA2</td>
</tr>
<tr>
<td>B9D2</td>
<td>CPS3</td>
<td>HBEGF</td>
<td>MTUS1</td>
<td>PRK1</td>
<td>TCHH</td>
</tr>
<tr>
<td>BAALC</td>
<td>CPS4</td>
<td>HADAC1</td>
<td>MTUS2</td>
<td>PRK2</td>
<td>TCHP</td>
</tr>
<tr>
<td>BARD1</td>
<td>CPS5</td>
<td>HADAC8</td>
<td>MYCBP</td>
<td>PRK3</td>
<td>TCP1</td>
</tr>
<tr>
<td>BBC3</td>
<td>CPS6</td>
<td>HERC2</td>
<td>MYO10</td>
<td>PRKDC</td>
<td>TEKT1</td>
</tr>
<tr>
<td>BBS1</td>
<td>CPS7A</td>
<td>HIPK2</td>
<td>MYO5A</td>
<td>PROCR</td>
<td>TEKT2</td>
</tr>
<tr>
<td>BBS4</td>
<td>CPS7B</td>
<td>HMMR</td>
<td>MZT1</td>
<td>PSEN1</td>
<td>TEF1</td>
</tr>
<tr>
<td>BBS5</td>
<td>CPS8</td>
<td>HOOK2</td>
<td>MZT2A</td>
<td>PSEN2</td>
<td>TFAM</td>
</tr>
<tr>
<td>BBS7</td>
<td>CPEB1</td>
<td>HOOK3</td>
<td>MZT2B</td>
<td>PSKH1</td>
<td>TFAP2A</td>
</tr>
<tr>
<td>BCL2L1</td>
<td>CRABP1</td>
<td>HSD1B14</td>
<td>NAA11</td>
<td>PSMA1</td>
<td>TIPO</td>
</tr>
<tr>
<td>BGN</td>
<td>CRB3</td>
<td>HSF1</td>
<td>NAF1</td>
<td>SMC5</td>
<td>TIAM1</td>
</tr>
<tr>
<td>BIRC5</td>
<td>CRMP1</td>
<td>HSP90AA1</td>
<td>NBN</td>
<td>PSMD1</td>
<td>TLN1</td>
</tr>
<tr>
<td>BLOC1S2</td>
<td>CRNLK1</td>
<td>HSPA1A</td>
<td>NCPAG</td>
<td>PSMD2</td>
<td>TMEM67</td>
</tr>
<tr>
<td>BRAF</td>
<td>CROCC</td>
<td>HSPA2</td>
<td>NDC80</td>
<td>PSMD3</td>
<td>TMUB1</td>
</tr>
<tr>
<td>BRCA1</td>
<td>CRYAB</td>
<td>HSPA8</td>
<td>NDE1</td>
<td>PSMD4</td>
<td>TNAIP3</td>
</tr>
<tr>
<td>BRCA2</td>
<td>CSK</td>
<td>HSPA9</td>
<td>NDEL1</td>
<td>PSME1</td>
<td>TKNS</td>
</tr>
<tr>
<td>BRINP3</td>
<td>CSNK1A1</td>
<td>HSPAB</td>
<td>NDN</td>
<td>PSME2</td>
<td>TOP2A</td>
</tr>
<tr>
<td>BRSK1</td>
<td>CSNK1D</td>
<td>HTT</td>
<td>NDRG1</td>
<td>PTEN</td>
<td>TOP2B</td>
</tr>
<tr>
<td>BRSK2</td>
<td>CSNK1E</td>
<td>HUNK</td>
<td>NDRG2</td>
<td>PT2K</td>
<td>TOPBP1</td>
</tr>
<tr>
<td>C10ORF90</td>
<td>CSNK2A1</td>
<td>HYLS1</td>
<td>NEDD1</td>
<td>PTK2B</td>
<td>TOPORS</td>
</tr>
<tr>
<td>C2CD3</td>
<td>CSNK2A2</td>
<td>ID1</td>
<td>NEDD9</td>
<td>PTPDC1</td>
<td>TP53</td>
</tr>
<tr>
<td>CAB39L</td>
<td>CSPP1</td>
<td>IFFO2</td>
<td>NEIL1</td>
<td>PTPN13</td>
<td>TPG51</td>
</tr>
<tr>
<td>CALM1</td>
<td>CTAG2</td>
<td>IFT20</td>
<td>NEK1</td>
<td>PTG1</td>
<td>TPG2</td>
</tr>
<tr>
<td>CALM2</td>
<td>CTBP1</td>
<td>IFT74</td>
<td>NEK2</td>
<td>PXK</td>
<td>TPX2</td>
</tr>
<tr>
<td>CALM3</td>
<td>CTCF</td>
<td>IFT80</td>
<td>NEK6</td>
<td>PXN</td>
<td>TRAF5</td>
</tr>
<tr>
<td>CAMK2B</td>
<td>CTCLFL</td>
<td>IFT81</td>
<td>NEK7</td>
<td>RAB11FIP3</td>
<td>TRIM22</td>
</tr>
<tr>
<td>CAMK2G</td>
<td>CTNNB1</td>
<td>IFT88</td>
<td>NEK8</td>
<td>RAB11FIP4</td>
<td>TRIM6</td>
</tr>
<tr>
<td>CAPG</td>
<td>CTNND1</td>
<td>IGFBP4</td>
<td>NEK9</td>
<td>RAB11FIP5</td>
<td>TRIOBP</td>
</tr>
<tr>
<td>CASP2</td>
<td>CTN</td>
<td>IK</td>
<td>NEURL1</td>
<td>RABGAP1</td>
<td>TRIP10</td>
</tr>
<tr>
<td>CASP3</td>
<td>CUL1</td>
<td>IKKBK</td>
<td>NEURL4</td>
<td>RAB3IP</td>
<td>TRIP11</td>
</tr>
<tr>
<td>CASP8</td>
<td>DAPK3</td>
<td>IKBK</td>
<td>NFE2L2</td>
<td>RAB6C</td>
<td>TRMU</td>
</tr>
<tr>
<td>CBX5</td>
<td>DARS2</td>
<td>ILIRN</td>
<td>NFU1</td>
<td>RAB8A</td>
<td>TRNT1</td>
</tr>
<tr>
<td>CBY1</td>
<td>DCAF12</td>
<td>ILK</td>
<td>NHLRC1</td>
<td>RABGAP1</td>
<td>TROAP</td>
</tr>
<tr>
<td>CCDC110</td>
<td>DCTN1</td>
<td>IPPK</td>
<td>NIN</td>
<td>RABL6</td>
<td>TSI1</td>
</tr>
<tr>
<td>CCDC124</td>
<td>DCTN2</td>
<td>IQCB1</td>
<td>NINL</td>
<td>RAN</td>
<td>TSC2</td>
</tr>
<tr>
<td>CCDC14</td>
<td>DCTN3</td>
<td>IQGAP1</td>
<td>NIT2</td>
<td>RANBP1</td>
<td>TSEN2</td>
</tr>
<tr>
<td>CCDC15</td>
<td>DCTN4</td>
<td>IST1</td>
<td>NME1</td>
<td>RANBP9</td>
<td>TSG101</td>
</tr>
<tr>
<td>CCDC34</td>
<td>DCTN5</td>
<td>JTB</td>
<td>NME2</td>
<td>RAP1GAP2</td>
<td>TSNS</td>
</tr>
<tr>
<td>CCDC61</td>
<td>DCTN6</td>
<td>KATNA1</td>
<td>NME7</td>
<td>RARRES3</td>
<td>TSN</td>
</tr>
<tr>
<td>CCDC64</td>
<td>DCX</td>
<td>KATNB1</td>
<td>NPHP4</td>
<td>RASSF1</td>
<td>TSYL2</td>
</tr>
<tr>
<td>CCDC77</td>
<td>DDHD2</td>
<td>KDM1A</td>
<td>NPM1</td>
<td>RASSF5</td>
<td>TSSK2</td>
</tr>
<tr>
<td>Gene</td>
<td>Gene</td>
<td>Gene</td>
<td>Gene</td>
<td>Gene</td>
<td>Gene</td>
</tr>
<tr>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>CCDC85B</td>
<td>DDX11</td>
<td>KDM4A</td>
<td>NTRK1</td>
<td>RB1</td>
<td>TTC19</td>
</tr>
<tr>
<td>CCDC88A</td>
<td>DDX9</td>
<td>KEAP1</td>
<td>NUDT21</td>
<td>RBM39</td>
<td>TCC8</td>
</tr>
<tr>
<td>CCHCR1</td>
<td>DIAPH3</td>
<td>KIAA0101</td>
<td>NUFS2</td>
<td>REC8</td>
<td>TTK</td>
</tr>
<tr>
<td>CCNA1</td>
<td>DIS3L</td>
<td>KIAA0368</td>
<td>NUMA1</td>
<td>RFWD2</td>
<td>TTTL5</td>
</tr>
<tr>
<td>CCNA2</td>
<td>DISC1</td>
<td>KIAA0586</td>
<td>NUP160</td>
<td>RGCC</td>
<td>TUBA4A</td>
</tr>
<tr>
<td>CCNB1</td>
<td>DXDC1</td>
<td>KIAA0753</td>
<td>NUP85</td>
<td>RGS14</td>
<td>TUBB3</td>
</tr>
<tr>
<td>CCNB2</td>
<td>DLGAP5</td>
<td>KIAAI107</td>
<td>ODF2</td>
<td>RHOA</td>
<td>TUBB4A</td>
</tr>
<tr>
<td>CNE2</td>
<td>DNAJB1</td>
<td>KIF11</td>
<td>ODF2L</td>
<td>RHOC</td>
<td>TUBD1</td>
</tr>
<tr>
<td>CCNF</td>
<td>DNM2</td>
<td>KIF13A</td>
<td>OFD1</td>
<td>RIC8A</td>
<td>TUBE1</td>
</tr>
<tr>
<td>CCNG2</td>
<td>DONSON</td>
<td>KIF14</td>
<td>OGG1</td>
<td>RIC8B</td>
<td>TUBG1</td>
</tr>
<tr>
<td>CCI10</td>
<td>DPF2</td>
<td>KIF15</td>
<td>OGT</td>
<td>RILPL1</td>
<td>TUBG2</td>
</tr>
<tr>
<td>CCAP</td>
<td>DTL</td>
<td>KIF18A</td>
<td>OLA1</td>
<td>RINT1</td>
<td>TUBGC2</td>
</tr>
<tr>
<td>CCT4</td>
<td>DUSP23</td>
<td>KIF18B</td>
<td>ORC1</td>
<td>RITA1</td>
<td>TUBGC3</td>
</tr>
<tr>
<td>CCT5</td>
<td>DYNC1H1</td>
<td>KIF20B</td>
<td>ORC2</td>
<td>RMDN1</td>
<td>TUBGC4</td>
</tr>
<tr>
<td>CCT8</td>
<td>DYNC1H1</td>
<td>KIF23</td>
<td>ORC3</td>
<td>RNF19A</td>
<td>TUBGC5</td>
</tr>
<tr>
<td>CDC14A</td>
<td>DYNC1I2</td>
<td>KIF2A</td>
<td>ORC4</td>
<td>ROCK1</td>
<td>TUBGC6</td>
</tr>
<tr>
<td>CDC14B</td>
<td>DYNC1L1</td>
<td>KIF2B</td>
<td>ORC5</td>
<td>ROCK2</td>
<td>TXNDC9</td>
</tr>
<tr>
<td>CDC16</td>
<td>DYNC1L2</td>
<td>KIF3A</td>
<td>PACSIN1</td>
<td>RPGR</td>
<td>UBA52</td>
</tr>
<tr>
<td>CDC20</td>
<td>DYNLL1</td>
<td>KIF3B</td>
<td>PACSIN2</td>
<td>RPGRIP1</td>
<td>UBE2C</td>
</tr>
<tr>
<td>CDC25A</td>
<td>DYNLL2</td>
<td>KIF3P3</td>
<td>PACSIN3</td>
<td>RPGRIP1L</td>
<td>UBE3A</td>
</tr>
<tr>
<td>CDC25B</td>
<td>DYNLRB1</td>
<td>KIFC1</td>
<td>PAF1</td>
<td>RPL41</td>
<td>UBXN6</td>
</tr>
<tr>
<td>CDC25C</td>
<td>DYNLRB2</td>
<td>KIFC3</td>
<td>PAFAH1B1</td>
<td>RPP25</td>
<td>UNC119</td>
</tr>
<tr>
<td>CDC27</td>
<td>DYNLT3</td>
<td>KIZ</td>
<td>PK1</td>
<td>RPS6KB2</td>
<td>USP2</td>
</tr>
<tr>
<td>CDC34</td>
<td>DYNIC1</td>
<td>KLHL41</td>
<td>PK2</td>
<td>RUNX3</td>
<td>USP21</td>
</tr>
<tr>
<td>CDC42</td>
<td>EFHC1</td>
<td>KPNB1</td>
<td>PAN3</td>
<td>RUVBL1</td>
<td>USP32</td>
</tr>
<tr>
<td>CDC42BPG</td>
<td>EIF3A</td>
<td>KRIT1</td>
<td>PARD3</td>
<td>RUVBL2</td>
<td>USP33</td>
</tr>
<tr>
<td>CDC45</td>
<td>EIF4A2</td>
<td>LAMTOR5</td>
<td>PARD6A</td>
<td>S100B</td>
<td>USP8</td>
</tr>
<tr>
<td>CDC4A</td>
<td>EIF4EBP1</td>
<td>LATS1</td>
<td>PARD6G</td>
<td>S1PR5</td>
<td>UTP6</td>
</tr>
<tr>
<td>CDH13</td>
<td>EN02</td>
<td>LATS2</td>
<td>PARG</td>
<td>SAC3D1</td>
<td>UXT</td>
</tr>
<tr>
<td>CDK1</td>
<td>EPB41</td>
<td>LCA5</td>
<td>PARK2</td>
<td>SAR1A</td>
<td>VAC14</td>
</tr>
<tr>
<td>CDK11B</td>
<td>EPM2A</td>
<td>LCK</td>
<td>PARP1</td>
<td>SASS6</td>
<td>VCL</td>
</tr>
<tr>
<td>CDK2</td>
<td>ERC1</td>
<td>LDLRAP1</td>
<td>PARP3</td>
<td>SAV1</td>
<td>VDAC3</td>
</tr>
<tr>
<td>CDK5RAP2</td>
<td>ESPL1</td>
<td>LIG3</td>
<td>PCGF3</td>
<td>SCLT1</td>
<td>VIM</td>
</tr>
<tr>
<td>CDK5RAP3</td>
<td>ESRP2</td>
<td>LIMK1</td>
<td>PCID2</td>
<td>SCYL1</td>
<td>VPS37A</td>
</tr>
<tr>
<td>CDKL2</td>
<td>EV15</td>
<td>LIMK2</td>
<td>PCIM</td>
<td>SDCCAG8</td>
<td>WASF1</td>
</tr>
<tr>
<td>CDKN1A</td>
<td>EWSR1</td>
<td>LMNA</td>
<td>PCNT</td>
<td>SEP12</td>
<td>WDR35</td>
</tr>
<tr>
<td>CDKN3</td>
<td>FAM10A</td>
<td>LM04</td>
<td>PDCD2</td>
<td>SEP7</td>
<td>WDR62</td>
</tr>
<tr>
<td>CDR2L</td>
<td>FAM110B</td>
<td>LLRC45</td>
<td>PDCD6IP</td>
<td>SEP9</td>
<td>WEE2</td>
</tr>
<tr>
<td>CEPNP</td>
<td>FAM110C</td>
<td>LLRCC1</td>
<td>PDE4DIP</td>
<td>SERPINB6</td>
<td>WRAP73</td>
</tr>
<tr>
<td>CEPNP1</td>
<td>FAM161A</td>
<td>LUZP1</td>
<td>PDE7A</td>
<td>SERPINI2</td>
<td>WRN</td>
</tr>
<tr>
<td>CEPNU</td>
<td>FANCA</td>
<td>LITZS2</td>
<td>PEBP1</td>
<td>SF11</td>
<td>XPO1</td>
</tr>
<tr>
<td>CEP104</td>
<td>FANCE</td>
<td>MAD1L1</td>
<td>PGD</td>
<td>SGOL1</td>
<td>XRCC1</td>
</tr>
<tr>
<td>CEP112</td>
<td>FBX7L</td>
<td>MAD2L1</td>
<td>PIBF1</td>
<td>SK2</td>
<td>XRCC5</td>
</tr>
<tr>
<td>CEP120</td>
<td>FBXO40</td>
<td>MAEA</td>
<td>PIK3CA</td>
<td>SIRT2</td>
<td>XRCC6</td>
</tr>
<tr>
<td>CEP131</td>
<td>FBXW11</td>
<td>MAK</td>
<td>PIK3R1</td>
<td>SKA1</td>
<td>YIF1A</td>
</tr>
<tr>
<td>CEP135</td>
<td>FBXW5</td>
<td>MAP1A</td>
<td>PIK3R2</td>
<td>SKP1</td>
<td>YPEL1</td>
</tr>
</tbody>
</table>
Supplementary Table 3.2: Gene-set analysis of centrosome genes in IQ, EA and SZ and post-hoc analysis in control phenotypes

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>No. of genes</th>
<th>BETA</th>
<th>SE</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>T2D</td>
<td>784</td>
<td>-0.00161</td>
<td>0.0289</td>
<td>0.52218</td>
</tr>
<tr>
<td>UC</td>
<td>784</td>
<td>0.0327</td>
<td>0.0338</td>
<td>0.16647</td>
</tr>
<tr>
<td>CD</td>
<td>784</td>
<td>0.0278</td>
<td>0.0344</td>
<td>0.20972</td>
</tr>
<tr>
<td>CAD</td>
<td>781</td>
<td>-0.0298</td>
<td>0.0314</td>
<td>0.82877</td>
</tr>
<tr>
<td>STR</td>
<td>781</td>
<td>-0.011</td>
<td>0.0302</td>
<td>0.64202</td>
</tr>
<tr>
<td>AD</td>
<td>783</td>
<td>-0.0179</td>
<td>0.0332</td>
<td>0.70547</td>
</tr>
<tr>
<td>OCD</td>
<td>784</td>
<td>0.028</td>
<td>0.0297</td>
<td>0.17254</td>
</tr>
<tr>
<td>MDD</td>
<td>785</td>
<td>-0.00437</td>
<td>0.0301</td>
<td>0.55773</td>
</tr>
<tr>
<td>EA</td>
<td>784</td>
<td>0.0821</td>
<td>0.0485</td>
<td>0.045305</td>
</tr>
<tr>
<td>IQ</td>
<td>803</td>
<td>0.101</td>
<td>0.0409</td>
<td>0.006813</td>
</tr>
<tr>
<td>SZ</td>
<td>790</td>
<td>0.0514</td>
<td>0.0414</td>
<td>0.10682</td>
</tr>
</tbody>
</table>

Notes: Output from MAGMA analysis, the 'Centrosome List', was tested for association with common variation in EA, IQ and SZ. Also presented post-hoc analysis of the centrosome list in control GWASs. No. of genes; number of genes in gene-set that were analysed, SE; standard error, T2D; type 2 diabetes, UC; ulcerative colitis, STR; stroke, CD; Crohn’s disease, CAD; coronary artery disease, AD; Alzheimer’s disease.
4 Functional Analysis of SDCCAG8

4.1 Introduction

The effect that SDCCAG8 genotype had on social cognition was the most significant finding in the neuropsychological analysis (Section 3.4). I chose to bring this centrosome/ciliary-associated gene forward for molecular analysis, as mutations in SDCCAG8 lead to ciliopathies and it has been associated with neurodevelopmental phenotypes. SDCCAG8 is mutated in several developmental disorders, notably the ciliopathy, Bardet-Biedl syndrome (Janssen et al., 2011; Otto et al., 2010; Schaefer et al., 2011), and Senior-Løken syndrome. Patients with SDCCAG8 mutations present with intellectual disability and sometimes seizures (Otto et al., 2010). Elements of obsessive compulsive behaviour can be seen in some patients with Bardet-Biedl syndrome; others have more severe behavioural phenotypes and develop autism or psychosis (Forsythe and Beales, 2013). SDCCAG8 has been shown to regulate the centrosomal accumulation of pericentriolar material and neuronal migration in the developing mouse cortex (Insolera et al., 2014). Furthermore, analyses of Sdccag8 gene-trap mice have implicated SDCCAG8 in the regulation of the DNA damage response, at least in kidney (Airik et al., 2014). GWAS analyses have implicated SDCCAG8 in SZ (Chang et al., 2015; Hamshere et al., 2012). SNPs surpassing genome-wide significance, a statistical threshold (P<5x10^-8) used to differentiate true positives from false positives in GWAS, have been reported at SDCCAG8 in the most recent and largest studies of SZ (Pardiñas et al., 2018), and EA (Lee et al., 2018). Of these SNPs, rs10803138 (associated with SZ) and rs2992632 (associated with EA) are in high linkage disequilibrium (LD; r^2=0.81), indicating that the same genetic variants are contributing to both phenotypes. These two SNPs are both reported to be associated, although not genome-wide significant, with IQ (P<0.001; Savage et al., 2018). For SZ risk SNP rs10803138, the G allele that is associated with increased SZ risk is also associated with reduced expression of SDCCAG8 (Table 3.3). These data imply that lower SDCCAG8 expression is associated with increased SZ risk.

