Advancing Our Understanding Of The Ciliopathy, Bardet-Biedl Syndrome: An Omics Approach

A thesis submitted for the degree of

Doctor of Philosophy

to

University College London

Rosalind Jane Davies

Genetics and Genomic Medicine

UCL Great Ormond Street Institute of Child Health

London

September 2017

Declaration of originality

I, Rosalind Davies, hereby declare that the work presented in this thesis is original and my own. All material in this document has been obtained in accordance with academic rules and ethical conduct. To the best of my knowledge and belief, it contains no material previously published or produced by another party. Information that has been derived from other sources has been included with permission and properly acknowledged in the text.

Rosalind Davies

September 2017

Abstract

Bardet-Biedl Syndrome (BBS) is a rare pleiotropic disorder, characterised by loss of vision, obesity, renal dysfunction, learning difficulties, and hypogonadism. This multisystem phenotype is caused by defects in genes that localize to the basal body of the primary cilia, where over 20 genes have now been attributed to cause BBS. However, ~32% of patients are affected by a recurrent missense variant, namely *BBS1* p.M390R, which contributes greatly to the overall burden of BBS. Despite recent advances in understanding the syndrome's genetic aetiology, much of the pathobiology of *BBS1* p.M390R remains elusive. This thesis aimed to use an innovative multi-omic strategy, implementing genomic, transcriptomic, and proteomic technologies, to uncover the molecular pathology of a cohort of 15 BBS patients, each carrying the common *BBS1* p.M390R variant.

Phenotypic variability is a hallmark feature of BBS, where clinical heterogeneity exists within the *BBS1* genotype, even if patients have the same underlying pathogenic mutation. It has been suggested that differences in disease expressivity are linked to secondary mutations that modify the manifestation of the primary locus. Here, the objective was to identify putative genetic modifying alleles from whole genome sequencing data generated from *BBS1* p.M390R patients expressing discordant disease presentation. A novel 4-tier variant categorisation system was developed where 37 candidate modifiers were detected in 13 patients. This included a known modifying variant previously shown to have a high penetrance with ciliopathies, *TTC21B* p.L1002V found in 2 patients. Furthermore, it was investigated whether the presence of these modifying variants contributed to the overall mutational burden of BBS. Mutational burden analysis determined that there was no significant enrichment of variants in primary cilia genes in BBS patients compared to control individuals.

There is a definitive lack of molecular biomarkers for rare diseases, such as BBS. Untargeted and targeted proteomic profiling assays were applied to plasma and urine, which aimed to uncover novel biomarkers specific to BBS. 8 significantly differentially expressed proteins were identified from urine, including PEDF, a secreted factor that is linked to fatty acid metabolism and insulin resistance in obesity (Log2FC: 2.56, $p = 0.015$). Similarly, plasma proteomic analysis identified putative biomarkers that are linked to secondary metabolic features of BBS, such as LEP and ApoM. Markers, such as these, will become subsequent targets for further validation in larger cohorts.

BBS patient-derived fibroblasts were profiled by transcriptomic and proteomic technologies, which aimed to identify dysregulated pathways at a cellular level compared to control cultures. Pathway analysis uncovered discordant expression of centrosomal genes between BBS and control cells, as well as a significant enrichment of genes associated with adipogenesis, which may provide insight into obesity manifested by BBS patients. Analysis of protein profiling data revealed dysregulation of processes not detected by RNA-seq, including actin cytoskeleton remodelling and hedgehog signalling. Finally, pathways were integrated to increase the power of analysis, identifying 17 pathways that were found to be impaired at both transcript and protein levels.

The phenotype of retinal dystrophy is one of the most detrimental effects on patient welfare, and affects over 90% of BBS patients. As retinal degeneration of BBS patients cannot be investigated *in vivo*, this project utilised BBS patient-derived induced pluripotent stem cells with the aim of developing a model for the study of retinal degeneration *in vitro*. For the first time, it was shown that BBS patient cells can differentiate into three-dimensional optic cups, which recapitulated the temporal expression of key retinal markers of *in vivo* mammalian eye development. Furthermore, immunohistochemistry and electron microscopy assays determined that BBS-derived optic cups could successfully undergo ciliogenesis, which was demonstrated by the formation of nascent photoreceptor outer segments.