I used reverse genetic analyses of SDCCAG8 to explore potential mechanisms by which genes that function at the centrosome could contribute to SZ and cognitive function.
4.2 Bioinformatic analysis of SDCCAG8

I used the NCBI database for analysis of the human \textit{SDCCAG8} locus. The \textit{SDCCAG8} gene spans 244.09 kbp on the p-arm of chromosome 1. Based on the current database there are 4 experimentally confirmed \textit{SDCCAG8} isoforms (Table 4.1). Kenedy et al. reported the presence of three different isoforms of \textit{SDCCAG8}, now known as isoforms 1, 3 and 4 (Kenedy et al., 2003). Isoform 1 represents the canonical sequence and is the longest transcript; isoforms 3 and 4 are splice variants of the full length protein missing exons 13 and 14 and exon 5, respectively. Isoform 2 is much shorter and encodes a truncated version of the protein. The remaining isoforms are solely predictions and there is no experimental confirmation of these isoforms. To date, isoform 1 is known as the predominant variant and has been most widely studied. This study was based on the analysis of isoform 1 of the \textit{SDCCAG8} gene.

\textbf{Table 4.1: Comparative analysis of predicted human SDCCAG8 isoforms}

<table>
<thead>
<tr>
<th>Isoform name</th>
<th>NCBI accession</th>
<th>Length (bp)</th>
<th>CDS length</th>
<th>Protein length (aa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>NM_001350246.1</td>
<td>2843</td>
<td>1238</td>
<td>412</td>
</tr>
<tr>
<td>2</td>
<td>NM_001350247.1</td>
<td>2731</td>
<td>1238</td>
<td>412</td>
</tr>
<tr>
<td>3</td>
<td>NM_001350248.1</td>
<td>2730</td>
<td>2237</td>
<td>745</td>
</tr>
<tr>
<td>4</td>
<td>NM_001350249.1</td>
<td>2647</td>
<td>1847</td>
<td>615</td>
</tr>
<tr>
<td>1</td>
<td>NM_006642.5</td>
<td>2581</td>
<td>2141</td>
<td>713</td>
</tr>
<tr>
<td>2</td>
<td>NM_001350251.1</td>
<td>3104</td>
<td>1238</td>
<td>412</td>
</tr>
<tr>
<td>X3</td>
<td>XM_011544026.3</td>
<td>2363</td>
<td>2030</td>
<td>676</td>
</tr>
<tr>
<td>X8</td>
<td>XM_011544028.3</td>
<td>2138</td>
<td>1805</td>
<td>601</td>
</tr>
<tr>
<td>X5</td>
<td>XM_024452548.1</td>
<td>6094</td>
<td>1943</td>
<td>647</td>
</tr>
<tr>
<td>X9</td>
<td>XM_024452547.1</td>
<td>2556</td>
<td>1751</td>
<td>583</td>
</tr>
<tr>
<td>X5</td>
<td>XM_024452540.1</td>
<td>2676</td>
<td>1943</td>
<td>647</td>
</tr>
<tr>
<td>X10</td>
<td>XM_005273018.2</td>
<td>2451</td>
<td>1718</td>
<td>572</td>
</tr>
<tr>
<td>X7</td>
<td>XM_017000105.2</td>
<td>5909</td>
<td>1904</td>
<td>634</td>
</tr>
<tr>
<td>X4</td>
<td>XM_005273013.5</td>
<td>6017</td>
<td>2012</td>
<td>670</td>
</tr>
<tr>
<td>X12</td>
<td>XM_005273022.4</td>
<td>5763</td>
<td>1220</td>
<td>406</td>
</tr>
<tr>
<td>X6</td>
<td>XM_017000104.2</td>
<td>2396</td>
<td>1925</td>
<td>641</td>
</tr>
<tr>
<td>X14</td>
<td>XM_005273023.5</td>
<td>1660</td>
<td>1082</td>
<td>360</td>
</tr>
<tr>
<td>X11</td>
<td>XM_024452549.1</td>
<td>2329</td>
<td>1610</td>
<td>536</td>
</tr>
<tr>
<td>X1</td>
<td>XM_024452537.1</td>
<td>7971</td>
<td>1919</td>
<td>639</td>
</tr>
<tr>
<td>X2</td>
<td>XM_024452539.1</td>
<td>2461</td>
<td>1847</td>
<td>615</td>
</tr>
<tr>
<td>X13</td>
<td>XM_011544030.3</td>
<td>6851</td>
<td>1196</td>
<td>398</td>
</tr>
</tbody>
</table>
As discussed previously, recessive mutations in \textit{SDCCAG8} cause Senior-Løken syndrome and Bardet-Biedl syndrome. To assess whether the positioning of these causative mutations could indicate important domains of the protein, all known disease-causing mutations were compiled and are visualized in Figure 4.1 and details of each mutation are presented in Table 4.2. The mutations are spread throughout the gene and are all patients are reported to be homozygotes or compound heterozygotes. All mutations result in a frameshift or aberrant splicing, resulting in premature termination of the protein and a complete absence of the full-length protein. This analysis does not pinpoint an essential or critical domain of the protein, but instead reinforces the idea that the expression of full-length SDCCAG8 is essential for normal development. Where data were available, the neurological symptoms patients presented with are presented in Table 4.2. These symptoms ranged from mild mental retardation (MR) to generalized seizures.

\textbf{Figure 4.1: SDCCAG8 genomic structure.}

Recessive mutations known to cause Bardet Biedl syndrome or Senior-Løken syndrome are indicated with arrows, and exons with blue boxes. Each mutation is numbered and referenced in Table 4.2, which provides further information on the mutations presented here.
<table>
<thead>
<tr>
<th>no</th>
<th>rsID</th>
<th>Exon</th>
<th>Start position</th>
<th>Mutation</th>
<th>Effect on protein</th>
<th>Clin var*</th>
<th>Disorder</th>
<th>Neurological Symp.</th>
<th>Inheritance</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>rs797045946</td>
<td>Intron 2</td>
<td>243434278</td>
<td>c.221+2A&gt;G</td>
<td>Splice acceptor variant</td>
<td>Path</td>
<td>SLS</td>
<td>No data</td>
<td>No data</td>
<td>Genetic Services Laboratory, University of Chicago</td>
</tr>
<tr>
<td>2</td>
<td>No data</td>
<td>5-7</td>
<td>243449574</td>
<td>c.421-?_740+?del</td>
<td>Deletion of a large region and a frame shift that results in premature termination.</td>
<td>No Data</td>
<td>SLS</td>
<td>Generalized seizures</td>
<td>Hom</td>
<td>Otto et al., 2011</td>
</tr>
<tr>
<td>3</td>
<td>rs797045947</td>
<td>5</td>
<td>243449634</td>
<td>c.481C&gt;T</td>
<td>AA Gln161 is predicted to change to a stop codon.</td>
<td>Path</td>
<td>SLS</td>
<td>No data</td>
<td>No data</td>
<td>Genetic Services Laboratory, University of Chicago</td>
</tr>
<tr>
<td>4</td>
<td>rs797045948</td>
<td>6</td>
<td>243456413</td>
<td>c.567G&gt;A</td>
<td>AA Tryptophan189 is predicted to change to a stop codon.</td>
<td>Path</td>
<td>SLS</td>
<td>No data</td>
<td>No data</td>
<td>Genetic Services Laboratory, University of Chicago</td>
</tr>
<tr>
<td>5</td>
<td>rs267607031</td>
<td>7</td>
<td>243468018</td>
<td>c.679A&gt;T</td>
<td>Substitution at AA 227, resulting in a stop codon</td>
<td>Path</td>
<td>BBS</td>
<td>Mild MR</td>
<td>Homo</td>
<td>Otto et al., 2011</td>
</tr>
<tr>
<td>6</td>
<td>rs772544112</td>
<td>7</td>
<td>243468035</td>
<td>c.696T&gt;G</td>
<td>Nonsense allele, at Tyr232 a stop codon is introduced.</td>
<td>No Data</td>
<td>BBS</td>
<td>MR</td>
<td>Compound het</td>
<td>Otto et al., 2011</td>
</tr>
<tr>
<td>7</td>
<td>No data</td>
<td>Intron 7</td>
<td>243469808</td>
<td>c.740+1delG</td>
<td>Splice site mutation</td>
<td>No Data</td>
<td>BBS</td>
<td>MR</td>
<td>Compund het</td>
<td>Otto et al., 2011</td>
</tr>
<tr>
<td>8</td>
<td>rs397515337</td>
<td>Intron 7</td>
<td>243468435</td>
<td>c.740+356C&gt;T</td>
<td>Loss of an exonic splice enhancer site, with the result of aberrant splicing introducing an in-frame stop codon</td>
<td>Path</td>
<td>BBS</td>
<td>Mild MR</td>
<td>Homo</td>
<td>Otto et al., 2011</td>
</tr>
<tr>
<td>9</td>
<td>No data</td>
<td>8</td>
<td>243471387</td>
<td>c.845_848delTTTG</td>
<td>4 Base pair deletion C283X, a premature stop codon is produced</td>
<td>No Data</td>
<td>BBS</td>
<td>Intellectual Disability Generalised seizures</td>
<td>Hom</td>
<td>Yamamura et al. 2017</td>
</tr>
<tr>
<td>10</td>
<td>No data</td>
<td>Intron 9</td>
<td>243480196</td>
<td>c.1068+1G&gt;A</td>
<td>Obligatory splice site mutation</td>
<td>No Data</td>
<td>SLS</td>
<td>Mild MR</td>
<td>Hom</td>
<td>Otto et al., 2011</td>
</tr>
<tr>
<td></td>
<td>rs</td>
<td></td>
<td></td>
<td></td>
<td>Likely</td>
<td>BBS</td>
<td>Developmental Delay</td>
<td>Hom</td>
<td>Schaefer et al., 2011</td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>-------</td>
<td>------</td>
<td>------------------</td>
<td>-----------------------</td>
<td>--------</td>
<td>----</td>
<td>---------------------</td>
<td>------</td>
<td>-----------------------</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>rs770084716</td>
<td>10</td>
<td>243493893</td>
<td>c.1120C&gt;T</td>
<td>Nonsense mutation, at Arg374 a stop codon is introduced</td>
<td>Likely</td>
<td>Path</td>
<td>BBS</td>
<td>Developmental Delay</td>
<td>Hom</td>
</tr>
<tr>
<td>12</td>
<td>rs387906218</td>
<td>11</td>
<td>243504458</td>
<td>c.1339-1340insG</td>
<td>Frameshift and premature termination at 463</td>
<td>Path</td>
<td>SLS</td>
<td>Mild MR</td>
<td>Homo</td>
<td>Otto et al., 2011</td>
</tr>
<tr>
<td>13</td>
<td>rs397515335</td>
<td>12</td>
<td>243507580</td>
<td>c.1420delG</td>
<td>Frameshift and premature termination at 493</td>
<td>Path</td>
<td>SLS</td>
<td>Brain scan normal</td>
<td>Homo</td>
<td>Otto et al., 2011</td>
</tr>
<tr>
<td>14</td>
<td>rs587777847</td>
<td>12</td>
<td>243507604</td>
<td>c.1444delA</td>
<td>Frameshift and premature termination at 493</td>
<td>Path</td>
<td>BBS</td>
<td>Mild MR</td>
<td>Compound het</td>
<td>Otto et al., 2011</td>
</tr>
<tr>
<td>15</td>
<td>rs587777846</td>
<td>14</td>
<td>243579014</td>
<td>c.1627_1630delGATA</td>
<td>Frameshift and premature termination at 566</td>
<td>Path</td>
<td>BBS</td>
<td>Mild MR</td>
<td>Compound het</td>
<td>Otto et al., 2011</td>
</tr>
<tr>
<td>16</td>
<td>No data</td>
<td>15</td>
<td>243581321</td>
<td>c.1796T&gt;G</td>
<td>Nonsense mutation, at Leu599 a stop codon is introduced</td>
<td>No data</td>
<td>SLS</td>
<td>No data</td>
<td>Homo</td>
<td>Otto et al., 2011</td>
</tr>
<tr>
<td>17</td>
<td>rs397515336</td>
<td>16</td>
<td>243589821</td>
<td>c.1946–1949delGTGT</td>
<td>Frameshift and premature termination at 658</td>
<td>Path</td>
<td>SLS</td>
<td>No data</td>
<td>Homo</td>
<td>Otto et al., 2011</td>
</tr>
</tbody>
</table>

SLS; Senior-Løken syndrome, BBS; Bardet-Biedl syndrome, het; heterozygous, hom; homozygous, MR; mental retardation, Path; Pathogenic

*Clinical significance value on Clin Var as recommended by the American College of Medical Genetics and Genomics and the Association for Molecular Pathology for variants interpreted for Mendelian disorders.
SDCCAG8 orthologues have been identified in several vertebrates and invertebrates. Using the Clustal Omega sequence alignment tool, I determined the level of protein homology between human SDCCAG8 and its orthologues in other species. I observed the highest level of conservation (≥70%) between human and other mammalian SDCCAG8 orthologues. When the human protein was compared to its fish orthologues (represented by zebrafish and West Indian Ocean coelacanth) there was a high level of sequence identity in only some parts of the protein sequence, resulting in an overall moderate homology (~45-56%). The lowest homology observed in this analysis was between human SDCCAG8 and the invertebrates analysed in this study, with less than 30% sequence identity observed (Figure 4.2). These analysis highlights the possibility of highly conserved domains with functional significance in SDCCAG8 and its orthologues.

To identify highly conserved domains within SDCCAG8 I first performed sequence analysis using SMART (http://smart.embl-heidelberg.de/), to identify predicted protein domains. The only domains identified based on the protein sequence were coiled coil regions. The information generated using SMART was overlaid with conservation across the length of the SDCCAG8 protein (Figure 4.3). The conservation was calculated based on the alignment of all vertebrate orthologues used in Table 4.2. The conservation score is an automatically calculated quantitative alignment annotation, which measures the number of conserved physico-chemical properties conserved for each column of the alignment. This score was generated in Jalview (http://www.jalview.org/). Its calculation is based on the one used in the AMAS method of multiple sequence alignment analysis (Livingstone and Barton, 1994), and the maximum conservation score is 11. In general there is a high level of conservation between SDCCAG8 vertebrate orthologues.
Figure 4.2: Comparative analysis of human SDCCAG8 and its orthologues

The level of conservation is measured by the % of identical amino acid residues found in the analysed sequences. The referred values do not account for computational errors associated with predicted protein sequence. The protein sequences used for this comparison and their NCBI protein database accession number are as follows: Human: NP_006633.1, House mouse: NP_084032.1, Zebrafish: NP_001333091.1, Cattle: NP_001180038.2, Tropical clawed frog: XP_012819175.1, Chicken: XP_004935496.1, Barn owl: XP_009961599.1, West Indian Ocean coelacanth: XP_005998101.1, Whale shark: XP_020392837.1, Green turtle: XP_007058024.1, Melon fly: JAD10820.1, Placozoan 16S haplotype: H2RDD44353.1, Flatworm: PAA89887.1, Fresh water polyp: CDG69114.1 and Pale anemone: KXJ24763.
Figure 4.3: Conserved regions of SDCCAG8 in vertebrates
Upper part of the image is the output from SMART, the human SDCCAG8 sequence; NP_006633.1 was used as input. Green regions are coiled coil domains no other domains were identified from the protein sequence. The bottom panel represents conservation across vertebrate orthologues of SDCCAG8; The protein sequences used for this comparison and their NCBI protein database accession number are as follows: Human: NP_006633.1, House mouse: NP_084032.1, Zebrafish: NP_001333091.1, Cattle: NP_001180038.2, Tropical clawed frog: XP_012819175.1, Chicken: XP_004935496.1, Barn owl: XP_009961599.1, West Indian Ocean coelacanth: XP_005998101.1, Whale shark: XP_020392837.1, and Green turtle: XP_007058024.1. Alignment was performed using Clustal omega and conservation score was generated for each alignment column by Jalview. Gaps in the human SDCCAG8 sequence generated during alignment were manually removed to allow conservation scores to be mapped back to the human protein.

4.3 Localisation of SDCCAG8 throughout the cell cycle
A mouse monoclonal antibody (66284-1-Ig, Proteintech) was used in asynchronous hTERT-RPE1 cells to define the localisation of endogenous SDCCAG8 (Figure 4.4). Immunofluorescence microscopy analysis of SDCCAG8 revealed that it is indeed a core centrosomal protein that is consistently localised to the centrosome at all stages of the cell cycle. The immunolocalisation studies also revealed a specific pattern in the appearance of SDCCAG8: in G1 cells there were two signals observed, suggesting SDCCAG8 is loaded on both centrioles. However, in the majority of cells there was one signal weaker than the other. In 76% of G1 cells, the weaker SDCCAG8 signal co-localised with CEP164, a mother centriole marker, suggesting that SDCCAG8 is preferentially loaded onto the daughter centriole. This trend continued throughout the
cell cycle, the preferential loading of SDCCAG8 onto the daughter centriole was observed in 62% of cells in metaphase.

Figure 4.4: SDCCAG8 localisation throughout the cell cycle
IF microscopy visualisation of SDCCAG8 in hTERT-RPE1 cells over the cell cycle. Asynchronous cells were fixed and stained with anti-SDCCAG8 (green), CEP164 (red) and DNA (blue). Scale bar, 5 µm and inset scale bar, 2.5 µm.

4.4 Genome editing of the SDCCAG8 locus in hTERT-RPE1 and SH-SY5Y cells using CRISPR/Cas9

To study the functions of SDCCAG8 in human cells, I generated SDCCAG8 deficient cells using CRISPR/Cas9 genome editing, a very robust and commonly used RNA-guided gene editing technique. Using the guide RNA (gRNA) database (available at http://arep.med.harvard.edu/human_crispr/) provided by Mali et al. (2013), I selected a gRNA sequence that can potentially disrupt the SDCCAG8 genomic locus. To ensure higher likelihood of disrupting all the SDCCAG8 transcripts/isoforms, I designed a gRNA against exon 3, which was chosen due to its proximity to the start codon, reducing the possibility of cells expressing a potentially stable coding transcript. hTERT-RPE1 and SH-SY5Y cells were transfected with a pX330 plasmid containing the gRNA sequence (the oligo sequence is listed in Table 2.12) along with a plasmid coding for neomycin antibiotic resistance. After 10 days of antibiotic selection with geneticin, colonies were
individually selected and subsequently examined by western blot and IF microscopy for their loss of SDCCAG8 expression (Figure 4.5). I screened 36 clones in hTERT-RPE1 cells and obtained 2 clones that lacked detectable SDCCAG8 by western blot and 19 clones in SH-SY5Y cells and obtained 6 clones. A representative immunoblot for the clones that were analysed further is shown in Figure 4.4 A. To confirm the loss of SDCCAG8 at the centrosome, I used a commercially available monoclonal anti-SDCCAG8 antibody for IF microscopy (Figure 4.5 B). I observed no SDCCAG8 signal at the centrosome, as visualised by staining for centrin. These data showed that I successfully targeted the human SDCCAG8 gene locus and ablated the protein expression in human hTERT-RPE1 and SH-SY5Y cells.

![Figure 4.5: Generation of SDCCAG8 deficient hTERT-RPE1 and SH-SY5Y cell lines](image)

A. Western blot confirming the loss of SDCCAG8 expression in hTERT-RPE1 and SH-SY5Y clones compared to their respective wildtype cells. The SDCCAG8 protein can be visualized at its expected molecular weight of 83kDa indicated on blot by arrow, the lower band appears to be isoform 4 of the protein which is also targeted in the KO clones generated, indicated on the blot by an arrowhead. Non-specific band on the blot indicated by ‘*’. Actin is shown as a loading control. B. Representative immunofluorescence images confirms the loss of SDCCAG8 (red) at the centrosome in KO hTERT-RPE1 clones. Cep135 (green) was used to visualise the centrosome and DNA is stained in blue (DAPI). C. Representative immunofluorescence images confirm the loss of SDCCAG8 (red) at the centrosome in SH-SY5Y KO clones. Cep135 (green) was used to visualise the centrosome and DNA is stained in blue (DAPI). Scale bar, 5µm and inset scale bar, 1µm.
To determine the exact nature of the Cas9-mediated DNA modification of the SDCCAG8 locus, I carried out PCR reactions on genomic samples from CRISPR knockout clones. The PCR amplicons were sequenced for the targeted region on exon 3 using the CRISPR screen primers (Table 2.12). The sequencing data shows that both hTERT-RPE1 clones (KO1 and KO2) have a single base pair deletion in exon 3 (Figure 4.6 B). This deletion resulted in the introduction of a premature stop codon. I sequenced two SH-SY5Y CRISPR clones and observed one clone with a 16 bp deletion on one allele and a 4 bp deletion on the other (Figure 4.6 C). This clone (referred to as ‘SH-SY5Y KO1’) was used in the remainder of my experimental investigations. The second clone sequenced had a 2 bp insertion. All clones generated have the same premature stop codon generated at the beginning of exon 4.

Generally, transcripts with a premature stop codon trigger entry into a pathway of mRNA decay, leading to rapid reduction of their availability for translation (Jacobson and Peltz, 1996). Furthermore, the predicted molecular weight of the truncated form of SDCCAG8 produced as a result of the premature stop codon introduced in all clones is 16.1kDa (146 aa). It is likely that the functionality of this truncated protein, even if expressed, would be insufficient to carry out any of SDCCAG8’s roles. Therefore, I concluded that the SDCCAG8 KO cell lines are complete nulls.
Figure 4.6: Sequencing confirmation of SDCCAG8 KO clones
A. Schematic representation of SDCCAG8 gene. The region to which the gRNA was designed is underlined below the exon, highlighting the targeted CRISPR exon. B. Nucleotide sequences and predicted protein sequence of wildtype and KO hTERT-RPE1 clones. Sequences are coloured according to their respective exons. Both clones have a deletion of 1 bp. C. Nucleotide sequences and predicted protein sequence of wildtype and KO SH-SY5Y clones. Sequences are coloured according to their respective exons. SH-SY5Y KO 1 has a different mutation in each allele, allele 1 (A1) was a 16 bp deletion, while allele 2 was a 4 bp deletion. SH-SY5Y KO 2 has a 2 bp insertion on both alleles. All mutations lead to a frameshift and the generation of the same premature stop codon (*) at the beginning of exon 4.

4.5 Generation of RPE1 rescue cell lines and proliferation analysis
To confirm that any phenotype observed as a result of loss of SDCCAG8 was indeed due to its absence and not an off-target effect, I generated SDCCAG8 rescue cell lines in RPE1 cells. This was achieved using pcDNA3.1-SDCCAG8, which was generated by subcloning SDCCAG8 cDNA from pEGFP-N1-SDCCAG8 into the expression plasmid pcDNA3.1(+) using NheI and BamHI. Successful cloning was verified by commercial DNA sequencing. Western blotting analysis of the clones picked after 14 days under blasticidin selection shows variation in the level of SDCCAG8 re-expressed among stable SDCCAG8-rescue clones (Figure 4.7 A). After screening, I selected the rescue clone with SDCCAG8 expression level comparable to the wild-type (Figure 4.7 A, clone 2) for further experiments in this project. Using immunofluorescence microscopy, I also
confirmed the re-expression of SDCCAG8 and its localisation to the centrosome in the SDCCAG8-rescue cells (Figure 4.7 B).

Figure 4.7: Stable re-expression of SDCCAG8 to generate rescue RPE1 cell line. A. Cell lysates from the indicated cell lines were analysed by western blotting using a monoclonal antibody against SDCCAG8. Red arrow indicates full-length SDCCAG protein, ‘*’ labels non-specific band. Ponceau staining was used as a loading control. B. IF microscopy images of RPE1 wildtype (WT), SDCCAG8 knockout (KO) and SDCCAG8 rescue (Rescue) cells after immunostaining with antibodies that detect SDCCAG8 (Red) and CEP135 (Green). Scale bars, 10 µm and inset scale bars, 1 µm.

Since SDCCAG8-deficient cells were viable in culture, I used cell proliferation analysis to detect any cell cycle aberrations in both SDCCAG8-deficient and rescue cell lines. As shown in Figure 4.8, the proliferation profiles remained unaltered in both cell lines and were comparable to the wild-type population.
116

Figure 4.8: SDCCAG8-deficient and rescue cell lines show no proliferation defect
Cells were plated at 5x10^4 cells and counted at every 24 h for 72 h. N=3 independent experiments and error bar indicates standard error of mean (SEM). No significant difference was observed between WT, KO and rescue cell line at any time point.

4.6 Reduction in PCM and satellite recruitment in SDCCAG8 deficient cells
The centrosome is composed of several hundred proteins that serve different functions at different stages of the centrosome cycle. SDCCAG8 localised to the centrosome in both interphase and mitosis, so I investigated the structural integrity of the centrosome upon absence of SDCCAG8 at these stages. I stained KO cells with various antibodies to known centrosomal components. I visualised the localisation pattern of CEP164, CEP290, Centrin, C-NAP1, centrobin, pericentrin and PCM-1 in G1; identified by the presence of a single centrosome (Figure 4.9) and metaphase (Figure 4.10). I studied the dynamics of distal appendages, the distal lumen and daughter centrioles and by staining for CEP164, Centrin and centrobin, respectively. I observed a single CEP164 signal at each centrosome, which was indicative of intact distal appendages on the mother centriole. Cep164 is a component of the distal appendage that resides only on the mother centriole and is essential for ciliogenesis (Graser et al. 2007a). Normal distribution and localisation of the maturation marker CEP164 suggests that SDCCAG8 deficiency does not affect centriole maturation. Centrin is visualised in SDCCAG8-deficient cells at both centrioles distal to CEP135, a proximal lumen marker (Kim et al., 2008b), which is consistent with centrin’s reported location at the distal lumen. Centrobin staining remains intact in the absence of SDCCAG8, a single signal at each centrosome was observed in all cases, which represents the loading of centrobin on the daughter centrosome, as previously described (Zou et al., 2005). I also visualised the integrity and dynamics of the PCM through pericentrin staining. The PCM serves as an essential hub for
microtubule nucleation and organisation in interphase cells and during spindle assembly in mitotic cells (Woodruff et al., 2014). In the SDCCAG8 KO cells, pericentrin localisation appeared grossly normal. However, quantification revealed a decline in pericentrosomal pericentrin levels to 76% of the wild-type controls (Figure 4.9 B), consistent with observations made in knockdown experiments in murine neurons (Insolera et al., 2014). I also studied the dynamics of satellite components, CEP290 and PCM-1. While the localisation of both satellite markers appeared grossly normal, the levels of pericentrosomal PCM-1 were reduced to 53% of the wild-type cell controls (Figure 4.9 C). CEP290 distribution is observed as granular structures adjacent to but not directly associated with the centrioles, as previously described (Kim et al., 2008 a) and this remains unchanged in SDCCAG8 deficient cells. Finally, I studied the dynamics of the centrosomal linker by observing the localisation pattern of C-NAP1. C-NAP1 localises to and acts as a tether at the proximal ends of both parental centrioles in interphase (Fry et al., 1998) as seen in Figure 4.6. Linker disassembly occurs at the beginning of mitosis (Hardy et al., 2014), as the localisation of C-NAP1 to the centrosome can no longer be observed during metaphase (Figure 4.10). This pattern is unchanged in SDCCAG8-deficient cells. These data together suggest that centriolar composition is grossly normal, although there is a reduction in the recruitment of the PCM component pericentrin and the satellite component PCM-1 in SDCCAG8-deficient cells.
Figure 4.9: Reduction in the recruitment of PCM and satellite components in G1 hTERT-RPE1 cells deficient of SDCCAG8.
A. Wild-type and SDCCAG8-deficient cells were fixed and stained with antibodies to CEP164, Cep290, Centrin, C-NAP1, CNTROB, PCNT and PCM-1 (all in red) as indicated, CEP135 (in green) and DNA in blue. Scale bars, 10 µm and inset scale bars, 1 µm. B. Quantification of fluorescence intensity of pericentrosomal pericentrin. Each individual data point, mean and SEM are depicted, for three independent experiments in which at least 30 cells were quantified. * P≤0.01 compared to RPE1 WT by unpaired t-test. C. Quantification of fluorescence intensity of pericentrosomal PCM-1. Each individual data point, mean and SEM are depicted, for two independent experiments in which at least 30 cells were quantified. *** P≤0.001 compared to RPE1 WT by unpaired t-test.
Figure 4.10: Loss of SDCCAG8 has no effect on localisation and composition of centrosomal proteins in mitotic hTERT-RPE1 cells. Wild-type and SDCCAG8-deficient cells were fixed and stained with antibodies to CEP164, Cep290, Centrin, CNAP-1, CNTROB, PCNT and PCM-1 (all in red) as indicated, CEP135 (in green) and DNA in blue. Scale bars, 10 μm and inset scale bars, 1 μm.
4.7 SDCCAG8-deficient cells show no centriole duplication defects

My IF analysis showed no major defect in the localisation, distribution or abundance of centrosomal proteins investigated in this study (section 4.5). Using Cep135 and centrin as centriole markers, I observed no increase in the frequency of acentriolar and monocentriolar cells in the SDCCAG8-deficient cells (Figure 4.11). Altered centrosome numbers or centrosome amplification arise in response to DNA damage (Saladino et al., 2009). I observed no significant increase in centrosome amplification 48 h after treatment with 5Gy ionising radiation. This suggests that SDCCAG8 is not required for centriole duplication.

![Image of centriole numbers and percentage of cells with different centriole numbers before and after IR treatment.](image)

**Figure 4.11 Regulation of centriole duplication remains intact in SDCCAG8 depleted cells**

Centrioles were visualized by centrin and CEP135 staining before and 48 hours after cells were treated with 5 Gy IR. Representative images are shown on the top, scale bar, 1 µm. Bar graph shows mean ± SEM of three separate experiments were the number of centrioles were counted in at least 100 cells.
4.8 Dynamics of SDCCAG8 localisation in ciliated cells

The first step in investigating the role in SDCCAG8 in ciliogenesis was to determine its localisation dynamics in serum-starved hTERT-RPE1 cells. Serum starvation induces ciliogenesis in ~80% of RPE1 wildtype cells. In proliferating cells, SDCCAG8 localises to both centrioles, with increased loading on the daughter centriole (Figure 4.4). In ciliated populations, I observed SDCCAG8 localisation to the daughter centriole but a proportion of the protein was also found in a discrete band pattern near the base of the primary cilium (Figure 4.12). In ciliated cells, mother centrioles were identified by the presence of the primary cilium marker, acetylated tubulin. Given the distinctive localisation pattern, I investigated whether SDCCAG8 associated with a transition zone marker and observed co-localisation of SDCCAG8 with CEP290 following serum starvation. In the mouse retina, SDCCAG8 has previously been shown to localise to the transition zone of photoreceptor cilia (Otto et al., 2010), which are uniquely specialised sensory cilia. However, no direct co-localisation with CEP290 was described in this analysis of photoreceptor cilia. This analysis suggests that a proportion of SDCCAG8 protein localises to the transition zone of primary cilia in human cells.

![Figure 4.12: SDCCAG8 localisation in ciliated cells](image)

hTERT RPE1 cells were serum starved for 48 h, microtubules were destabilised and cells were fixed and stained with antibodies to SDCCAG8 (green) and Acetylated tubulin or CEP290 (red). DNA was stained with DAPI. Scale bar, 5 µm, inset scale bar, 1 µm.
4.9 Loss of SDCCAG8 reduces ciliation frequency in human cells

Primary cilia are microtubule-based antenna-like organelles that assemble after the docking of mother centriole to the plasma membrane. Knockout of SDCCAG8 in mouse studies has demonstrated no gross ciliogenesis defects, in either radial glial progenitors in the ventricular zone of the cortex (Insolera et al., 2014) or renal tubule cells (Airik et al., 2014). However, the use of SDCCAG8 siRNA in cultured human cells has been shown to impair ciliation (Airik et al., 2016). I tested the impact of SDCCAG8 ablation on the assembly of primary cilia, in the SDCCAG8 deficient hTERT-RPE1 cell line, as well as in the SDCCAG8 deficient SH-SY5Y cells (Figure 4.13).

Cells were serum-starved for 48 h and fixed for immunofluorescence microscopy. As shown in Figure 4.13 A, immunofluorescence microscopy of serum-starved wild-type cells showed the presence of acetylated tubulin which marks the ciliary axoneme. SDCCAG8 deficient cells showed a significant reduction in ciliation frequency, which was rescued by the stable re-expression of SDCCAG8 in the deficient RPE1 cell line (Figure 4.13 B). This ciliogenesis defect was also evident in the SH-SY5Y cells. Cells were serum-starved and fixed as before, and in WT cells ARL13b clearly marks the ciliary membrane (Figure 4.12 C). To the best of my knowledge, this is the first time that the ciliation capacity of SH-SY5Y cells has been demonstrated. The loss of SDCCAG8 in SH-SY5Y cells leads to a significant reduction in primary cilium frequency (Figure 4.13 D).

In addition to the reduction in cilia number observed here, the cilia that did form in SDCCAG8-deficient cells were also shorter than those in wild-type controls, a phenotype which was also rescued by restoration of SDCCAG8 expression (Figure 4.14). These data together suggest a regulatory role for SDCCAG8 in primary ciliogenesis.
Figure 4.13: Loss of SDCCAG8 significantly impacts cilia formation

A. hTERT-RPE1 cells were serum starved, microtubules were destabilized, and the cells were then fixed and stained with acetylated tubulin (red), a primary cillum marker, and CEP135 (green), a centriolar marker. Representative images of hTERT-RPE1 wildtype and KO cells are shown. Scale bar, 5 µm, inset scale bar, 1 µm. B. Quantification of ciliation frequency. Histogram depicts the mean ± SEM of 4 independent repeats of which at least 100 cells were counted. ***, $P \leq 0.001$; in comparison to RPE1 WT cells by unpaired t-test.

C. SH-SY5Y cells were serum starved, microtubules were destabilized, and the cells were then fixed. The ciliary membrane was stained with an antibody against ARL13B (red), while the centrosome was marked with centrin (green) and DNA (blue) was stained with DAPI. Representative images of SH-SY5Y wildtype and KO cells are shown. Scale bar, 5 µm, inset scale bar, 1 µm. D. Quantification of ciliation frequency. Histogram depicts the mean ± SEM of 3 independent repeats of which at least 100 cells were counted. **, $P \leq 0.01$; in comparison to SH-SY5Y WT cells by unpaired t-test.
Figure 4.14: Loss of SDCCAG8 significantly impacts ciliary length
Quantitation of cillum length in RPE1 cells of the indicated genotype. Individual data points are shown, as is the mean cillum length + SEM from at least 88 ciliated cells quantitated in 3 separate experiments. **, P ≤ 0.01; ***, P ≤ 0.001; ns, non-significant in comparison to WT RPE1 cells by unpaired t-test.

4.10 Localisation of key ciliary regulators in the absence of SDCCAG8

In an attempt to understand the molecular mechanism by which SDCCAG8 depletion inhibits primary cilium formation, I serum starved RPE1 cells for 48 h and analysed the localisation and abundance of known regulators of ciliation in WT and SDCCAG8-deficient cells. As shown in Figure 4.15, the centriolar localisation of the negative regulators of ciliogenesis, CP110 and CEP97, remains grossly normal. In cells that have not undergone ciliogenesis they form a cap, which is removed from mother centrioles during the extension of primary cilia (Spektor et al., 2007). However CP110 pericentrosomal abundance is significantly reduced in asynchronous SDCCAG8-deficient cells and significantly increased in serum-starved SDCCAG8-deficient cells (Figure 4.16 A). Cells were also stained for IFT88, an intraflagellar transport component that is essential for ciliogenesis (Goetz et al., 2012). IFT88 can be visualised along the axoneme, indicative of a normal protein localisation. Quantification also revealed no significant changes in protein abundance along the length of the primary cilium (Figure
4.16 B). The centriolar satellite protein, PCM-1, was also visualised, expression of which is required for ciliogenesis (Wang et al., 2016). The satellite localisations of PCM-1 remains intact in SDCCAG8 deficient cells. I demonstrated earlier that asynchronous SDCCAG8-deficient cells have a significant reduction of PCM-1 abundance (Figure 4.9). However, the pericentrosomal PCM-1 levels normalise in SDCCAG8 cells that have been serum starved for 48 h (Figure 4.16 C). TTBK2, a known positive regulator of ciliogenesis (Goetz et al., 2012), remains localised to the basal body of ciliated cells, and no reduction in expression was observed (Figure 4.16 D). Most of the markers tested in this analysis all appear grossly normal, thus mislocalisation of the markers tested is an unlikely source of the ciliogenesis defect. However, there is a significant alteration to the abundance of CP110, a known negative regulator of ciliogenesis in SDCCAG8 deficient cells.
**Figure 4.15: Localisation of ciliary regulators remains intact in SDCCAG8 deficient cells**

RPE1 WT and SDCCAG8 KO cells were serum starved for 48 h, fixed and stained with acetylated tubulin (red) and the markers listed (green). Scale bar, 5 µm, inset scale bar, 2.5 µm.
Figure 4.16: CP110 abundance is affected by SDCCAG8 deficiency

A. Quantification of the fluorescence intensity of pericentrosomal CP110, in asynchronous and 48 h serum-starved cells. Individual data points, mean and standard error of the mean are displayed from two independent experiments in which at least 30 cells were quantified. *** P ≤ 0.001 compared to RPE1 WT by unpaired t-test. ** P ≤ 0.01 compared to RPE1 WT by unpaired t-test. * P ≤ 0.05 compared to RPE1 WT by unpaired t-test. B. Quantification of the fluorescence intensity of IFT88 along the cilium, divided by cilium length (µm), in 48 h serum-starved cells. Individual data points, mean and standard error of the mean are displayed from two independent experiments in which at least 20 cells were quantified.