Table of contents

Declaration of originality 3

Abstract 4

Table of contents 6

List of figures 16

List of tables 20

Acknowledgements 22

List of abbreviations 23

Chapter 1: Introduction 29

1.1 Eukaryotic cilia 29

1.2 The core of the primary cilium 30

1.3 Intraflagellar transport 30

1.4 Key processes enabled by primary cilia 33

1.4.1 G-Protein Coupled Receptors 33

1.4.2 Hedgehog signalling 34

1.4.3 Mechanosensation 34

1.4.4 Wnt signalling and planar cell polarity 36

1.5 The nonmotile ciliopathies 37

1.5.1 Syndromic nonskeletal ciliopathies 37

1.5.2 Skeletal ciliopathies 39

1.5.3 Nonsyndromic ciliopathies 40

1.6 A spectrum of overlapping heterogeneous disorders 40

1.7 The Ciliome 43

1.8 Bardet-Biedl Syndrome 45

1.8.1 Identification of BBS loci 46

1.8.2 BBS causality and localisation 49

1.8.3 Complexes, roles and interactions 49

1.8.4 Molecular basis of BBS 52

1.8.4.1 Retinal dystrophy 52

1.8.4.2 Obesity 52

1.8.4.3 Polydactyly 53

1.8.4.4 Genital abnormalities 54
1.8.4.5 Renal dysfunction 55
1.8.4.6 Learning difficulties 55
1.8.5 Genotype-Phenotype Correlation in BBS 56
1.8.6 Aetiology of *BBS1* p.M390R 57
1.9 In-depth Phenotyping of Rare Diseases 58
1.9.1 High definition, in-depth phenotyping at Great Ormond Street Hospital 58
1.9.2 Study of rare diseases 58
1.10 The age of next generation sequencing and omic analysis 59
1.10.1 Precision medicine 60
1.10.2 Integrative personal omics profile study: Proof of principle 61
1.11 Introduction to omics techniques 61
1.11.1 Whole genome and whole exome sequencing 61
1.11.1.1 Types of human genetic variation 62
1.11.1.2 Classification of pathogenicity 63
1.11.2 Transcriptomics by RNA-sequencing 65
1.11.3 Proteomics 66
1.11.3.1 Mass spectrometry 66
1.11.3.2 Targeted proteomic techniques 67
1.12 Pathway analysis 68
1.12.1 Gene Set Enrichment Analysis 68
1.12.2 PANTHER Statistical Overrepresentation Test 69
1.12.3 PANTHER Statistical Enrichment Test 69
1.12.4 Pathway topology mapping 70
1.13 Induced pluripotent stem cells 70
1.14 Objectives of this thesis 71
Chapter 2: Materials, Methods, and Patients 74
2.1 Sample Collection 74
2.1.1 Ethics 74
2.1.2 Plasma isolation of whole blood 74
2.1.3 Skin biopsy 74
2.1.4 Urine collection 75
2.2 Fibroblast culture 75
2.2.1 Culturing fibroblasts from a skin biopsy 75
2.2.1.1 Plating the skin biopsy 75