C. Quantification of the fluorescence intensity of pericentrosomal PCM-1, in asynchronous and 48 h serum-starved cells. Individual data points, mean and standard error of the mean are displayed from two independent experiments in which at least 30 cells were quantified.

D. Quantification of the fluorescence intensity of pericentrosomal TTBK2, in asynchronous and 48 h serum-starved cells. Individual data points, mean and standard error of the mean are displayed from two independent experiments in which at least 30 cells were quantified.
4.11 Disruption of ciliary signalling in the absence of SDCCAG8

Primary cilia are essential organelles in the hedgehog (Hh) signal transduction pathway. Smoothened (SMO) is a highly conserved seven-transmembrane protein that relocates to the primary cilium to transduce signals in response to Sonic Hh (SHh) activation (Corbit et al. 2005, Goetz and Anderson 2010). To investigate the signalling capacity of residual cilia in SDCCAG8-deficient cells, I examined the responsiveness of these cilia to external stimuli. After treatment with a smoothened agonist (SAG), which activates the SHh pathway, translocation of SMO to the axoneme was used as a readout of SHh pathway activation. I saw a high level of pathway activation in RPE1 WT cells, with a marked decrease in SMO translocation to the axoneme in SDCCAG8 deficient cells, which can be rescued partially by re-expression of SDCCAG8 (Figure 4.15, A and B). This shows that the residual cilia in SDCCAG8 knockout cells are defective in ciliary signalling pathways.

Figure 4.17: Defective Smo localisation in residual SDCCAG8 deficient cilia
A. Localisation of Smoothened (green) in hTERT-RPE1 WT and SDCCAG8 deficient 48-h serum starved cells, after 4 h treatment with 100 mM SAG. Scale bar, 5 µm, inset scale bar, 2.5 µm. B. Quantification of SMO localisation to cilia. Histogram depicts the mean ± SEM of 3 independent repeats in which at least 5 ciliated cells were counted. ***, P ≤ 0.001; in comparison to RPE1 WT cells by unpaired t-test.

4.12 Loss of SDCCAG8 significantly impacts cell migration

Knockout of SDCCAG8 expression has been shown previously to impair neuronal migration in the developing cortex of mice (Insolera et al., 2014). Given that some of the most dysregulated pathways in the absence of SDCCAG8 were generation of neurons and axon guidance signalling (discussed in detail in Section 5.7), I hypothesised that
migration would be impaired in the SDCCAG8-deficient SH-SY5Y cells. Migration of SH-SY5Y cells was assessed using transwell inserts between serum-free and serum-containing conditions. At 16 h after exposure to serum as an attractant, most cells remained within pores and had not completed migration to the underside of the insert, although by 24 h migration was complete and WT cells regained their morphology (Figure 4.16 A). Colorimetric quantification of crystal violet uptake by migrated cells revealed that the migration of SDCCAG8-deficient cells was significantly defective (Fig 4B), consistent with previous findings in mice and with the impact of SDCCAG8 deficiency on gene expression.

**Figure 4.18: Defective migration in SDCCAG8 depleted SH-SY5Y cells**
A. Representative images of the underside of the transwell inserts after cells had migrated for the time indicated and had been stained with crystal violet. Scale bar = 250µm. B. Quantification of cell migration based on crystal violet staining intensity. Absorbance was read at 595nm, and percentage of migrated cells was calculated relative to positive controls where all cells were plated on the other side of the transwell insert. Lines represent the mean of individual date points shown on the graph. **, P ≤ 0.01; in comparison to SH-SY5Y WT cells by unpaired t-test.

4.13 Loss of SDCCAG8 reduces SH-SY5Y cells’ ability to differentiate

Complex neuronal morphology is required in order for connections to be formed within the brain. Centrosomes have been shown to play a role in neurite formation and axon development (Andersen and Halloran, 2012). GO terms that were highly enriched for SDCCAG8-dysregulated genes (Section 5.7) included cell differentiation and generation of neurons. I therefore hypothesised that the transcriptional dysregulation caused by SDCCAG8 depletion may affect the cells’ ability to differentiate. To test this idea, both
wild-type and SDCCAG8-deficient SH-SY5Y cells were differentiated with retinoic acid for 7 days, based on a previously optimised protocol (Dwane et al., 2013). After 7 days, phase contrast images of the cells were taken to assess morphological appearance of the cells. Undifferentiated WT and SDCCAG8−/− SH-SY5Y cells have a flat phenotype with few projections, while differentiated WT cells show extensive and elongated neuron projections (Figure 4.17 A). Cells that lack SDCCAG8 failed to undergo this morphological change and remain very flat. Furthermore, the formation of neurite projections, a measure of differentiation, was greatly reduced in SDCCAG8-deficient cells (Figure 4.17 B). These data demonstrate a marked deficiency in neuronal differentiation capacity in the absence of SDCCAG8.
Figure 4.19 Loss of SDCCAG8 affects cells ability to differentiate

A. Representative images of differentiation in SHSY5Y cells of the indicated genotype following treatment with vehicle or retinoic acid for 7 days. Scale bar, 250 µm. B. Quantitation of neuritic projections as a measure of differentiation in SHSY5Y cells of the indicated genotype. Data show mean length + SEM of N=292 (WT) and N=244 (KO) projections analysed in 2 separate experiments. ***, P ≤ 0.0001; in comparison to WT SH-SY5Y cells by unpaired t-test.
4.14 Discussion

The bioinformatic analysis of SDCCAG8 demonstrated the high level of homology among the vertebrate SDCCAG8 orthologues. They share synteny and a high degree of amino acid conservation, which was lowest at both the C and N termini of the protein. The pattern of conservation observed may be indicative of domain conservation and potentially conserved functionality. However, the crystal structure of the protein has not been solved and the only domains predicted by bioinformatic analysis were coiled-coiled domains which did not overlap with regions of high or low conservation in this analysis.

Multiple proteins that localise to the centrosome have been found to contain a high degree of coiled-coil domains (Salisbury, 2003). Coiled-coil domains are thought to be important in creating a scaffold, to which other proteins can be recruited, allowing the centrosome to become an effective signalling hub (Kuhn et al., 2014). While many centriolar proteins maintained grossly normal localisations in the absence of SDCCAG8, in this analysis the abundance of PCM-1 and PCNT proteins were both reduced. This may suggest that the coiled-coil domains within SDCCAG8 contribute to the formation of a scaffolding to which PCM and satellite components can be recruited more effectively.

SDCCAG8 orthologues were also found in arthropods, cnidarians and even placozoa, although with much lower protein identities than vertebrates. Previous evolutionary analysis of centriole components have revealed that there is a surprising amount of malleability in centriole protein composition. Some species have extensive loss of what are considered core centriole proteins or the divergence of the proteins themselves is sufficient to make them unrecognizable from one species to the next (Hodges et al., 2010). This extensive protein divergence could explain my inability to detect SDCCAG8 orthologues in Drosophila melanogaster and Caenorhabditis elegans, amongst other species. SDCCAG8 evolutionary distribution suggests that it was not a core protein in ancestral centrioles. SDCCAG8 orthologues seem to be limited to metazoa, whilst orthologues for core centriolar components are found in at least four major eukaryotic groups including metazoa, fungi, excavata, amoebozoa and plantae (Hodges et al., 2010).
To my knowledge, this study is the first to use CRISPR/Cas9 technology to disrupt the human SDCCAG8 locus, achieving a full depletion of functional SDCCAG8. Reproducible phenotypes from my SDCCAG8-deficient clones have been observed in two separate cell types and also in clones with different mutations, confirming the authenticity of the reported phenotypes. In addition to this, the phenotypes investigated here in the RPE1 cell line have been rescued by re-expressing of SDCCAG8, confirming that the observed phenotypes are specific to loss of functional SDCCAG8. Previous studies have used siRNA mediated depletion of SDCCAG8 in human cells. However, off-target effects, transient depletion and residual functionality of proteins, can hamper the reproducibility and validity of reported findings from siRNA studies (Boettcher and McManus, 2016, Birmingham et al., 2006, Jackson et al., 2003). In addition to this siRNA mediated knockdown and knockout of centriolar components have been shown to have different effects on ciliary function (Hall et al., 2013). These studies highlight the importance of generating knock-outs when functionally characterising centrosome proteins and their role in human disease.

SDCCAG8 was shown to localise to the transition zone of primary cilia and loss of the protein leads to defects in ciliogenesis in human cells. The transition zone was been ascribed a gate keeper function, permitting ciliary proteins to cross the barrier and excluding other proteins (Czarnecki and Shah, 2012). It is possible that loss of SDCCAG8 disrupts the TZ, so that ciliary proteins may be excluded from the axoneme and aggregate or non-ciliary proteins may enter erroneously. Mutations of many known components of the TZ lead to the development of ciliopathies, including SLS and BBS among others (Gonçalves and Pelletier, 2017), demonstrating that normal function of the TZ is critical for efficient ciliogenesis. In line with a possible TZ dysfunction, SMO was not incorporated efficiently into the axoneme after treatment with a SAG agonist in the SDCCAG8-deficient cells. However in the SDCCAG8-deficient cells, the abundance of IFT88 along the length of the primary cilium was not decreased, suggesting that, at least for this protein, there was no exclusion from the axoneme. Perhaps a timecourse analysis may have revealed a reduction in the efficiency by which IFT88 was granted access to the axoneme. Quantification of other ciliary proteins may clarify whether SDCCAG8
deficiency leads to a disruption of TZ function. However, these investigations were beyond the scope of this study.

Another possible explanation for the localisation of SDCCAG8 at the transition zone could be that the protein plays a role in effectively recruiting other proteins to this location rather than directly influencing TZ function. I have shown that SDCCAG8 deficiency leads to defective recruitment of pericentrin and PCM-1 in asynchronous cells, even though the recruitment of PCM1 normalises following 48 h serum starvation. This suggests that SDCCAG8 regulates the efficiency of satellite recruitment, although pericentrosomal recruitment is eventually possible in its absence. SDCCAG8 regulation of satellite components may contribute to the observed ciliogenesis defect. Multiple TZ components have been shown to interact with satellite components, which suggests that satellite proteins play an important role in trafficking TZ components to the primary cilium (Quarantotti et al., 2019, Gupta et al., 2015, Klinger et al., 2013). SDCCAG8 has been shown to interact with the known satellite components, PCM1, OFD1 and CEP131 (Airik et al., 2016, Insolera et al., 2014), all of which have been shown to contribute to ciliogenesis. PCM1 sequesters an E3 ligase, Mib1, and removal of PCM1 leads to increased ubiquitination of satellite component CEP131 and impaired ciliogenesis (Wang et al., 2016). CEP131 localises to satellites but also to the TZ of ciliated cells (Hall et al., 2013), and CEP131 has been shown to contribute to ciliogenesis (Hall et al., 2013, Wilkinson et al., 2009). OFD1 is also a regulator of ciliogenesis, which has been shown to be necessary for distal appendage formation and IFT88 recruitment (Singla et al., 2010). Removal of OFD1 from centriolar satellites by autophagy promotes ciliogenesis (Tang et al., 2013). I also observed a disruption in CP110 localisation in SDCCAG8 deficient cells. CP110 is a negative regulator of ciliogenesis (Spektor et al., 2007). Msd1 is a centriolar satellite protein that tethers MTs to the centrosome (Hori et al., 2014). Upon Msd1 depletion, CP110 was shown to localise to centriolar satellites, suggesting that satellite proteins play a role in trafficking CP110 to the centrosome (Hori et al., 2015). CP110 has been shown to interact with the satellite proteins PCM1 and CEP290 (Gupta et al., 2015, Tsang et al., 2008) and its interaction with CEP290 has been hypothesised to be part of the mechanism by which CP110 supresses ciliation (Tsang et al., 2008). Depletion of another satellite component, WDR8, impairs CP110 cap removal.
and ciliogenesis (Kurtulmus et al., 2015). Taken together, these data suggest that
alterations in satellite proteins in SDCCAG8-deficient cells may be the primary cause of
the ciliogenesis defect. Strikingly, a primary cilium formation defect has been reported
in cells derived from patients with SZ and bipolar disorder (Muñoz-Estrada et al., 2018),
although further research is needed to identify the mechanism by which this defect
occurs.

SDCCAG8 deficient cells do not lose their ability to ciliate completely, although the cilia
that do form demonstrate a deficit in SHh signalling. Interestingly, a pathway-wide
association study found Hh signalling to be one of the five top pathways associated with
SZ (Liu et al., 2017). The Hh pathway plays an important role in brain development and
the most characterized ligand, SHh, has been shown to regulate axonal guidance (Hor et
al., 2010). A defect in migration was observed in SH-SY5Y cells in response to a serum
gradient, consistent with previous findings in mice (Insolera et al., 2014) and with the
impact of SDCCAG8 deficiency on gene expression, which will be discussed in detail in
the next chapter. Neuronal migration is a key feature of brain development. In some
migrating neurons, the centrosome is positioned ahead of the nucleus. The organelle
moves into the leading process first, followed rapidly by translocation of the nucleus into
the leading process, suggesting that it may contribute to the forward movement of the
nucleus (Tsai et al., 2007, Solecki et al., 2004, Xie et al., 2003). Cells derived from
patients with SZ have been shown to have impairments in migration in multiple studies
(Tee et al., 2016, 2017, Sei et al., 2007 and Fan et al., 2013), suggesting that cell motility
may be important in the pathogenesis of the disorder.

Cells lacking SDCCAG8 is this study failed to undergo the morphological changes which
are the hallmark of differentiated cells, extensive and elongated neuron projections, and
instead remained very flat, with few projections. This demonstrated a marked deficiency
in the ability of SDCCAG8-deficient cells to differentiate. This finding is consistent with
data that demonstrates the importance of the centrosome in differential neurite formation
(Andersen and Halloran, 2012, de Anda et al., 2005, Zmuda and Rivas, 1998) and with
the impact of SDCCAG8 deficiency on gene expression. Induced pluripotent stem cell
based SZ case and control studies have previously demonstrated a delayed hippocampal differentiation of patient derived cells (Ahmad et al., 2018). Suggesting that neuronal differentiation may be impaired in SZ.

Taken together, these results suggest that disruption to SDCCAG8 may contribute to SZ pathophysiology and variance in cognition through defective ciliogenesis, disrupted SHh signalling and cell migration and differentiation impairments.
5 Transcriptional dysregulation of genes associated with SZ, IQ and EA in SDCCAG8-deficient cells.

5.1 Introduction

Primary cilia serve as a signalling hub in human cells and a large number of signalling pathways are co-ordinated through the primary cilium including Wnt, Notch and Hedgehog (Hh). Many of these pathways mediate their effects via altered gene expression (Wheway et al., 2018). Given that ciliation capacity is reduced in cells that lack SDCCAG8 and that the residual cilia display defective cilium-dependent signaling, I hypothesised that gene expression would be dysregulated in SDCCAG8-deficient cells.

To identify the transcriptional consequence of SDCCAG8 deficiency, I performed RNA-Seq; an innovative tool which quantifies the abundance of all mRNAs. There are a number of advantages that RNA-Seq has over other transcriptional profiling techniques. It is not limited to detecting transcripts that are previously characterized and there is dynamic range within transcript abundances that can be detected where transcripts that are either lowly expressed or very abundant can all be profiled, and the method has been shown to be highly reproducible (Wang et al., 2009b).

Using RNA-Seq, I compared SDCCAG8-deficient RPE1 cells to wild-type RPE1 cells under baseline conditions and following 48 hours serum starvation, which induces ciliogenesis. Three replicates were performed for each condition. This experiment was repeated using the same conditions in the immature neuronal cell line SH-SY5Y cells. I performed the RNA extractions and library preps for all experiments (further details available in section 2.9). DNA sequencing was performed by the Genomics Core Technology Unit at Queen’s University Belfast, who also aligned all reads to the genome, aggregated the reads by gene and assessed the quality of the data generated.

The aims of the study were: (i) to use RNA-Seq to identify the genes and pathways that are dysregulated as a consequence of SDCCAG8 deficiency and (ii) to investigate whether pathways affected by SDCCAG8 deficiency are enriched for genes associated with SZ and cognition.
5.2 Quality assessment of RPE1 RNA-Seq data

The quality of the data generated from RNA-Seq analysis of RPE1 cells was assessed using MultiQC (https://multiqc.info/), a tool that aggregates results from bioinformatics analyses across many samples into a single report. Each nucleotide that is called during the sequencing of DNA is assigned a Phred score, which is a measure of the quality of the identification of that nucleotide. The DNA reads generated from sequencing the library prep generated from RPE1 cells were of very high quality: the mean Phred score was above 30 at every position of the 75 bp reads that were generated in every sample that was sequenced (Figure 5.1 A). This means that the mean base call accuracy was above 99.9% for all generated reads in each sample, and the probability of an incorrect base call is 1 in 1,000. Phred scores are logarithmically linked to error probability so a decrease in Phred score quality to 20 would increase the probability of an incorrect base call to 1 in 100.

The MultiQC tool also assessed the alignment scores, which are generated when aligning sequenced DNA reads back to the human genome. The alignment scores show that the majority of reads generated during sequencing mapped back to genes (Figure 5.2 B), which would be expected in an RNA-Seq experiment as mRNA is purified during the library preparation. A small proportion of reads in each sample were aligned to multiple regions of the genome. These reads were removed from all further analysis and only reads that mapped unambiguously were retained in order to avoid increasing the read counts for multiple genes erroneously.

Based on the MultiQC report generated by the Genomics Core Technology Unit, I was satisfied that the sequencing data generated were of high quality and I proceeded to analyse them further.
Figure 5.1: Data generated in RNA-Seq experiments of RPE1 cells is of good quality

A. Sequence quality histogram for RPE1 RNA-Seq data, generated using MultiQC (https://multiqc.info/). Each line represents a sample analysed, to enable multiple samples to be plotted on the same graph, only the mean quality scores are plotted. The y-axis on the graph shows the quality scores. The higher the score, the better the base call. The background of the graph divides the y axis into very good quality calls (green), calls of reasonable quality (orange), and calls of poor quality (red). The x-axis on the graph shows base pair (bp) position, from 5’ to 3’.

B. Alignment scores for RPE1 RNA-Seq data, generated using MultiQC. For each sample sequenced, a percentage of reads that mapped to genes (Overlapping Genes), intergenic regions (No Feature), reads that overlap multiple genes (Ambiguous Features), mapped to multiple loci (Multimapping) or were Unmapped in the alignment of the reads to the human genome using STAR is presented. (The sequencing and this assessment of sequencing quality was performed by the Genomics Core Technology Unit at Queen’s University Belfast.)
5.3 Differentially expressed genes in SDCCAG8-deficient RPE1 cells

Raw HTSeq-count data was imported into R and normalised using DESeq2. In order to visualise the variation present between the samples that were analysed in this part of the study, I performed a principal component analysis (PCA). PCA plots data based on the directions in the dataset that explain the most variance. The PCA analysis is visualised in Figure 5.2 and shows that the principal component that explains the most variance in the data is the genotype of the cells, accounting for 66% of all the variance in the samples analysed. The second principal component is the condition, which was whether or not the cells had been serum starved. This component accounts for 31% of the overall variance in the data analysed. Replicates cluster closely together, as seen on the PCA plot, which indicates that there is a low level of variation between these samples and that they are similar to one another.