2.2.1.2 Passaging fibroblasts 75
2.2.1.3 Freezing and thawing fibroblasts 76
2.2.2 Treated and untreated fibroblasts for RNA extraction 76
2.2.3 Treated and untreated fibroblasts for protein extraction 76
2.3 Nucleic acid extraction 77
2.3.1 Isolation of DNA 77
2.3.2 RNA extraction 77
2.3.3 Assessment of nucleic acid quantity and purity 78
2.3.4 gDNA agarose gel electrophoresis 78
2.4. Protein extraction 79
2.4.1 Protein extraction from urine 79
2.4.2 Protein enrichment by ProteoMiner kit 79
2.4.3 Homogenisation of fibroblasts 80
2.4.4 Protein quantitation by BCA assay 80
2.5 Protein preparation for unlabelled liquid chromatography mass spectrometry 80
2.5.1 Acetone protein precipitation 80
2.5.2 Digest and clean-up of protein from urine 81
2.5.3 Protein preparation from plasma 81
2.5.3.1 SDS-PAGE 81
2.5.3.2 In-gel trypsin digest and clean up 82
2.5.4 Protein preparation for SOMAscan 83
2.5.5 Protein preparation for iTRAQ mass spectrometry 83
2.6 Sequencing and profiling 84
2.6.1 Whole genome sequencing 84
2.6.2 Whole exome sequencing 84
2.6.3 RNA sequencing 84
2.6.4 Unlabelled tandem mass spectrometry 84
2.6.4.1 Synapt G2 Si QToF 85
2.6.4.2 MALDI QToF Premier 85
2.6.5 iTRAQ TripleTOF mass spectrometry 85
2.7 Data processing 86
2.7.1 Whole genome and whole exome sequencing 86
2.7.2 RNA sequencing 87
2.7.3 Protein identification and quantitation of unlabelled MS data 87
2.7.4 Protein identification and quantitation of iTRAQ data 88

2.7.5 Differential Gene Expression 88
2.7.6 Pathway analysis 88
2.7.7 Gene Set Enrichment Analysis 88
2.8 Identification of modifying alleles from NGS data 90
2.8.1 Identification of modifying alleles 90
2.8.2 Burden testing 90
2.9 Induced pluripotent stem cell culture 91
2.9.1 Reprogramming of fibroblasts into induced pluripotent stem cells 91
2.9.2 Maintenance of iPSC colonies 92
2.9.2.1 Thawing iPSC colonies 92
2.9.2.2 Passaging iPSC colonies 92
2.9.2.3 Freezing iPSC colonies 93
2.9.3 Differentiation into 3D optic cups 93
2.9.3.1 Experimental outline 93
2.9.3.2 Embryoid body (EB) formation 94
2.9.3.3 Changing to neural induction media and laminin plating 94
2.9.3.4 Changing to retinal differentiation media 95
2.9.3.5 Mechanical lifting of neural retina domains 95
2.9.3.6 Addition of exogenous factors to retinal differentiation media 96
2.9.3.7 Sample collection 96
2.10 Polymerase Chain Reaction 97
2.10.1 Reverse Transcription from RNA to cDNA 97
2.10.2 Designing primers 97
2.10.3 Polymerase chain reaction 97
2.10.4 Primers 98
2.10.5 2% agarose gel electrophoresis 98
2.10.6 Quantitative PCR 99
2.10.7 Sanger sequencing of p.M390R 100
2.11 Immunofluorescence and imaging 100
2.11.1 Collection and cryosectioning optic cups at day 98, 147 and 198 100
2.11.2 Fixing of cultured cells 101
2.11.3 Immunofluorescence staining 101
2.11.4 Antibodies 101
2.11.5 Imaging 102
2.11.6 Calculating the percentage of ciliated fibroblasts 102

2.11.7 Transmission Electron Microscopy (TEM) 102
2.12 Software and Databases 103
2.13 Raw Data 103
2.14 Patients and controls 105
2.14.1 Recruitment criteria 105
2.14.2 Clinical details 105
2.14.2.1 BBS001 105
2.14.2.2 BBS002 105
2.14.2.3 BBS003 106
2.14.2.4 BBS004 106
2.14.2.5 BBS005 106
2.14.2.6 BBS006 106
2.14.2.7 BBS007 107
2.14.2.8 BBS008 107
2.14.2.9 BBS009 107
2.14.2.10 BBS010 107
2.14.2.11 BBS011 107
2.14.2.12 BBS012 108
2.14.2.13 BBS013 108
2.14.2.14 BBS014 108
2.14.2.15 BBS015 108
2.14.3 Control individuals 111
Chapter 3: Use of next generation sequencing data to investigate the presence of genetic modifiers in BBS 112
3.1. Background 112
3.1.1. Genetic modifiers in ciliopathies 112
3.1.2. Identification of modifiers 113
3.1.3. Characterisation of ciliopathy modifiers 114
3.1.4. Mutational burden 115
3.1.5. Aims of chapter 116
3.2. Results 118
3.2.1. Quality control for whole genome sequencing 118
3.2.2. Identification of modifiers 119
3.2.2.3. Known modifiers 129
3.2.2.4. Candidate modifiers 130
3.2.3. Mutational burden analysis 130