**Figure 5.2: Principal component analysis of RPE1 expression data**

Principal component analysis performed in R using variance stabilized transformed read count data from RPE1 wild-type cells (WT), and SDCCAG8 deficient RPE1 cells (KO), under baseline (Bas) and serum starved (SS) conditions. The plot was generated using R and demonstrates the directions that explain the most variance in the samples analysed. The percentage of variance explained by each principal component (PC) is labelled on each axis and the legend is displayed to the right.
I performed differential expression analysis using DESeq2, comparing WT RPE1 cells to SDCCAG8 deficient cells, both before and after serum starvation. This analysis revealed an extensive list of genes that were differentially expressed under each condition, the complete results are presented in Supplementary Tables 5.1 and 5.2. Due to the large number of SDCCAG8 dysregulated genes, I decided to focus on those protein-coding genes most dysregulated by the absence of SDCCAG8, which I defined as genes that had an absolute log$_2$ fold change ≥3 and a Benjamini-Hochberg adjusted p-value ≤0.05. Using this quite stringent threshold, I identified 585 and 604 differentially expressed genes (DEGs) in RPE1 cells under baseline conditions and after serum starvation respectively. The ten genes with the greatest log$_2$ fold change in each condition are presented in Table 5.1.

**Table 5.1: Differentially expressed genes in RPE1 cells depleted of SDCCAG8**

<table>
<thead>
<tr>
<th>Condition</th>
<th>Gene</th>
<th>Log$_2$ Fold Change</th>
<th>Adjusted P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>RCAN2</td>
<td>14.21</td>
<td>7.01-$^{32}$</td>
</tr>
<tr>
<td></td>
<td>SULF1</td>
<td>-12.70</td>
<td>1.42-$^{25}$</td>
</tr>
<tr>
<td></td>
<td>COL11A1</td>
<td>-12.69</td>
<td>1.64-$^{25}$</td>
</tr>
<tr>
<td></td>
<td>PCDH7</td>
<td>-12.11</td>
<td>2.42-$^{23}$</td>
</tr>
<tr>
<td></td>
<td>STXBP6</td>
<td>-11.08</td>
<td>1.82-$^{19}$</td>
</tr>
<tr>
<td></td>
<td>AQP4</td>
<td>10.76</td>
<td>1.68-$^{18}$</td>
</tr>
<tr>
<td></td>
<td>TRAML1L</td>
<td>-10.73</td>
<td>3.09-$^{18}$</td>
</tr>
<tr>
<td></td>
<td>TIMP3</td>
<td>-10.58</td>
<td>1.34-$^{23}$</td>
</tr>
<tr>
<td></td>
<td>PRR26</td>
<td>-10.39</td>
<td>4.33-$^{17}$</td>
</tr>
<tr>
<td></td>
<td>LRRK2</td>
<td>-10.15</td>
<td>2.92-$^{16}$</td>
</tr>
<tr>
<td>Serum-starved</td>
<td>SULF1</td>
<td>-14.70</td>
<td>1.79-$^{24}$</td>
</tr>
<tr>
<td></td>
<td>COL11A1</td>
<td>-14.06</td>
<td>9.94-$^{21}$</td>
</tr>
<tr>
<td></td>
<td>TIMP53</td>
<td>-13.62</td>
<td>1.83-$^{19}$</td>
</tr>
<tr>
<td></td>
<td>ZNF385D</td>
<td>-11.95</td>
<td>5.3-$^{15}$</td>
</tr>
<tr>
<td></td>
<td>AQP4</td>
<td>11.59</td>
<td>1.85-$^{20}$</td>
</tr>
<tr>
<td></td>
<td>IQCA1</td>
<td>-11.38</td>
<td>1.26-$^{13}$</td>
</tr>
<tr>
<td></td>
<td>RCAN2</td>
<td>11.36</td>
<td>8.88-$^{31}$</td>
</tr>
<tr>
<td></td>
<td>POTEC</td>
<td>11.18</td>
<td>4.81-$^{19}$</td>
</tr>
<tr>
<td></td>
<td>PCDH7</td>
<td>-11.16</td>
<td>1.13-$^{49}$</td>
</tr>
<tr>
<td></td>
<td>COL4A3</td>
<td>-10.96</td>
<td>1.05-$^{24}$</td>
</tr>
</tbody>
</table>

Notes: Partial output of analysis using DESeq2, comparing SDCCAG8 deficient cells to RPE1 WT cells at baseline conditions and after serum starvation. Data sorted by absolute log$_2$ fold change. Adjusted P-value; Wald test P-value after Benjamin-Hochberg correction.
5.4 Pathway analysis of DEGs in SDCCAG8 deficient RPE1 cells

In order to understand my experimental results within the context of biological systems, I used Ingenuity Pathway Analysis (IPA) to perform canonical pathway enrichment analysis. The ten most significantly affected pathways in SDCCAG8 deficient RPE1 cells under baseline conditions and after serum starvation are visualised in Figure 5.3 and Figure 5.4 respectively. Many of the identified pathways are shared amongst the two conditions tested here and details on all significant pathways can be found Supplementary Tables 5.3 and 5.4.

Figure 5.3: Canonical pathways enriched for SDCCAG8 dysregulated genes in RPE1 cells under baseline conditions

Output from IPA analysis, showing the ten pathways with the strongest enrichment for DEGs in SDCCAG8 deficient RPE1 cells at baseline conditions. Bar charts represent the percentage of overlap between SDDCAG8 dysregulated genes with each pathway, axis is situated at the top of the graph. Figures at the end of the bar chart show the number of genes within each pathway. The axis at the bottom of the graph denotes -log (p-value), a measure of the significance of the enrichment of pathways in the list of SDDCAG8 dysregulated genes.
Chapter 5

**Figure 5.4: Canonical pathways enriched for SDCCAG8 dysregulated genes in RPE1 cells following serum starvation**

Output from IPA analysis, showing the ten pathways with the strongest enrichment for DEGs in SDCCAG8 deficient RPE1 cells after serum starvation. Bar charts represent the percentage of overlap between SDDCAG8 dysregulated genes with each pathway, axis is situated at the top of the graph. Figures at the end of the bar chart show the number of genes within each pathway. The axis at the bottom of the graph denotes $-\log(p\text{-value})$, a measure of the significance of the enrichment of pathways in the list of SDDCAG8 dysregulated genes.

In addition to the IPA analysis, I performed gene ontology (GO) enrichment analysis of the DEGs to identify the molecular functions, biological pathways and cellular components that may be disrupted as a consequence of SDCCAG8 deficiency. The most enriched GO terms for DEGs from the SDCCAG8-deficient RPE1 cells, for both baseline conditions and after serum starvation, related to processes involved in extracellular matrix composition and structure. The ten most significantly associated pathways under baseline conditions and after serum starvation are presented in Table 5.2 and Table 5.3 respectively, and details on all GO terms with a p-value of less than 0.01 are listed in Supplementary Tables 5.5 and 5.6.
Table 5.2: GO terms that are enriched for DEGs in SDCCAG8 deficient RPE1 cells at baseline conditions

<table>
<thead>
<tr>
<th>p-value</th>
<th>q-value</th>
<th>GO term ID</th>
<th>GO category</th>
<th>GO level</th>
<th>GO term name</th>
<th>No. of genes in GO term</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.42E-12</td>
<td>1.04E-09</td>
<td>GO:0048731</td>
<td>B</td>
<td>3</td>
<td>System development</td>
<td>4713</td>
</tr>
<tr>
<td>1.79E-11</td>
<td>3.84E-09</td>
<td>GO:0009887</td>
<td>B</td>
<td>3</td>
<td>Animal organ morphogenesis</td>
<td>981</td>
</tr>
<tr>
<td>7.94E-11</td>
<td>6.2E-08</td>
<td>GO:0001501</td>
<td>B</td>
<td>4</td>
<td>Skeletal system development</td>
<td>518</td>
</tr>
<tr>
<td>9.83E-11</td>
<td>1.41E-08</td>
<td>GO:0043062</td>
<td>B</td>
<td>3</td>
<td>Extracellular structure organization</td>
<td>404</td>
</tr>
<tr>
<td>1.63E-10</td>
<td>1.85E-08</td>
<td>GO:0062023</td>
<td>C</td>
<td>3</td>
<td>Collagen-containing extracellular matrix</td>
<td>374</td>
</tr>
<tr>
<td>3.88E-10</td>
<td>1.51E-07</td>
<td>GO:0030198</td>
<td>B</td>
<td>4</td>
<td>Extracellular matrix organization</td>
<td>349</td>
</tr>
<tr>
<td>9.39E-10</td>
<td>2.45E-07</td>
<td>GO:0048705</td>
<td>B</td>
<td>4</td>
<td>Skeletal system morphogenesis</td>
<td>239</td>
</tr>
<tr>
<td>1.8E-09</td>
<td>1.93E-07</td>
<td>GO:0048513</td>
<td>B</td>
<td>3</td>
<td>Animal organ development</td>
<td>3452</td>
</tr>
<tr>
<td>1.1E-08</td>
<td>9.45E-07</td>
<td>GO:0030154</td>
<td>B</td>
<td>3</td>
<td>Cell differentiation</td>
<td>4089</td>
</tr>
<tr>
<td>9.43E-08</td>
<td>5.33E-06</td>
<td>GO:0031226</td>
<td>C</td>
<td>3</td>
<td>Intrinsic component of plasma membrane</td>
<td>1712</td>
</tr>
<tr>
<td>1.06E-07</td>
<td>2.06E-05</td>
<td>GO:0060349</td>
<td>B</td>
<td>4</td>
<td>Bone morphogenesis</td>
<td>112</td>
</tr>
</tbody>
</table>

Notes: Output from consensus DB (http://cpdb.molgen.mpg.de/) over-representation analysis of DEGs from RPE1 cells deficient in SDCCAG8 at baseline conditions with an absolute log2 fold change ≥3 and adjusted p-value ≤0.05. Data sorted by q-value. GO category: M; molecular function, C; Cellular Component, B; Biological Process. GO levels 3-5 were analysed in this study.
Table 5.3: GO terms that are enriched for DEGs in SDCCAG8 deficient RPE1 cells after serum starvation

<table>
<thead>
<tr>
<th>p-value</th>
<th>q-value</th>
<th>GO term ID</th>
<th>GO category</th>
<th>GO level</th>
<th>GO term name</th>
<th>No. of genes in GO term</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.93E-19</td>
<td>1.35E-16</td>
<td>GO:0048731</td>
<td>B</td>
<td>3</td>
<td>System development</td>
<td>4713</td>
</tr>
<tr>
<td>3.98E-18</td>
<td>9.18E-16</td>
<td>GO:0009887</td>
<td>B</td>
<td>3</td>
<td>Animal organ morphogenesis</td>
<td>981</td>
</tr>
<tr>
<td>5.42E-17</td>
<td>8.35E-15</td>
<td>GO:0051239</td>
<td>B</td>
<td>3</td>
<td>Regulation of multicellular organismal process</td>
<td>2989</td>
</tr>
<tr>
<td>2.84E-16</td>
<td>3.28E-14</td>
<td>GO:0048513</td>
<td>B</td>
<td>3</td>
<td>Animal organ development</td>
<td>3452</td>
</tr>
<tr>
<td>7.29E-14</td>
<td>8.46E-12</td>
<td>GO:0062023</td>
<td>C</td>
<td>3</td>
<td>Collagen-containing extracellular matrix</td>
<td>374</td>
</tr>
<tr>
<td>1.76E-13</td>
<td>1.63E-11</td>
<td>GO:0030154</td>
<td>B</td>
<td>3</td>
<td>Cell differentiation</td>
<td>4089</td>
</tr>
<tr>
<td>1.39E-12</td>
<td>1.18E-09</td>
<td>GO:0001501</td>
<td>B</td>
<td>4</td>
<td>Skeletal system development</td>
<td>518</td>
</tr>
<tr>
<td>3.04E-12</td>
<td>1.29E-09</td>
<td>GO:0030198</td>
<td>B</td>
<td>4</td>
<td>Extracellular matrix organization</td>
<td>349</td>
</tr>
<tr>
<td>4.08E-12</td>
<td>3.14E-10</td>
<td>GO:0043062</td>
<td>B</td>
<td>3</td>
<td>Extracellular structure organization</td>
<td>404</td>
</tr>
<tr>
<td>3.6E-11</td>
<td>2.09E-09</td>
<td>GO:0009888</td>
<td>B</td>
<td>3</td>
<td>Tissue development</td>
<td>1954</td>
</tr>
</tbody>
</table>

Notes: Output from consensus DB (http://cpdb.molgen.mpg.de/) over-representation analysis of DEGs from RPE1 cells deficient in SDCCAG8 after serum starvation, with an absolute log2 fold change ≥3 and adjusted p-value ≤0.05. Data sorted by q-value. GO category: M; molecular function, C; Cellular Component, B; Biological Process. GO levels 3-5 were analysed in this study.
5.5 Generation and quality assessment of SH-SY5Y RNA-Seq data

I also assessed the transcriptional consequence of SDCCAG8 deficiency in SH-SY5Y cells, an immature neuronal cell line. As described previously DNA sequencing was performed by the Genomics Core Technology Unit at Queen’s University Belfast, as well as genome alignment, read aggregation and data quality assessment. The DNA reads generated from sequencing the library prep generated from SH-SY5Y cells were of similarly high quality, the mean phread score was above 30 at every position of the 75 bp reads that were generated in every sample that was sequenced (Figure 5.5 A). This means that the mean base call accuracy was above 99.9% for all generated reads in each sample.

The alignment scores show that the majority of reads generated during sequencing mapped back to genes (Figure 5.5 B), as expected. A small proportion of reads in each sample could have been aligned to multiple regions of the genome, these reads were removed from all further analysis and only reads that map unambiguously were retained, in order to avoid increasing the read counts for multiple genes erroneously.

Based on the MultiQC report generated by the Genomics Core Technology Unit, I was satisfied that the sequencing data generated was of high quality and I proceeded to analyse the data further.
Figure 5.5: SH-SY5Y RNA-Seq data is of high quality
A. Sequence quality histogram for SH-SY5Y RNA-Seq data, generated using MultiQC (https://multiqc.info/). Each line represents a sample analysed, to enable multiple samples to be plotted on the same graph, only the mean quality scores are plotted. The y-axis on the graph shows the quality scores. The higher the score, the better the base call. The background of the graph divides the y axis into very good quality calls (green), calls of reasonable quality (orange), and calls of poor quality (red). The x-axis on the graph shows base pair (bp) position, from 5’ to 3’.
B. Alignment scores for SH-SY5Y RNA-Seq data, generated using MultiQC. For each sample sequenced, a percentage of reads that mapped to genes (Overlapping Genes), intergenic regions (No Feature), reads that overlap multiple genes (Ambiguous Features), mapped to multiple loci (Multimapping) or were Unmapped in the alignment of the reads to the human genome using STAR is presented. (The sequencing and this assessment of sequencing quality was performed by the Genomics Core Technology Unit at Queen’s University Belfast.)
5.6 Differentially expressed genes in SDCCAG8 deficient SH-SY5Y cells

As described previously, raw HTSeq-count data was imported into R and normalised using DESeq2. The variance in the data was assessed using PCA, which demonstrates that the principal component that explains the most variance in the data is the genotype of the cells, accounting for 61% of all the variance in the samples analysed (Figure 5.6). The second principal component is the condition, which was whether or not the cells had been serum starved this component accounts for 35% of the overall variance in the data analysed. Replicates cluster closely together, as seen on the PCA plot, which indicates that there is a low level of variation between these samples and that they are highly similar to one another.

![Figure 5.6: Principal component analysis of SH-SY5Y expression data](image)

Principal component analysis performed in R using variance stabilized transformed read count data from SH-SY5Y wild-type cells (WT), and SDCCAG8 deficient SH-SY5Y cells (KO), under baseline (Bas) and serum starved (SS) conditions. Plot was generated using gplot package in R and demonstrates the directions that explain the most variance in the samples analysed. The percentage of variance explained by each principal component is indicated on each axis and the legend for the sample colouring is displayed to the right.
I performed differential expression analysis using DESeq2, comparing wild-type SH-SY5Y cells to SDCCAG8-deficient cells, both before and after serum starvation. Again I decided to focus on those protein-coding genes most dysregulated by the absence of SDCCAG8, which I defined as genes that had an absolute log$_2$ fold change of three or greater and a Benjamini-Hochberg adjusted p-value of 0.05 or less. There are 585 and 604 DEGs in SDCCAG8-deficient SH-SY5Y cells under baseline conditions and after serum starvation respectively at this threshold, however full results are available Supplementary Tables 5.7 and 5.8. The ten genes with the greatest absolute log$_2$ fold change in each condition are presented in Table 5.4.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Gene</th>
<th>Log$_2$ Fold Change</th>
<th>Adjusted P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>CELF2</td>
<td>-12.49</td>
<td>3.46$^{24}$</td>
</tr>
<tr>
<td></td>
<td>MGST1</td>
<td>-12.41</td>
<td>7.46$^{24}$</td>
</tr>
<tr>
<td></td>
<td>SLC6A11</td>
<td>-12.11</td>
<td>1.08$^{22}$</td>
</tr>
<tr>
<td></td>
<td>CDH4</td>
<td>-12.04</td>
<td>2.18$^{22}$</td>
</tr>
<tr>
<td></td>
<td>PPP2R2C</td>
<td>-10.50</td>
<td>1.80$^{44}$</td>
</tr>
<tr>
<td></td>
<td>CACNG2</td>
<td>-10.07</td>
<td>2.25$^{15}$</td>
</tr>
<tr>
<td></td>
<td>VIP</td>
<td>-10.07</td>
<td>1.25$^{15}$</td>
</tr>
<tr>
<td></td>
<td>HPDL</td>
<td>-9.44</td>
<td>6.56$^{13}$</td>
</tr>
<tr>
<td></td>
<td>SERPINA5</td>
<td>-9.01</td>
<td>1.87$^{77}$</td>
</tr>
<tr>
<td></td>
<td>BMP7</td>
<td>-8.84</td>
<td>2.00$^{108}$</td>
</tr>
<tr>
<td>Serum-starved</td>
<td>SLC6A11</td>
<td>-12.55</td>
<td>6.32$^{25}$</td>
</tr>
<tr>
<td></td>
<td>PPP2R2C</td>
<td>-11.75</td>
<td>2.85$^{30}$</td>
</tr>
<tr>
<td></td>
<td>FAM135B</td>
<td>-10.87</td>
<td>1.29$^{18}$</td>
</tr>
<tr>
<td></td>
<td>MGST1</td>
<td>-10.83</td>
<td>1.37$^{18}$</td>
</tr>
<tr>
<td></td>
<td>SERPINA5</td>
<td>-10.45</td>
<td>1.23$^{55}$</td>
</tr>
<tr>
<td></td>
<td>BASP1</td>
<td>-10.44</td>
<td>5.91$^{34}$</td>
</tr>
<tr>
<td></td>
<td>KCNA3</td>
<td>-9.99</td>
<td>1.11$^{15}$</td>
</tr>
<tr>
<td></td>
<td>ADAMTS2</td>
<td>-9.73</td>
<td>6.34$^{15}$</td>
</tr>
<tr>
<td></td>
<td>HMX1</td>
<td>-9.69</td>
<td>1.00$^{14}$</td>
</tr>
<tr>
<td></td>
<td>CRTAC1</td>
<td>-9.69</td>
<td>9.76$^{15}$</td>
</tr>
</tbody>
</table>

Notes: Outputs of analysis using DESeq2, comparing SDCCAG8 deficient cells to SH-SY5Y WT cells at baseline conditions and after serum starvation. Data sorted by absolute log$_2$ fold change. Adjusted P-value; Wald test P-value after Benjamin-Hochberg correction.
5.7 Pathway Analysis of DEGs in SDCCAG8 deficient SH-SY5Y cells

To understand my experimental results within the context of biological systems, I used Ingenuity Pathway Analysis (IPA) to perform canonical pathway enrichment analysis. The ten most significantly affected pathways in SDCCAG8 deficient SH-SY5Y cells under baseline conditions and after serum starvation are visualised in Figures 5.7 and 5.8, respectively. Many of the identified pathways are shared amongst the two conditions tested here and details on all significant pathways can be found in Supplementary Tables 5.9 and 5.10.