3.2.3.1. Mutational burden of patients from WES data 131
3.2.3.2. Mutation burden of WGS data 133
3.2.3.3. Combination of WES and WGS 133
3.2.3.4. Mutational burden within a larger gene panel 134
3.2.3.5. Burden analysis with control gene panel 135
3.2.3.6. Technical investigations 135
3.3. Discussion 139
3.3.1. Identification of potential modifier alleles 139
3.3.1.1. High confidence candidate modifier alleles 139
3.3.1.2. Low confidence candidate modifier alleles 142
3.3.1.3. Undiscovered modifiers 144
3.3.2. Future experiments for validation of modifiers 145
3.3.2.1. Screening in larger cohort 145
3.3.2.2. Functional validation 146
3.3.3. Mutational Burden Analysis 147
3.3.3.1. Mutational burden associated with PDMs in WES data 148
3.3.3.2. Differences found between WES and WGS burden analysis 148
3.3.3.3. Combination of datasets 149
3.3.4. Conclusions of burden analysis 149
3.3.4.1. Number of variants does not correlate with disease expressivity 150
3.3.4.2. Private mutations drive statistical significance in burden testing 150
3.3.5. Further investigations and experimental limitations 152
3.3.6. Final conclusion 153
Chapter 4: Utilisation of proteomic profiling techniques for identification of biomarkers in BBS 154
4.1 Background 154
4.1.1 Biomarkers 154
4.1.1.1 Biomarkers for Bardet-Biedl Syndrome 155
4.1.1.2 Strategy for biomarker identification 156
4.1.2 Proteomic profiling of biofluids 156
4.1.2.1 Urine 156
4.1.2.2 Plasma 157
4.1.3 Proteomic techniques for biomarker identification 158
4.1.3.1 Ultra-performance liquid chromatography mass spectrometry 158
4.1.3.2 Peptide identification and quantification from MS data 160
4.1.3.3 Peptide identification 160

4.1.3.4 Label-free quantitation 160
4.1.3.5 Challenges of Mass spectrometry 161
4.1.3.6 Alternative proteomic technique: SOMAlogic 164
4.1.4 Aims of investigation 166
4.2 Results 167
4.2.1 Label-free UPLC-MS/MS of urine 167
4.2.1.1 Quantitation of protein from urine 167
4.2.1.2 Identification of peptides and inferring proteins 168
4.2.1.3 Quantification metrics 169
4.2.1.4 Statistical testing and identification of differentially expressed proteins 172
4.2.1.5 Pathway analysis 180
4.2.2 Proteomic profiling of plasma by mass spectrometry 182
4.2.2.1 Reduction of plasma complexity 183
4.2.2.2 Label-free UPLC-MS/MS analysis of plasma 184
4.2.3 Proteomic profiling by a targeted technology: SOMAscan 186
4.2.3.1 Targeted proteomic profiling of plasma 186
4.2.3.2 Comparison of proteins by targeted and untargeted techniques 190
4.3 Discussion 191
4.3.1 Proteomic profiling of urine 191
4.3.1.1 Statistical analysis of proteins from urine 192
4.3.1.2 Identification of urinary protein biomarkers 192
4.3.1.3 Cellular components of urine 194
4.3.2 Proteomic profiling of plasma 195
4.3.2.1 Investigation of plasma proteome by UPLC-MS/MS 195
4.3.2.2 Protein profiling using a targeted technique: SOMAscan 197
4.3.3 Comparison of targeted and untargeted proteomic approaches 198
4.3.4 Considerations for future biomarker experiments 199
4.3.4.1 Further work for biomarker discovery and validation 199
Chapter 5: Exploration of cellular pathways in BBS patient fibroblasts through implementation of complete mRNA transcriptomic and deep proteomic profiling 201
5.1 Background 201
5.1.1 Regulatory pathways for control of ciliogenesis 201
5.1.1.1 The relationship between cilia and the cell cycle 201
5.1.1.2 Regulation by the actin cytoskeleton 203
5.1.1.3 The ciliary function in proteostasis 204
5.1.2 Strategy for pathway identification in BBS 205

5.1.2.1 iTRAQ proteomics 206
5.1.3 Aims of investigation 208
5.2 Results 210
5.2.1 Experimental design and quality control 210
5.2.1.1 Ciliation of fibroblasts 210
5.2.1.2 Quantification and quality control of RNA 211
5.2.1.3 Quantitation of protein from fibroblasts 214
5.2.2 Analysis of RNA-seq data 216
5.2.2.1 Identification of differentially expressed genes 216
5.2.2.2 Gene set enrichment analysis 217
5.2.2.3 Comparison of RNA-seq pathways between starved vs fed 220
5.2.2.4 Comparison of treatment with respect to disease 222
5.2.3 Identifications of proteins from iTRAQ-MS/MS 228
5.2.3.1 Differentially expressed proteins from iTRAQ experiments 228
5.2.3.2 Gene set enrichment analysis with iTRAQ proteomics data 230
5.2.3.3 Pathway analysis for iTRAQ BBS vs control experiments 231
5.2.4 Integration of transcriptomic and proteomic datasets 233
5.3 Discussion 238
5.3.1 Enrichment of pathways associated with adipogenesis in serum fed BBS cultures 238
5.3.2 Analysis of the effects of serum starvation on BBS and control fibroblasts 241
5.3.3 Many upregulated proteins linked to ciliary pathways in starved BBS vs control 243
5.3.4 Experimental limitations 246
5.3.4.1 Pathway enrichment analysis 246
5.3.4.2 Pathway topology mapping analysis 247
5.3.4.3 Database annotation 247
5.3.4.4 Limits of the study design and future work 248
Chapter 6: Generation of an *in vitro* retinal model from BBS patient-derived induced pluripotent stem cells for the study of retinal dystrophy 250
6.1 Background 250
6.1.1 Development of the eye 250
6.1.2 The retina 251
6.1.3 The photoreceptor and the cilium 253
6.1.4 Impacts of impaired connecting cilium 255
6.1.5 Protocols for iPSC differentiation into optic cups 257
6.1.6 Aims of investigation 259

6.2 Results 261
6.2.1 Maintenance and experimental design 261
6.2.1.1 Culturing of iPSC lines 261
6.2.1.2 Validation of disease causing mutation in iPSC lines 261
6.2.1.3 Identification of a suitable reference gene 262
6.2.2 Differentiation of iPSCs into 3D optic cups 264
6.2.2.1 Embryoid body formation 264
6.2.2.2 Formation of neural retina and retinal pigmented epithelium 267
6.2.2.3 Expression of photoreceptor specific cells 269
6.2.3 Differentiation of iPSCs from BBS008 and CTRL037 272
6.2.4 BBS-specific staining 276
6.2.5 Transmission Electron microscopy 278
6.3 Discussion 280
6.3.1 Formation of optic cups from BBS iPSCs 280
6.3.1.1 Recapitulation of the neural retina 280
6.3.1.2 Formation of connecting cilia and outer segments 280
6.3.2 Expression of the BBSome and IFT machinery 281
6.3.3 Variability within iPSC cultures 283
6.3.3.1 Embryoid body formation 283
6.3.3.2 Epigenetic variability 284
6.3.3.3 Genetic variability 284
6.3.3.4 Media and culturing techniques 285
6.3.4 Future role for this model 286
6.3.4.1 Optimisation of culture system 286
6.3.4.2 Model for therapeutic use 287
Chapter 7: Final Conclusions 289
7.1 Summary of findings 289
7.1.1 There is evidence of secondary modifying variants in BBS patients, which may point towards the underlying causes of phenotypic variability. However, these variants do not contribute to the overall mutational burden of BBS 290
7.1.2 Proteomic profiling uncovered molecular biomarkers that may be indicative of secondary features of obesity 291
7.1.3 A multi-omic strategy identified dysregulated pathways in BBS cell cycling and ciliated fibroblasts compared to controls 292