**Figure 5.7: Canonical pathways enriched for DEGS in SDCCAG8 deficient SH-SY5Y cells at baseline conditions**

Output from IPA analysis showing the ten pathways with the strongest enrichment for DEGs in SDCCAG8-deficient SH-SY5Y cells under baseline conditions. Bar charts represent the percentage of overlap between SDCCAG8 dysregulated genes with each pathway, with the axis situated at the top of the graph. Figures at the end of the bar chart show the number of genes within each pathway. The axis at the bottom of the graph denotes $-\log(p\text{-value})$, a measure of the significance of the enrichment of pathways in the list of SDCCAG8-dysregulated genes.
Chapter 5

Figure 5.8: Canonical pathways enriched for DEGs in SDCCAG8 deficient SH-SY5Y cells after serum starvation

Output from IPA analysis, showing the ten pathways with the strongest enrichment for DEGs in SDCCAG8 deficient SH-SY5Y cells after serum starvation. Bar charts represent the percentage of overlap between SDCCAG8 dysregulated genes with each pathway, axis is situated at the top of the graph. Figures at the end of the bar chart show the number of genes within each pathway. The axis at the bottom of the graph denotes -\log(p\text{-value}) , a measure of the significance of the enrichment of pathways in the list of SDCCAG8 dysregulated genes.

In addition to the IPA analysis, I performed GO enrichment analysis of the DEGs to identify the molecular functions, biological pathways and cellular components that may be disrupted as a consequence of SDCCAG8 deficiency. The most enriched GO terms for DEGs from the SDCCAG8-deficient SH-SY5Y cells related to synaptic signalling as well as nervous system development, although terms related to the cell-cycle progression were also enriched in the GO analysis of DEGs from the serum-starved SHSY5Y cells. The ten most significantly dysregulated GO terms under baseline conditions and after serum starvation are presented in Tables 5.5 and 5.6, respectively, and details on all GO terms with a p-value of less than 0.01 are listed in Supplementary Tables 5.11 and 5.12.
Table 5.5: GO terms that are enriched for DEGs in SDCCAG8 deficient SH-SY5Y cells at baseline conditions

<table>
<thead>
<tr>
<th>p-value</th>
<th>q-value</th>
<th>GO term ID</th>
<th>GO category</th>
<th>GO level</th>
<th>GO term name</th>
<th>No. of genes in GO term</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.68E-21</td>
<td>9.32E-19</td>
<td>GO:009953</td>
<td>B</td>
<td>4</td>
<td>Trans-synaptic signaling</td>
<td>675</td>
</tr>
<tr>
<td>1.79E-21</td>
<td>5.79E-19</td>
<td>GO:009953</td>
<td>B</td>
<td>3</td>
<td>Synaptic signaling</td>
<td>676</td>
</tr>
<tr>
<td>9.27E-21</td>
<td>3.94E-18</td>
<td>GO:009891</td>
<td>B</td>
<td>5</td>
<td>Anterograde trans-synaptic signaling</td>
<td>672</td>
</tr>
<tr>
<td>1.08E-16</td>
<td>8.88E-15</td>
<td>GO:004300</td>
<td>C</td>
<td>4</td>
<td>Neuron projection</td>
<td>1228</td>
</tr>
<tr>
<td>1.91E-15</td>
<td>1.74E-13</td>
<td>GO:003647</td>
<td>C</td>
<td>3</td>
<td>Somatodendritic compartment</td>
<td>769</td>
</tr>
<tr>
<td>7.53E-14</td>
<td>2.17E-12</td>
<td>GO:003042</td>
<td>C</td>
<td>4</td>
<td>Dendrite</td>
<td>572</td>
</tr>
<tr>
<td>7.93E-14</td>
<td>2.17E-12</td>
<td>GO:009744</td>
<td>C</td>
<td>4</td>
<td>Dendritic tree</td>
<td>573</td>
</tr>
<tr>
<td>1.39E-13</td>
<td>3.86E-11</td>
<td>GO:000739</td>
<td>B</td>
<td>4</td>
<td>Nervous system development</td>
<td>2304</td>
</tr>
<tr>
<td>1.63E-11</td>
<td>3.46E-09</td>
<td>GO:004869</td>
<td>B</td>
<td>5</td>
<td>Generation of neurons</td>
<td>1442</td>
</tr>
<tr>
<td>2.86E-11</td>
<td>4.63E-09</td>
<td>GO:005080</td>
<td>B</td>
<td>3</td>
<td>Synapse organization</td>
<td>305</td>
</tr>
</tbody>
</table>

Notes: Output from consensus DB (http://cpdb.molgen.mpg.de/) over-representation analysis of DEGs from SH-SY5Y cells deficient in SDCCAG8 at baseline conditions, with an absolute log2-fold change $\geq 3$ and adjusted p-value $\leq 0.05$. Data sorted by q-value. GO category: M; molecular function, C; Cellular Component, B; Biological Process. GO levels 3–5 were analysed in this study.
Table 5.6: GO terms that are enriched for DEGs in SDCCAG8 deficient SH-SY5Y cells after serum starvation

<table>
<thead>
<tr>
<th>p-value</th>
<th>q-value</th>
<th>GO term ID</th>
<th>GO category</th>
<th>GO level</th>
<th>GO term name</th>
<th>No. of genes in GO term</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.47E-22</td>
<td>2.7E-19</td>
<td>GO:0099537</td>
<td>B</td>
<td>4</td>
<td>Trans-synaptic signaling</td>
<td>675</td>
</tr>
<tr>
<td>3.79E-22</td>
<td>1.68E-19</td>
<td>GO:0099536</td>
<td>B</td>
<td>3</td>
<td>Synaptic signaling</td>
<td>676</td>
</tr>
<tr>
<td>1.23E-21</td>
<td>8.09E-19</td>
<td>GO:0098916</td>
<td>B</td>
<td>5</td>
<td>Anterograde trans-synaptic signaling</td>
<td>672</td>
</tr>
<tr>
<td>2.04E-14</td>
<td>2.42E-12</td>
<td>GO:0043005</td>
<td>C</td>
<td>4</td>
<td>Neuron projection</td>
<td>1228</td>
</tr>
<tr>
<td>2.43E-14</td>
<td>9.41E-12</td>
<td>GO:0007399</td>
<td>B</td>
<td>4</td>
<td>Nervous system development</td>
<td>2304</td>
</tr>
<tr>
<td>1.38E-12</td>
<td>3.57E-10</td>
<td>GO:0099177</td>
<td>B</td>
<td>4</td>
<td>Regulation of trans-synaptic signaling</td>
<td>336</td>
</tr>
<tr>
<td>1.62E-12</td>
<td>3.6E-10</td>
<td>GO:0048731</td>
<td>B</td>
<td>3</td>
<td>System development</td>
<td>4667</td>
</tr>
<tr>
<td>2.36E-12</td>
<td>2.5E-10</td>
<td>GO:0036477</td>
<td>C</td>
<td>3</td>
<td>Somatodendritic compartment</td>
<td>769</td>
</tr>
<tr>
<td>2.54E-12</td>
<td>4.92E-10</td>
<td>GO:0000819</td>
<td>B</td>
<td>4</td>
<td>Sister chromatid segregation</td>
<td>245</td>
</tr>
<tr>
<td>3.64E-12</td>
<td>5.38E-10</td>
<td>GO:0098813</td>
<td>B</td>
<td>3</td>
<td>Nuclear chromosome segregation</td>
<td>313</td>
</tr>
</tbody>
</table>

Notes: Output from consensus DB (http://cpdb.molgen.mpg.de/) over-representation analysis of DEGs from SH-SY5Y cells deficient in SDCCAG8 after serum starvation, with an absolute log₂fold change ≥3 and adjusted p-value ≤0.05. Data sorted by q-value. GO category: M; molecular function, C; Cellular Component, B; Biological Process. GO levels 3-5 were analysed in this study.
5.8 Overlap of DEGs upon SDCCAG8 disruption

Across the four different conditions analysed in this study, *SDCCAG8* disruption resulted in the differential expression of 1,403 unique protein coding genes, at an absolute log$_2$ fold change of three or greater and Benjamini-Hochberg adjusted p-value of less than 0.05. The Venn diagram in Figure 5.9 shows the overlap of these genes between the four conditions; RPE1 cells at baseline conditions, RPE1 cells after 48h serum starvation, SH-SY5Y cells at baseline and SH-SY5Y cells after 48h serum starvation. Ten genes were differentially expressed under all the conditions analysed. These ten genes are listed below along with some details about their known functions and roles in cognition and neurodevelopment. Using the association test result for each gene from the MAGMA analysis (following correction for the number of genes tested), both *PHACTR1* and *ZFR2* are significantly associated with EA (Lee et al., 2018).

![Figure 5.9: Overlap of DEGs upon SDCCAG8 disruption](image)

Venn diagram demonstrating the overlap of DEGs upon SDCCAG8 disruption across the four conditions analysed in this study; RPE1 cells at baseline conditions (‘Bas’), RPE1 cells after 48h serum starvation (‘SS’), SH-SY5Y cells at baseline and SH-SY5Y cells after 48h serum starvation. All DEGs had an absolute log$_2$ fold change of 3 or greater and Benjamini-Hochberg adjusted p-value of less than 0.05. Listed are the gene symbols for the ten genes that are differentially expressed under all four conditions analysed.
ADARB2 (Adenosine Deaminase RNA Specific B2 (Inactive)) encodes a member of the double-stranded RNA adenosine deaminase family of RNA-editing enzymes. The protein has no detectable RNA-editing activity and instead is hypothesized to have a regulatory role in RNA editing (Chen et al., 2000). A SNP within the gene has been associated with cognitive decline in Alzheimer's disease (Lee et al., 2017) and the gene has been shown to be associated with temporal lobe volume in AD (Kohannim et al., 2012). RNA editing has recently been shown to be dysregulated in the brains of individuals diagnosed with autism (Tran et al., 2019).

GALNT14 (Polypeptide N-Acetylgalactosaminyltransferase 14) encodes a Golgi protein which is a member of the polypeptide N-acetylgalactosaminyltransferase protein family. These enzymes catalyze the transfer of N-acetyl-D-galactosamine (GalNAc) to the hydroxyl groups on serines and threonines in target peptides. The GALNT14 locus is a human accelerated region, the locus is highly different between humans and other species which may suggest the gene has a role in the evolution of human specific traits (Doan et al., 2016).

KCNQ3 (Potassium Voltage-Gated Channel Subfamily Q Member) encodes a protein that functions in the regulation of neuronal excitability. Loss of function mutations in the gene can result in benign familial neonatal seizures 2, a condition which results in seizures within the first few days of life (Ryan et al., 1991 and Charlier et al., 1998,). Gain of function variants in KCNQ3 have been shown to cause neurodevelopmental disability and autism spectrum disorder (Sands et al., 2019).

MARCH11 (Membrane Associated Ring-CH-Type Finger 11) is a member of the MARCH family of membrane-bound E3 ubiquitin ligases. These enzymes add ubiquitin to target lysines in substrate proteins, thereby signaling their intracellular transport. March11 appears to have a role in ubiquitin-mediated protein sorting in the trans-Golgi network (TGN)-multivesicular body (MVB) transport pathway (Morokuma et al., 2007).
MARCH11 contains a differentially methylated position in patients with SZ, compared to controls and this methylation is affected by atypical antipsychotic treatments (Montano et al., 2016).

PHACTR1 (Phosphatase And Actin Regulator 1) encodes a protein that can interact with actin and regulate cell morphology and motility. The protein plays a role in cortical migration and synaptic formation (Hamada et al., 2018). Mutations in the gene have been found in patients with early infantile epileptic encephalopathy-70, which have been deemed to be pathogenic based on molecular studies (de Ligt et al., 2012 and Hamada et al., 2018). These patients present with seizures and developmental delay and one patient was also diagnosed with ASD.

RRAGD (Ras-Related GTP Binding D) is a guanine nucleotide-binding protein. The gene is up-regulated in mouse genetic models of Lissencephaly (Lis1, Dcx, Ndel1 and Ywhae mutants). Lissencephaly is a neuronal migration defect in humans and causes mental retardation and intractable epilepsy (Pramparo et al., 2011).

SERPINA5 (Serpin Family A Member 5) encodes a glycoprotein that inhibits serine proteases. SERPINA5 was found in serum from patients with SZ but was absent from controls and patients with bipolar disorder (Smirnova et al., 2019). SERPINA5 also had a serum peak in children with ASD from China (Yang et al., 2018b).

THBS2 (Thrombospondin 2) encodes a disulfide-linked homotrimeric glycoprotein that mediates cell-to-cell and cell-to-matrix interactions. THBS2 is expressed by immature astrocytes and promotes synaptogenesis (Christopherson et al., 2005). Deletions and duplications of chromosome 6q27 cause extensive structural brain abnormalities; THBS2 has been suggested to be responsible for these phenotypes (Burnside et al., 2015, Peddibhotla et al., 2014).
UNC5D (Unc-5 Netrin Receptor D) encodes a cell surface netrin-1 receptor. Upon cell stress and absence of netrin-1, UNC5D translocates to the nucleus to initiate the activation of proapoptotic genes (Zhu et al., 2013). UNC5D has been implicated in ASD and the gene contained a higher burden of runs of homozygosity in patients with ASD than in controls (Gamsiz et al., 2013 and Casey et al., 2012). DISC1 mutations cause mental illness. iPSCs harbouring DISC1 mutations were differentiated and a reduction in UNC5D expression was observed. These neurons demonstrated a lack of neurite outgrowth, which was also observed in WT cells where UNC5D was depleted and rescued by re-expression of UNC5D. This suggest that a potential mechanism by which DISC1 mutations cause mental illness is by dysregulating UNC5D (Srikanth et al., 2018).

ZFR2 (Zinc Finger RNA Binding Protein 2), a brain-expressed chimaera of MATK and ZFR2, was found in patients with SZ and absent from controls and a hypothesised mechanism of pathogenicity was the altered expression of ZFR2 (Rippey et al., 2013). A CNV resulting in deletion of ZFR2 has been reported in an individual with ASD (Woodbury-Smith et al., 2015).

5.9 Gene-set analysis of GO terms enriched for SDCCAG8 dysregulated genes

I next investigated if the GO terms enriched for the DEGs from the RPE1 cells or the SH-SY5Y cells were also enriched for genes associated with SZ, IQ or EA using MAGMA gene-set analysis of GWAS data for these phenotypes. A gene-set analysis is a statistical method for simultaneously analysing GWAS results for multiple genes to test if variants in those genes jointly affect a phenotype. As there is considerable overlap between related GO terms, I used ancestral GO charts to identify the most enriched but non-overlapping GO terms from the analysis of DEGs (five GO terms from each cell line, highlighted in bold in Supplementary Tables 5.5, 5.6, 5.11 and 5.12). For RPE1 cells, those enriched GO terms were collagen-containing extracellular matrix, extracellular matrix organization, skeletal system morphogenesis, cell differentiation and regulation of multicellular organismal process. For SH-SY5Y cells, those enriched GO terms were trans-synaptic signalling, dendrite, generation of neurons, synapse organization and sister chromatid segregation. After correcting for multiple testing (0.05/30, P= 0.0017),
all SH-SY5Y GO terms except sister chromatid segregation were significantly enriched for genes associated with SZ, IQ and EA (Figures 5.10, 5.11 and 5.12). There were no significant enrichments for any RPE1 GO terms. Neuronally-expressed genes are a major contributor to the phenotypes tested here. It is possible that the enrichments detected for SZ, IQ and EA could be due to the SH-SY5Y GO terms representing a subset of neuronally-expressed genes. To control for this I also conditioned the analyses on ‘SH-SY5Y expressed genes’, which is a gene-set containing all genes for which expression was detected in wild-type (WT) SH-SY5Y cells in the RNA-Seq experiment. All enrichments remained significant in these conditional analyses (Supplementary Tables 5.13-5.15). As a further control, all significantly enriched SH-SY5Y GO terms were tested for enrichment in “control” GWAS datasets from non-neurodevelopmental diseases (Alzheimer’s disease (AD), Crohn’s disease (CD), stroke (STR), type 2 diabetes (T2D), coronary artery disease (CAD) and ulcerative colitis (UC)). As no significant results were detected (Table 5.7), I conclude that the GO term enrichments observed for SZ, IQ and EA are not a property of polygenic phenotypes in general.

**Figure 5.10:** Gene expression pathways impacted by SDCCAG8 deficiency affect EA

Gene-set analysis of the SH-SY5Y and RPE1 pathways in EA. GO terms are listed on the y-axis. P-values are shown above each data point, which represent beta-values (regression coefficients) plotted on the x-axis. Horizontal bars indicate standard error. Significant p-values after multiple test correction are indicated in bold.
Figure 5.11: Gene expression pathways impacted by SDCCAG8 deficiency affect IQ
Gene-set analysis of the SH-SY5Y and RPE1 pathways in EA. GO terms are listed on the y-axis. P-values are shown above each data point, which represent beta-values (regression coefficients) plotted on the x-axis. Horizontal bars indicate standard error. Significant p-values after multiple test correction are indicated in bold.