7.1.4 BBS-patient derived iPSCs were used to develop a model for investigation of retinal dystrophy *in vitro*, which successfully recapitulated gene expression of the native developing retina 293

7.2 Considerations for further work 294

Appendix 296

Supplemental data 301

Bibliography 304

List of figures

Figure 1.1 – Cross sections of cilia. 31
Figure 1.2 – Examples of key processes enabled by primary cilia 35
Figure 1.3 – The multifaceted interactions of CEP290 45
Figure 1.4 – Molecular assembly and interactions of the BBSome 51
Figure 1.5 – Overall study outline that displays how biological materials 72
Figure 2.1 – Graphical representation of gene set enrichment analysis 89
Figure 2.2 – An outline of the design adopted for maintenance and differentiation of iPSCs 93
Figure 3.1 – 1% agarose gel of total genomic DNA 119
Figure 3.2 – Strategy for genetic modifiers 4-tier categorisation 121
Figure 3.3 – The location of proteins encoded by qualifying genes within the primary cilium 127
Figure 3.4 – The interacting network of BBS1 with the protein products of genes identified through 4-tier categorisation 128
Figure 3.5 – Conservation of ARL6 p.T32M 130
Table 3.10 – Burden analysis comparison with SCGSv1 gene panel 134
Figure 3.6 – The mean number of variants throughout IVA filtering 136
Figure 3.7 – Mean coverage data to assess the quality of WGS data 138
Figure 4.1 – The hierarchy for biomarker discovery 156
Figure 4.2 – Schematic representation of a quadrupole time of flight mass spectrometer 159
Figure 4.3 – The profile of a well fragmented peptide versus a low efficiency fragmentation profile 160
Figure 4.4 – Example of an ion intensity plot from Progenesis QIP 162
Figure 4.5 – Principle component analysis (PCA) of metabolomic analysis by mass spectrometry 164
Figure 4.7 – Urine proteomics QC metrics read-outs 170
Figure 4.8 – Histogram of mean relative abundances for quantified proteins from urine HPLC-MS/MS 172
Figure 4.9 – Principle component analysis (PCA) plots generated from protein quantitation data obtained from urine HPLC-MS/MS analysis 173
Figure 4.10 – Volcano plot displaying DE proteins between patient and controls (excluding BBS004) 178
Figure 4.11 – Heatmap of DE proteins with FDR<0.1 179
Figure 4.12 – PANTHER overrepresentation testing pathway analysis from urinary proteins 181
Figure 4.13 – Protein-protein interaction analysis of 39 proteins with FDR<0.1 182

Figure 4.14 – Standard curve generated for quantitation of protein extracted from plasma. 183
Figure 4.15 – SDS-PAGE gel for products from flow-through (FT), wash 1, and wash 2 of the ProteoMiner protocol 184
Figure 4.16 – The distribution of peptides from plasma over 8 fractions 185
Figure 4.17 – Histogram of mean relative abundances for quantified proteins from plasma LC-MS/MS 185
Table 4.5 – Top 10 most abundant proteins identified from mean of patient BBS010 and BBS011 plasma by LC-MS/MS 186
Figure 4.18 – Volcano plot displaying differential expression of proteins between patient and controls from plasma 190
Figure 4.19 – The number of total protein identifications from plasma for SOMAscan, untargeted mass spectrometry, and shared between both technologies 190
Figure 5.1 – Schematic representation of the centrosome's dual function linking the cell cycle with ciliogenesis 204
Figure 5.2 – Experimental and methodological design for iTRAQ mass spectrometry 207
Figure 5.4 – Increase in ciliogenesis induced by serum starvation 211
Figure 5.5 – Electropherogram traces of RNA run on the TapeStation chip 213
Figure 5.6 – Single lane electrophoresis outputs from TapeStation 214
Figure 5.7 – Standard curve generated by Bradford assay that was used for quantitation of fibroblast-derived protein 215
Figure 5.8 – MA plots displaying significantly differentially expressed transcripts 217
Figure 5.9 – Heatmap featuring results of enriched Hallmark gene sets generated by GSEA 219
Figure 5.11 – Gene expression for each gene of the leading edge for disease relevant findings from GSEA 221
Figure 5.12 – GSEA profiles for treatment comparisons 222
Figure 5.13 – Reactome pathways significantly positively or negatively enriched by PANTHER enrichment test in starved vs fed cultures 223
Figure 5.14 – Interactions between DE genes from the Reactome pathway "assembly of the primary cilium" 225
Figure 5.16 – Volcano plots displaying differentially expressed proteins (blue) for each comparison 229
Figure 5.17 – Heatmap featuring enriched Hallmark gene sets generated by GSEA with iTRAQ proteomics data 230

Figure 5.18 – GSEA profiles for enriched Hallmark gene sets related to mTOR signalling from proteomics data 231
Figure 5.19 – Reactome pathways found to be significantly positive or negatively enriched by PANTHER statistical enrichment test for starved BBS vs control comparison from iTRAQ proteomics data 232
Figure 5.20 – Protein-protein interaction network of upregulated pathways from starved BBS vs control comparison 234
Figure 5.21 – Protein-protein interaction map of the Hallmark adipogenesis gene set identified as statistically positively enriched in fed BBS vs control by GSEA of 2 omic technologies 237
Figure 6.1 – A schematic to depict the formation of the retina and lens from the optic vesicle and surface ectoderm during development 251
Figure 6.2 – Specification and organisation of the optic cup 252
Figure 6.3 – The laminated structure of the retina 253
Figure 6.4 – Schematic of a rod photoreceptor and the connecting cilia machinery 254
Figure 6.5 – Confirmation by Sanger sequencing of homozygous BBS1 p.M390R in iPSC lines derived from patients BBS001, BBS007, and BBS008 262
Figure 6.6 – Relative fold change of expression for 6 putative reference genes from optic cup samples taken over a 124-day differentiation process 263
Figure 6.7 – Gene expression of pluripotency genes from patient BBS001 samples 265
Figure 6.8 – Gene expression analysis of EFTF genes from patient BBS001 samples 266
Figure 6.9 –The progression of differentiating optic cups from patient BBS001 samples at day 28 through to day 198 268
Figure 6.10 – Gene expression of neural retina specification markers from patient BBS001 samples 269
Figure 6.11 – Gene expression analysis of phototransduction markers from patient BBS001 samples 270
Figure 6.12 – Immunohistochemical analysis of phototransduction proteins in optic cups from patient BBS001 272
Figure 6.13 A-H – Relative fold change of gene expression profiles by qPCR for key genes involved in the development of the retina 274
Figure 6.14 – Immunohistochemical analysis of phototransduction proteins in optic cups from CTRL037 and BBS008 276
Figure 6.15 – Immunohistochemical staining to observe localisation of primary cilia proteins 277

Figure 6.16 – Transmission electron microscopic analysis of optic cups at day 198 from patient BBS001 279
Figure 7.1 – Summary of findings for each chapter 290
Figure A.1 – iPSC colonies and embryoid bodies from lines used in chapter 6 300