Figure 5.12: Gene expression pathways impacted by SDCCAG8 deficiency affect SZ
Gene-set analysis of the SH-SY5Y and RPE1 pathways in SZ. GO terms are listed on the y-axis. P-values are shown above each data point, which represent beta-values (regression coefficients) plotted on the x-axis. Horizontal bars indicate standard error. Significant p-values after multiple test correction are indicated in bold.
Table 5.7: Gene-set analysis of GO terms impacted by SDCCAG8 deficiency and associated with IQ, EA and SZ in control GWAS

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Trans-synaptic signaling</th>
<th>Dendrite</th>
<th>Generation of neurons</th>
<th>Synapse organization</th>
</tr>
</thead>
<tbody>
<tr>
<td>T2D</td>
<td>686</td>
<td>-0.0114</td>
<td>0.0318</td>
<td>0.6402</td>
</tr>
<tr>
<td>UC</td>
<td>686</td>
<td>0.0298</td>
<td>0.037</td>
<td>0.2104</td>
</tr>
<tr>
<td>STR</td>
<td>682</td>
<td>0.0371</td>
<td>0.0327</td>
<td>0.1288</td>
</tr>
<tr>
<td>CD</td>
<td>687</td>
<td>-0.0359</td>
<td>0.0389</td>
<td>0.822</td>
</tr>
<tr>
<td>CAD</td>
<td>681</td>
<td>0.0329</td>
<td>0.0342</td>
<td>0.1682</td>
</tr>
<tr>
<td>AD</td>
<td>686</td>
<td>0.0304</td>
<td>0.037</td>
<td>0.2055</td>
</tr>
</tbody>
</table>

Notes: Output from MAGMA analysis, Gene-sets that were significantly enriched for associations with SZ, IQ or EA were tested for associations in control GWAS. No. genes; number of genes in gene-set that were analysed, SE; standard error, T2D; type 2 diabetes, UC; ulcerative colitis, STR; stroke, CD; Crohn’s disease, CAD; coronary artery disease, AD; Alzheimer’s disease.
5.10 Discussion

The RNA-Seq analysis performed in this chapter reveals that a large number of genes are transcriptionally dysregulated in the absence of SDCCAG8. In order to facilitate the interpretation of this transcriptional dysregulation, I performed pathway analysis using IPA and GO term enrichment. The pathways that are enriched for SDCCAG8 dysregulated genes in RPE1 cells are related to processes involved in extracellular matrix (ECM) deposition and turnover. For example, hepatic fibrosis is a chronic liver disease associated with a pathological accumulation of ECM proteins (Baiocchini et al., 2016) and cAMP is a second messenger central to a number of biological processes that regulates ECM gene expression and metabolism (Rockel et al., 2009). Rheumatoid arthritis (RA) is a chronic inflammatory autoimmune disease and the extracellular matrix is hypothesised to play a key role in RA by promoting a pro-inflammatory niche (Ruhmann and Midwood, 2013). Matrix metalloproteinases (MMP) are the main ECM degradation enzymes and GP6 is a platelet membrane glycoprotein, a major signaling receptor for collagen, and can also activate lamin and fibrin among other ECM components (Ezumi et al., 2000). GO terms that were enriched for SDCCAG8 dysregulated genes in RPE1 cells are related to processes involved in extracellular matrix composition and structure including; extracellular structure organisation, collagen-containing ECM and ECM organisation. ECM dysregulation is also a clear trend present in the pathway analysis performed on SDCCAG8 dysregulated genes in SH-SY5Y cells. Hepatic fibrosis and cAMP signaling are again highly enriched in concordance with RPE1 data. Also enriched are heparan sulfate, chondratin sulfate and dermatan sulfate biosynthesis pathways. Heparan sulfate, chondratin sulfate and dermatan sulfate are negatively charged glycosaminoglycans that are often covalently linked to ECM proteins, forming proteoglycans. The glycosaminoglycan component of these proteins provides hydration and swelling pressure which enables tissue to withstand compressional forces (Yanagishita, 1993). These data taken together indicate that SDCCAG8 deficiency in human cells promotes ECM dysregulation.

The extracellular space in the human brain is estimated to occupy 20% of the brain’s volume (Lei et al., 2017). The brain ECM has a unique composition and is enriched with proteoglycans, glycoproteins, small linker proteins and enzymes that regulate ECM
deposition and degradation (Zimmermann and Dours-Zimmermann, 2008). Perineurinal nets (PNN) are specialized ECM structures intricately woven around neurons as well as other cell types. PNNs are important for the stabilisation of synapses and have been proposed as a key element that underlies long-term memory consolidation (Krishnaswamy et al., 2019). Numerous studies have showed a decrease in PNNs in patients with SZ, which is not associated with a decrease in cell number, indicating a perturbation of the ECM structure (Mauney et al., 2013, Pantazopoulos et al., 2010 and 2015). This disruption to PNNs is hypothesised to destabilise synaptic connectivity. Proteolytic ECM remodelling has been shown to play a role in synaptic plasticity, and is mediated in part by metalloproteases. One such metalloprotease, MMP-9, has been shown to be increased in blood samples of patients with SZ (Okulski et al., 2007, Domenici et al., 2010), and patients with treatment-resistant SZ (Yamamori et al., 2013).

This evidence is part of a growing body of literature reviewed here (Sethi and Zaia, 2017, Pantazopoulos and Berretta, 2016), which suggests that the brain ECM structure is dysregulated in psychiatric disorders. The primary cilia of many cells orient toward the ECM and receptors for a number of ECM components localise to the primary cilium (Veland et al., 2014). Patients with ciliopathies demonstrate extensive fibrosis of affected organs (Seeger-Nukpezah and Golemis, 2012). The primary cilium is therefore involved in sensing ECM components, responding to alterations in ECM components and potentially influencing deposition of newly synthesised ECM. It is possible that the disruption of ciliary signalling in SDCCAG8-deficient cells (discussed in chapter 4) may be responsible for the transcriptional dysregulation of ECM-related pathways.

A large proportion of canonical pathways that were enriched for SDCCAG8 dysregulated genes in SH-SY5Y cells were directly related to neuronal function. These include axonal guidance signaling, which is the process by which axon extensions migrate to reach their synaptic targets and therefore forming neuronal connections. Generating precise patterns of connectivity in the brain relies on the action of members of this pathway (Bashaw and Klein, 2010). Calcium is a key regulator of axonal guidance signaling (Gomez and Zheng, 2006) and calcium signalling was also an enriched canonical pathway. A very similar GO term, neuron projection, was also significantly enriched. Defects in axonal guidance signaling are associated with many conditions including dyslexia, autism spectrum disorder, epilepsy (Van Battum et al., 2015) and SZ (Wang et al., 2018a).
Primary cilia have been hypothesised to play a role in axon guidance signaling. The strongest evidence supporting this hypothesis is a defect in the directional navigation of interneurons upon disruption of Arl13b (Higginbotham et al., 2012). In addition, patients with Joubert syndrome, a ciliopathy with a number of causative genes that encode centrosome/ciliary proteins, have been shown to have some nerve fibre tracts that fail to decussate, suggesting an axon guidance defect in these patients (Poretti et al., 2007). The transcriptional dysregulation of neuronal projection pathways observed upon ablation of SDCCAG8 is consistent with the defect in migration in SDCCAG8-deficient cells discussed in section 4.12 and the differentiation defect discussed in 4.13.

Other enriched canonical pathways in this study directly related to neuronal function are GABA receptor and glutamate receptor signaling. This is consistent with the GO term enrichment in which synaptic signaling featured heavily. GABA is the chief inhibitory neurotransmitter in the central nervous system, while glutamate is by far the most abundant excitatory neurotransmitter. Coordination of these two systems is essential for normal functioning of the central nervous system (Foster and Kemp, 2006). Dysfunction of both signaling pathways have been implicated in the pathophysiology of SZ (Koshiyama et al., 2018) and ASD (Horder et al., 2018). Previous studies have demonstrated that primary cilia can influence the connectivity and functioning of both glutamatergic (Kumamoto et al., 2012) and GABAergic neurons (Guo et al., 2017).

A subset of highly enriched GO terms were taken forward for gene-set enrichment analysis of GWAS data for the phenotypes of interest; SZ, IQ and EA. After correcting for multiple testing, all SH-SY5Y GO terms except sister chromatid segregation were significantly enriched for genes associated with SZ, IQ and EA. These analyses indicate that the specific biological pathways and cellular components that are dysregulated in immature neurons upon depletion of SDCCAG8 contain genes that contribute to SZ risk and cognitive function in the general population. Similar pathways were not affected in the RPE1 cells, supporting the cell-specific nature of the GO terms identified by the gene expression analysis.
In addition to harbouring common genetic variants that are associated with SZ and cognition, many SDCCAG8 dysregulated pathways have been implicated in SZ pathophysiology and normal brain development previously, as discussed above. Importantly, many of these pathways have also been shown to be regulated at least in part by primary cilia. The link between cell migration, cell differentiation and the ciliary control of such activities through the regulation of gene expression that is evidenced by my analyses of SDCCAG8 provides a model for how centrosomal genes may contribute to SZ risk and cognitive function.
### 5.11 Supplementary data

Supplementary Tables 5.1-5.12 are available on the attached CD. Supplementary Tables 5.13-5.15 are presented below.

**Supplementary Table 5.13: Gene-set analysis of GO terms impacted by SDCCAG8 deficiency in SZ**

<table>
<thead>
<tr>
<th>Source</th>
<th>Gene-set</th>
<th>No. of Genes</th>
<th>Beta</th>
<th>SE</th>
<th>P-value</th>
<th>Conditioned on 'SH-SY5Y expressed genes'</th>
<th>Beta</th>
<th>SE</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>RPE1</td>
<td>Collagen-containing extracellular matrix</td>
<td>397</td>
<td>-0.11</td>
<td>0.0616</td>
<td>0.96292</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPE1</td>
<td>extracellular matrix organization</td>
<td>350</td>
<td>-0.0766</td>
<td>0.0562</td>
<td>0.91338</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPE1</td>
<td>skeletal system morphogenesis</td>
<td>238</td>
<td>0.103</td>
<td>0.0709</td>
<td>0.072841</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPE1</td>
<td>cell differentiation</td>
<td>3935</td>
<td>0.0348</td>
<td>0.0195</td>
<td>0.037692</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPE1</td>
<td>regulation of multicellular organismal process</td>
<td>2969</td>
<td>0.0514</td>
<td>0.0215</td>
<td>0.008415</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SH-SY5Y</td>
<td>trans-synaptic signaling</td>
<td>695</td>
<td>0.175</td>
<td>0.0407</td>
<td>8.45E-06</td>
<td>0.16</td>
<td>0.0459</td>
<td>0.00024221</td>
<td></td>
</tr>
<tr>
<td>SH-SY5Y</td>
<td>dendrite</td>
<td>609</td>
<td>0.272</td>
<td>0.0446</td>
<td>5.62E-10</td>
<td>0.311</td>
<td>0.0501</td>
<td>2.763E-10</td>
<td></td>
</tr>
<tr>
<td>SH-SY5Y</td>
<td>generation of neurons</td>
<td>1473</td>
<td>0.145</td>
<td>0.0292</td>
<td>3.19E-07</td>
<td>0.145</td>
<td>0.0326</td>
<td>0.000004407</td>
<td></td>
</tr>
<tr>
<td>SH-SY5Y</td>
<td>synapse organization</td>
<td>411</td>
<td>0.169</td>
<td>0.0537</td>
<td>0.00834</td>
<td>0.207</td>
<td>0.061</td>
<td>0.00034098</td>
<td></td>
</tr>
<tr>
<td>SH-SY5Y</td>
<td>sister chromatid segregation</td>
<td>187</td>
<td>0.0652</td>
<td>0.0774</td>
<td>0.19666</td>
<td>0.0803</td>
<td>0.0872</td>
<td>0.17879</td>
<td></td>
</tr>
</tbody>
</table>

Notes: Output from MAGMA analysis, a subset of highly enriched GO terms in Supplementary Tables 5.5, 5.6, 5.11 and 5.12 were tested for association with SZ. A conditional analysis was also performed for terms that were enriched for DEGs in SHSY5Y cells, 'SH-SY5Y expressed genes'; a gene-set containing all genes expressed in SH-SY5Y cells was used as a covariate in the enrichment analysis, to ensure the association signal was not been driven by neuronally expressed genes in general. No. of genes; number of genes in gene-set that were analysed, SE; standard error.
### Supplementary Table 5.14: Gene-set analysis of GO terms impacted by SDCCAG8 deficiency in IQ

<table>
<thead>
<tr>
<th>Source</th>
<th>Gene-set</th>
<th>No. of Genes</th>
<th>Beta</th>
<th>SE</th>
<th>P-value</th>
<th>Beta</th>
<th>SE</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>RPE1</td>
<td>Collagen-containing extracellular matrix</td>
<td>408</td>
<td>-0.1</td>
<td>0.0611</td>
<td>0.94917</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPE1</td>
<td>extracellular matrix organization</td>
<td>365</td>
<td>-0.109</td>
<td>0.0615</td>
<td>0.96177</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPE1</td>
<td>skeletal system morphogenesis</td>
<td>243</td>
<td>-0.0892</td>
<td>0.0757</td>
<td>0.88065</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPE1</td>
<td>cell differentiation</td>
<td>4043</td>
<td>0.0597</td>
<td>0.0215</td>
<td>0.002725</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPE1</td>
<td>regulation of multicellular organismal process</td>
<td>3035</td>
<td>0.0356</td>
<td>0.0236</td>
<td>0.065962</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SH-SY5Y</td>
<td>trans-synaptic signaling</td>
<td>707</td>
<td>0.179</td>
<td>0.0451</td>
<td><strong>3.72E-05</strong></td>
<td>0.176</td>
<td>0.0451</td>
<td><strong>4.84E-05</strong></td>
</tr>
<tr>
<td>SH-SY5Y</td>
<td>dendrite</td>
<td>620</td>
<td>0.208</td>
<td>0.0484</td>
<td><strong>8.85E-06</strong></td>
<td>0.204</td>
<td>0.0484</td>
<td><strong>1.29E-05</strong></td>
</tr>
<tr>
<td>SH-SY5Y</td>
<td>generation of neurons</td>
<td>1494</td>
<td>0.175</td>
<td>0.0324</td>
<td><strong>3.72E-08</strong></td>
<td>0.169</td>
<td>0.0325</td>
<td><strong>1.04E-07</strong></td>
</tr>
<tr>
<td>SH-SY5Y</td>
<td>synapse organization</td>
<td>417</td>
<td>0.191</td>
<td>0.0586</td>
<td><strong>0.000554</strong></td>
<td>0.186</td>
<td>0.0585</td>
<td><strong>0.00075</strong></td>
</tr>
<tr>
<td>SH-SY5Y</td>
<td>sister chromatid segregation</td>
<td>190</td>
<td>0.119</td>
<td>0.083</td>
<td>0.076481</td>
<td>0.113</td>
<td>0.0829</td>
<td>0.085498</td>
</tr>
</tbody>
</table>

Notes: Output from MAGMA analysis, a subset of highly enriched GO terms in Supplementary Tables 5.5, 5.6, 5.11 and 5.12 were tested for association with IQ. A conditional analysis was also performed for terms that were enriched for DEGs in SH-SY5Y cells, 'SH-SY5Y expressed genes'; a gene-set containing all genes expressed in SHSY5Y cells was used as a covariate in the enrichment analysis, to ensure the association signal was not been driven by neuronally expressed genes in general. No. of genes; number of genes in gene-set that were analysed, SE; standard error.
Supplementary Table 5.15: Gene-set analysis of GO terms impacted by SDCCAG8 deficiency in EA

<table>
<thead>
<tr>
<th>Source</th>
<th>Gene-set</th>
<th>No. of Genes</th>
<th>Beta</th>
<th>SE</th>
<th>P-value</th>
<th>Conditioned on 'SH-SY5Y expressed genes'</th>
<th>Beta</th>
<th>SE</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>RPE1</td>
<td>Collagen-containing extracellular matrix</td>
<td>392</td>
<td>-0.221</td>
<td>0.074</td>
<td>0.9987</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPE1</td>
<td>extracellular matrix organization</td>
<td>357</td>
<td>-0.281</td>
<td>0.075</td>
<td>0.9999</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPE1</td>
<td>skeletal system morphogenesis</td>
<td>242</td>
<td>-0.0083</td>
<td>0.093</td>
<td>0.5356</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPE1</td>
<td>cell differentiation</td>
<td>3920</td>
<td>0.0574</td>
<td>0.0258</td>
<td>0.0130</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPE1</td>
<td>regulation of multicellular organismal process</td>
<td>2940</td>
<td>0.0607</td>
<td>0.0288</td>
<td>0.0176</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SH-SY5Y</td>
<td>trans-synaptic signaling</td>
<td>686</td>
<td>0.306</td>
<td>0.0553</td>
<td>1.67E-08</td>
<td></td>
<td>0.301</td>
<td>0.0552</td>
<td>2.6E-08</td>
</tr>
<tr>
<td>SH-SY5Y</td>
<td>dendrite</td>
<td>592</td>
<td>0.287</td>
<td>0.0577</td>
<td>3.24E-07</td>
<td></td>
<td>0.281</td>
<td>0.0576</td>
<td>5.45E-07</td>
</tr>
<tr>
<td>SH-SY5Y</td>
<td>generation of neurons</td>
<td>1442</td>
<td>0.264</td>
<td>0.0387</td>
<td>4.39E-12</td>
<td></td>
<td>0.25</td>
<td>0.0386</td>
<td>5.15E-11</td>
</tr>
<tr>
<td>SH-SY5Y</td>
<td>synapse organization</td>
<td>391</td>
<td>0.512</td>
<td>0.0745</td>
<td>3.31E-12</td>
<td></td>
<td>0.503</td>
<td>0.0744</td>
<td>7.2E-12</td>
</tr>
<tr>
<td>SH-SY5Y</td>
<td>sister chromatid segregation</td>
<td>184</td>
<td>0.00215</td>
<td>0.104</td>
<td>0.49179</td>
<td></td>
<td>-0.016</td>
<td>0.104</td>
<td>0.56092</td>
</tr>
</tbody>
</table>

Notes: Output from MAGMA analysis, a subset of highly enriched GO terms in Supplementary Tables 5.5, 5.6, 5.11 and 5.12 were tested for association with EA. A conditional analysis was also performed for terms that were enriched for DEGs in SHSY5Y cells, 'SH-SY5Y expressed genes'; a gene-set containing all genes expressed in SH-SY5Y cells was used as a covariate in the enrichment analysis, to ensure the association signal was not been driven by neuronally expressed genes in general. No. of genes; number of genes in gene-set that were analysed, SE; standard error.
6. General Discussion

6.1 Summary of main findings

A number of genes mapping to risk loci for SZ encode proteins that function at the centrosome. Five of the six candidate genes harboured SZ risk SNPs (SDCCAG8; rs14403, MPHOSPH9; rs2851447, PRKDL1; rs2068012, MAD1L1; rs10650434, and MAPK3; rs12691307) that were found to be associated with performance in tests from four of the five cognitive domains tested; IQ, episodic memory, working memory and social cognition. Building on these single gene associations, my centrosomal list of genes showed enrichment for common variants associated with IQ. This list of centrosomal genes was also enriched for rare de novo variants reported in individuals with ASD and intellectual disability. These results taken together suggest that the molecular mechanisms that underpin neurodevelopmental disease and cognitive function may include disruption of biological processes influenced by the centrosome.

Following these analyses, I focused on a specific centrosomal protein, SDCCAG8, based on its association with social cognition in the Irish dataset, and its causal role in ciliopathies where patients present with intellectual disability. Reverse genetic analyses of SDCCAG8 revealed that the protein plays an important role in ciliation and ciliary dependent signaling, differentiation and cell migration.

A large number of genes are transcriptionally dysregulated in the absence of SDCCAG8. Pathways enriched for DEGs in SDCCAG8-deficient cells are involved in the deposition and turnover of ECM components, suggesting that the ECM may be dysregulated in the absence of SDCCAG8. In addition, pathways directly related to neuronal functioning, including GABA and glutamate signalling and axon guidance signalling, were enriched for DEGs in SDCCAG8. I have shown that the pathways enriched for SDCCAG8 DEGs in SH-SY5Y cells are enriched for common variation associated with SZ risk and cognitive ability. Many SDCCAG8-dysregulated pathways have been implicated previously in SZ pathophysiology and normal brain development. Importantly, many of these pathways have also been shown to be regulated, at least in part, by primary cilia.
6.2 Strengths and weaknesses

6.2.1 GWAS

A strength of the analyses in this thesis is that they are based on data from GWAS that used very large sample sizes. These sample sizes provide adequate statistical power to detect common variants that have small effects on phenotype. This approach has been utilised to identify risk loci associated with a large number of traits and conditions and GWAS loci can lead to the discovery of novel biological function (Jhamb et al., 2019, Visscher et al., 2017). GWAS continues to be utilised extensively by the psychiatric and cognitive genetics fields, leading to the frequent release of studies with growing sample sizes. My studies progressed with the updates to these fields and used the largest GWAS datasets available. In the future these analyses can be repeated in larger datasets as they become available, which may further strengthen the findings from this thesis.

A general limitation in GWAS is that the causal variant is not always detected and the mechanism by which genetic variation affects a gene product at a risk locus and how this contributes to risk are not immediately evident from GWAS results. LD is useful in the initial identification of a locus, but can also make it difficult to discern the true causal variant. In addition to this, most association signals map to non-coding regions of the genome, making biological interpretation challenging. Following GWAS, additional analyses are often required to identify causal variants and their target genes. A strength of this analysis was the use of eQTL data to help associate risk SNPs with functional impact on local genes. The two main datasets I used to do this were Braineac (http://braineac.org/) and GTEx (https://gtexportal.org/home/). These datasets do have limitations; the sample sizes are small for analyses of brain samples, the analysis is performed on post-mortem tissue of control subjects and the tissue samples collected are heterogeneous collections of multiple cell types within a brain region. Larger datasets have now become available to aid in the detection of brain eQTLs. The Common Mind Consortium (https://www.synapse.org) has collected brain eQTL data for SZ cases and healthy controls. Other types of QTL data have also been utilised in an attempt to understand how genetic variation contributes to SZ risk, including methylation,
splicing and chromatin accessibility QTLs (Bryois et al., 2018, Takata et al., 2017, Hannon et al., 2015). Additional tools have also been developed to aid in fine mapping of risk loci to genes. The FUMA tool (Watanabe et al., 2017) is a platform designed to annotate, prioritize, visualise and interpret GWAS results. Summary data based Mendelian Randomization (Wu et al., 2018b) tests for association between gene expression and complex traits using GWAS summary statistics and eQTL data. It determines if SNP effect size is mediated by gene expression levels. These tools are helping to prioritise genes for molecular analysis based on data generated by GWAS.

A limitation of GWAS of psychiatric conditions is the phenotypic heterogeneity. In order to increase sample sizes, all patients are placed into a single case grouping. However, these individuals can have drastically different presentations of SZ, e.g. the degree to which they experience cognitive impairment and the severity of positive and negative symptoms can vary greatly between patients. Future GWAS that stratify patients into more phenotypically similar groupings may uncover genetic variants that are relevant to specific symptomology.

6.2.2 Neuropsychological analysis
Endophenotypes have been used in psychiatric genetics because they allow individuals to be stratified based on a measurable trait such as cognitive performance. A strength of our study was the depth of the neuropsychological data that we had available for cases and controls in our Irish dataset. Five domains of cognition were analysed, where each domain was represented by a number of different tests. However, it is extremely difficult to perform extensive testing like this on a large sample of individuals and so our sample size was small, limiting my power to detect statistical differences. There are many different measures that can be used to assess cognitive ability, which became an issue when I attempted to replicate findings in the Irish dataset in other international samples. Comparing different measures of cognitive ability can weaken any replication found or result in a failure to find similar associations. To truly capture the genetic architecture of endophenotypes in psychiatry, a very large sample size, with consistent measures of a broad range of symptoms and cognitive measures, would be required.
6.2.3 Model systems

I chose to perform our molecular analyses in human cell lines, to characterise the mechanisms by which SDCCAG8 deficiency impacts upon cellular function, which may be relevant to SZ pathophysiology and general cognition. RPE1 cells were chosen because of their high ciliation frequency, allowing me to investigate the role of SDCCAG8 in ciliogenesis. SH-SY5Y cells were chosen because the immature neuronal cell line has a gene expression profile more relevant to our phenotypes of interest and due to the ability to differentiate these cells in culture. I used CRISPR/Cas9 to disrupt the *SDCCAG8* locus, as this had not been previously performed in human cells and off-target effects, transient depletion and residual functionality of proteins, can hamper the reproducibility and validity of reported findings from siRNA studies (Boettcher and McManus, 2016, Birmingham et al., 2006, Jackson et al., 2003). Furthermore, siRNA-mediated knockdown and gene knockout of centriolar components have been shown to have different phenotypic effects on ciliary function (Hall et al., 2013), highlighting the importance of performing knockout studies when interpreting the role of a protein in human disease. These functional studies allowed us to better characterise SDCCAG8 and its role in human disease. While loss-of-function (LoF) variants in *SDCCAG8* cause Bardet-Biedl and Senior–Løken syndrome, the SNPs in *SDCCAG8* associated with SZ risk and cognitive ability are intronic variants, although eQTL data suggest that they may decrease expression of *SDCCAG8*. While the cell lines I have generated have LoF variants in *SDCCAG8*, similar to that found in patients with ciliopathies, they can be used to understand what SDCCAG8-related molecular functions may be relevant to SZ pathophysiology and cognition.

A huge advantage of using animal models instead of cell lines is the opportunity to study effects on the full organism, including genetic, proteomic, cellular, tissue and behavioural changes. However, a major challenge in psychiatric research is the lack of suitable model systems, as the hallmarks of these conditions are behaviours which are specific to humans—hallucinations, delusions and paranoia. Some behaviours, however, can be reproduced, such as abnormal social interaction, cognitive impairment and reduced motivation, but the model remains incomplete. A number of mouse models for SZ have been developed, using four induction methods; developmental, drug-induced, lesion or genetic manipulation. However the
development of more comprehensive models that more adequately replicate human deficits and help to understand causal factors is required (Jones et al., 2011). In mice lacking SDCCAG8 there was no obvious microcephaly, but some mice had enlargement of the lateral ventricle (Insolera et al., 2014). Lateral ventricle enlargement has been consistently and reproducibly associated with SZ (Wright et al., 2000). However, no cognitive measures were analysed in SDCCAG8 deficient mice, limiting the insights to SZ-relevant phenotypes.

6.2.4 RNA-seq
I used RNA-Seq to characterise alterations to gene expression in the absence of SDCCAG8. The strengths of this approach are that it is hypothesis free, expression profiles are genome-wide and transcript detection is not limited to those that have been previously characterized. There is dynamic range within transcript abundances that can be detected; transcripts that are either lowly expressed or very abundant can all be profiled and the method has been shown to be highly reproducible (Wang et al., 2009b). The main concern when using this approach is distilling useful biological insights from the wealth of information that is generated. The large number of genes that were differentially expressed in my analyses made it very difficult to focus on all these genes individually. Instead I relied on data generated from pathway analysis to understand how SDCCAG8 deficiency was influencing cellular processes. I used two different tools; IPA and ConsensusPathDB, for these enrichment analyses, and these different methodologies produced largely consistent data, which supports the authenticity of the reported pathway findings.

6.3 Mechanistic insight
Data from this thesis suggests that SDCCAG8 is a positive regulator of ciliogenesis and ciliary dependent signalling, which is in line with previously published work (Airik et al., 2016). I have shown that SDCCAG8 deficiency leads to defective recruitment of pericentrin and PCM1 in asynchronous cells. Multiple TZ components have been shown to localise with satellite components, suggesting that satellite components play a role in trafficking TZ proteins (Quarantotti et al., 2019, Gupta et al., 2015, Klinger et al., 2013). SDCCAG8 has been shown to interact with the satellite components OFD1, PCM1 and CEP131, all of which contribute to
ciliogenesis (Airik et al., 2016, Insolera et al., 2014, Wang et al., 2013, Hall et al., 2013 Singla et al., 2010). SDCCAG8 localises to the transition zone in ciliated cells. These data suggest that SDCCGA8 plays a role in trafficking proteins to the transition zone both through its localisation at the TZ and the regulation of satellite components. This hypothesis is graphically represented in Figure 6.1. Further characterisation of additional TZ and satellite components or live cell imaging of these components in SDCCAG8-deficient cells undergoing ciliogenesis could help prove this hypothesis. Strikingly, a primary cilium formation defect has been reported in cells derived from patients with SZ and bipolar disorder (Muñoz-Estrada et al., 2018). Further immunofluorescence characterisation of patient derived cells for key regulators of ciliogenesis, including satellite and TZ components could help the cause of the ciliogenesis defect and provide important insight into the role of the centrosome in SZ.

![Figure 6.1: The role of SDCCAG8 in ciliogenesis](image)

SDCCAG8 localises to the transition zone of ciliated cells. SDCCAG8 deficient cells have reductions in PCM and satellite components. I hypothesise that this satellite dysregulation leads to impaired protein recruitment at the transition zone, causing the ciliary defects in SDCCAG8 deficient cells namely; reduced ciliation frequency and ciliary length, and impairments in ciliary dependent signalling.
There is extensive transcriptional dysregulation in SDCCAG8-cells, consistent with the defect in ciliogenesis and ciliary dependent signalling. The transcriptional dysregulation of neuronal projection pathways observed upon ablation of SDCCAG8 is consistent with the defect in migration and differentiation seen in SDCCAG8-deficient cells. Migration and differentiation have both been shown to be defective in cells derived from patients with SZ (Ahmad et al., 2018, Tee et al., 2017, 2016, Fan et al., 2013 and Sei et al., 2007). I have shown that biological pathways and cellular components that are dysregulated in immature neurons upon depletion of SDCCAG8 contain genes that contribute to SZ risk and cognitive function in the general population. The link between cell migration, cell differentiation and the ciliary control of such activities through the regulation of gene expression that is evidenced by our analyses of SDCCAG8 provides a model for how centrosomal genes may contribute to SZ risk and cognitive function.

6.4 Future directions
The focus of my molecular analysis of SDCCAG8 was to model how centrosomal proteins may contribute to SZ and cognition. However, a number of other genes that encode proteins that are localised to the centrosome have been also been associated with SZ risk including GIGYF2, MPHOSPH9, MAPK3, PRKD1 and MAD1L1, which are the remaining five candidate genes from my original analysis of centrosomal genes in SZ. Functional analysis of these genes and other candidates identified by future larger genetic studies would help clarify if centrosome proteins influence SZ risk and cognition by converging on the same molecular pathways.

My candidate genes were selected from GWAS that investigated the relationship between common variation and a phenotype. The majority of variants identified by GWAS are intronic or intergenic, with small effect sizes, making it difficult to assess their functional impact on local genes and possibly genes located elsewhere in the genome. Rare variants mapping to coding regions are more likely to alter gene function. Exome sequencing studies have indicated that there is an increased abundance of ultra-rare (population frequency <0.01%) disruptive coding mutations in SZ patients compared to healthy controls (Genovese et al., 2016, Purcell et al., 2014). However, the rarity of individual mutations and their broad distribution
throughout the genome has made it difficult to implicate specific genes. In the future, larger sample sizes combined with exome or genome sequencing will give us a better understanding of the role of rare variants in SZ, either through extended case-parent trios-based studies of de novo variants or population case-control association studies of rare mutations. It will be interesting when larger datasets are available, to investigate the contribution of rare variants in centrosomal and ciliary genes to neuropsychiatric conditions and cognition.

Another cellular approach to take is obtaining tissue from SZ patients. Human induced pluripotent stem cells (iPSCs), usually generated from patient fibroblasts, can provide good cellular models of developmental disorders (Ardhanareeswaran et al., 2017). Both neural progenitor cells (NPCs) and neurons can be derived from IPSCs and will carry the risk mutation/s as well as the entire genetic profile of the donor. Comparing patient NPCs with control NPCs can provide insight into disease pathobiology. If rare mutations in a centrosomal or ciliary gene were found to be a risk factor for SZ, it would be interesting to investigate ciliation and ciliary dependent signalling in patient- and control-derived cells, and also to investigate the abundance of PCM and satellite proteins that I found to be altered in this study. However, the process of iPSC generation has many caveats including low efficiency, risk of new mutations being inserted with transcriptional reprogramming, tumour growth, and incomplete reprogramming due to factors such as epigenetic modifications (Omole and Fakoya, 2018). Recentl methods have been developed to generate brain organoids—self-assembled, three-dimensional aggregates generated from pluripotent cells with cell types and cytoarchitectures that resemble the embryonic human brain (Qian et al., 2019). This system would allow even more complex primary cilia functions to be evaluated in a CNS-like environment.

6.5 Translational applications and drug discovery
The field of psychiatric genetics has rapidly evolved over the past decade with the identification of thousands of risk loci for common disorders, including SZ. Despite these findings, there remains a large gap between variant discovery and meaningful translation to therapeutics and disease prevention. The neuroscientist and former head of the National Institutes of Mental Health (NIMH) Thomas Insel said “I spent 13 years at NIMH really pushing on the
neuroscience and genetics of mental disorders, and when I look back on that I realize that while I think I succeeded at getting lots of really cool papers published by cool scientists at fairly large costs—I think $20 billion—I don’t think we moved the needle in reducing suicide, reducing hospitalizations, improving recovery for the tens of millions of people who have mental illness” (Rogers, 2017). The role of the primary cilia and centrosome in neuropsychiatric conditions and cognition should be investigated further, in part because the primary cilium is a druggable target. A recent study has shown that defects caused by CEP290 mutation, which causes a number of different ciliopathies, can be rescued by the administration of the flavonoid, eupatilin. Eupatilin relieved ciliogenesis and ciliary receptor defects by inhibiting calmodulin binding of NPHP5, a TZ component that interacts with CEP290, which led to NPHP5 relocating to the ciliary base (Kim et al., 2018). As mentioned previously, the development of cancer is often accompanied by loss of primary cilia. One screen identified 118 different compounds that restored primary cilia in pancreatic ductal cancer cells (Khan et al., 2016). These compounds include glucocorticoids, fibrates and other nuclear receptor modulators, neurotransmitter regulators, ion channel modulators, tyrosine kinase inhibitors, DNA gyrase/topoisomerase inhibitors, antibacterial compounds, protein inhibitors, microtubule modulators and COX inhibitors. Most of the compounds attenuated cell proliferation, at least partially due to the restoration of primary cilia (Khan et al., 2016). Chondrosarcoma is a bone tumour characterized by excretion of cartilage-like ECM. Inhibition of HDAC6, which as mentioned earlier destabilises axonemal MTs, can restore primary cilium assembly and suppress chondrosarcoma cell proliferation and invasion capacities (Xiang et al., 2017). In addition, obesity-impaired adipose-derived mesenchymal stem cells (ASCs) have been shown to have impairments in ciliation. Low dose inhibition of Aurora A, an activator of HDAC6 or ERK1/2, restored primary cilia and the invasion and migration potential capacity of obese ASCs, as well as improving their differentiation (Ritter et al., 2019). These compounds could be screened on cells derived from SZ patients to investigate their effect on migration, differentiation and SHh signalling impairments.

6.6 Concluding remarks

SZ is a neurodevelopmental disorder affecting 1% of adults and is a major global health problem, but the pathophysiology remains largely unknown. SZ is highly heritable and recent
GWAS have identified over 100 common low-risk SNPs across the genome. Similarly, GWAS has identified over 100 loci for EA. EA is used as a “proxy” for general cognition because it correlates with cognitive ability ($r \sim 0.5$). GWAS data indicates that there is a significant genetic correlation between SZ and EA pointing to the genetic basis of cognitive deficits in SZ. My PhD focused on identifying genes that contribute to these cognitive deficits to understand the molecular mechanisms at play and identify candidate risk genes and drug targets. My specific focus was genes that encode proteins with centrosomal functions, the centrosome mediates normal brain development which is relevant to SZ and cognition.

This thesis built on single gene associations detected in GWAS of SZ to show that genes that encode proteins with centrosomal functions are enriched for common variants associated with IQ, and are enriched for rare variants that increase risk of ASD and intellectual disability. Reverse genetic analysis of SDCCAG8 was used to model the potential role of centrosome genes in SZ pathophysiology and cognition. The link between cell migration, cell differentiation and the ciliary control of such activities through the regulation of gene expression that is evidenced by my analyses of SDCCAG8 provides a new model for how centrosomal genes may contribute to SZ risk and cognitive function.
References


References


References


Cambridge Cognition (2019). CANTAB® [Cognitive assessment software]. All rights reserved. www.cantab.com


References


References


References


Irish Schizophrenia Genomics Consortium and The Wellcome Trust Case Control Consortium 2 (2012). Genome-wide association study implicates HLA-C*01:02 as a risk factor at the major histocompatibility complex locus in schizophrenia. Biol Psychiatry 72, 620-628.


References


References


References


References


Syndrome and Are Associated with Penetrant Renal Disease and Absent Polydactyly. Molecular syndromology 1, 273-281.


References


References


Ware, J.S., Samocha, K.E., Homsy, J., and Daly, M.J. (2015). Interpreting de novo Variation in Human Disease Using denovolyzeR. Current protocols in human genetics 87, 7.25.21-27.25.15.


References
References


Appendices

Appendix A: List of Presentations and Awards

Oral presentations:
December 2018, Galway Neuroscience Centre’s Research Day, NUIG “Molecular analysis of SDCCAG8, a schizophrenia risk gene that functions at the centrosome”

September 2017, 20th Annual Scientific Meeting of the Irish Society of Human Genetics, Dublin “A molecular analysis of SDCCAG8, a schizophrenia risk gene that functions at the centrosome.”


Poster presentations:
October 2018, World Congress of Psychiatric Genetics, Glasgow “Molecular analysis of SDCCAG8, a schizophrenia risk gene that functions at the centrosome”.

September 2018, Irish Society of Human Genetics, Dublin “Molecular analysis of SDCCAG8, a schizophrenia risk gene that functions at the centrosome”

June 2018, European Society of Human Genetics, Milan, “Molecular analysis of SDCCAG8, a schizophrenia risk gene that functions at the centrosome”.

October 2017 World Congress of Psychiatric Genetics, Orlando, Florida “A molecular analysis of SDCCAG8, a schizophrenia risk gene that functions at the centrosome.”

September 2016, Irish Society of Human Genetics, Belfast “The Role of the Centrosome in Schizophrenia”.
Awards:
2018 Thomas Crawford Hayes research fund award (€2,430, NUIG).
2017 Irish Society of Human Genetics Young Investigator Award for Best Postgraduate Oral Presentation (Dublin). The award included a scholarship to the European Society of Human genetics in Milan June 2018.
2017 Thomas Crawford Hayes research fund award (€5,000, NUIG)
2016 Runner up, Best Oral Presentation at Galway Neuroscience Research Day (Galway).
2016 Best paper at the 6th UL/NUIG Annual Postgraduate Research Day (Limerick).
2016 Government of Ireland Postgraduate Scholarship (GOIPG/2016/506; replacing College of Science Scholarship).
2015 Graduate Research Scholarship, College of Science, NUI Galway.

Appendix B: Submissions
**Flynn M**, Whitton L, Donohoe G, Morrison CG and Morris DW. Altered gene regulation as a candidate mechanism by which ciliopathy gene SDCCAG8 contributes to schizophrenia and cognitive function (submitted for publication).

Appendix C: Study Funding and Acknowledgements
Recruitment and genotyping was supported by Science Foundation Ireland (SFI) (grants 12.IP.1359 and 08/IN.1/B1916) and the Wellcome Trust Case Control Consortium 2 project (grant 085475/B/08/Z) and the Wellcome Trust (grants 072894/Z/03/Z, 090532/Z/09/Z, and 075491/Z/04/B).

The Wellcome Trust Case Control Consortium 2 investigators include:
Peter Donnelly, Lesley Bates, Ines Barroso, Jenefer M. Blackwell, Elvira Bramon, Matthew A. Brown, Juan P. Casas, Aiden Corvin, Panos Deloukas, Audrey Duncanson, Janusz Jankowski, Hugh S. Markus, Christopher G. Mathew, Colin N. A. Palmer, Robert Plomin, Anna Rautanen,
Appendices


Thank you to all patients and their support staff, and all healthy volunteers for participating in the data collection on which part of this thesis is based.