List of tables

Table 1.1 – Proteins found in different key complexes required for IFT 32
Table 1.2 – The clinical features of the different nonmotile ciliopathies 38
Table 1.3 – The genetic heterogeneity of nonmotile ciliopathies 42
Table 1.4 – Summary of causative BBS genes 48
Table 1.5 – Types of human genetic variation that affect protein coding genes 63
Table 2.1 – Seeding densities for fibroblasts in different flasks or plates used 75
Table 2.2 – Chromatography trapping (sample loading) and elution conditions 86
Table 2.3 – The recipe for embryoid body media 94
Table 2.4 – The recipe for neural induction media 95
Table 2.5 –The recipe for retinal differentiation media 95
Table 2.6 – The recipe for retinal differentiation media with extra supplementation from FBS and taurine 96
Table 2.7 – Forward (F) and reverse (R) primer sequences 98
Table 2.8 – List of TaqMan probes and assay IDs used for qPCR 100
Table 2.9 – Software, databases, and open source software utilised within this thesis 104
Table 2.10 – Patient clinical information 110
Table 2.11 – Sample information about control individuals used within this study 111
Table 3.1 – Concentrations of BBS patient genomic DNA 118
Table 3.2 – Filtering criteria and resulting number of calls via IVA analysis 119
Table 3.3 – Putative modifying variants identified through the 4-tier categorisation strategy 125
Table 3.4 – A summary of the number of potentially modifying variants per patient that were detected at each tier 126
Table 3.5 – The candidate modifying variant for each patient with a confidence grading of high or low 130
Table 3.6 – Known genetic diagnoses for 10 of 16 patients sequenced by WES 131
Table 3.7 – Burden analysis comparison against WES patient and control datasets with ciliopathy gene panel 132
Table 3.8 – Burden analysis comparison against WGS patient and control datasets with ciliopathy gene panel 133
Table 3.9 – Burden analysis comparison against integrated datasets with ciliopathy gene panel 133
Table 3.11 – Comparison between WGS, WES, and combined datasets with a control gene panel 135

Table 4.1 – Values from sample quantitation after BCA assay of proteins extracted from urine 168
Table 4.2 – Top 10 DE proteins ranked by FDR between patients and controls (including BBS004) from urine 174
Table 4.3 – DE proteins between patients and controls (excluding BBS004) with FDR<0.1 from urine HPLC-MS/MS experiment 177
Table 4.4 – Values from sample quantitation after BCA assay of protein obtained from plasma 183
Table 4.6 – The 20 most abundant and least abundant proteins identified by SOMAscan from plasma 188
Table 4.7 – The top 25 DE proteins ranked by p-value identified from plasma by SOMAscan 189
Table 5.1 – Sample quantitation and quality control results from RNA extracted from serum fed or serum starved fibroblasts 212
Table 5.2 – Protein concentrations and yields from sample quantitation by Bradford assay 216
Table 5.3 – The number of transcripts detected as differentially expressed for each experimental comparison 217
Table 5.4 – Data for each biological group for the 10 genes from the Reactome pathway "assembly of the primary cilium" 227
Table 5.5 – The number of proteins detected as differentially expressed for each experimental comparison 229
Table 5.6 – Significantly enriched pathways or gene sets that were identified in both analysis of RNA-seq and iTRAQ 236
Table 6.1 – Table presenting patient and control iPSC line information 262
Table A.1 – Genes known to cause nonmotile ciliopathies, referred to as the ciliopathy panel 296
Table A.2 – Nonmotile cilia genes from the SYSCILIA gold standard version 1 gene panel 297
Table A.3 – Genes associated with causing Charcot-Marie-Tooth disease 297
Table A.4 – Whole genome and whole exome sequencing samples and batches used in burden analysis, with disease information 299
Table A.5 – Proteins detected by both SOMAscan and LC-MS/MS from plasma 300
Table S.1 – Supplemental data and material available on the USB stick attached to this thesis 303

Acknowledgements

There are many people that have been involved, in one way or another, in the creation of this thesis. Firstly, I'd like to thank my supervisors, Professor Phil Beales and Dr Chiara Bacchelli, for the opportunities and support they have provided to allow me to progress into a well-rounded scientist. They both have offered endless encouragement to help me develop this project. I would particularly like to thank Chiara for setting aside time for regular discussions, as well as providing mentorship throughout this process.

I would like to extend my appreciation to Dr Dan Kelberman, who acted as a mentor, role model, and friend during my PhD studentship. He has offered excellent scientific expertise, advice, and insight throughout this process, as well as providing wine and cake when times became more challenging. I'd also like to thank the members of the HIGH-5 and GOSgene groups, especially the bioinformatics team, for which this project would not have been possible. In particular, Dr Georg Otto for RNA-seq analysis, Dr Xueting Wang for proteomics analysis, Dr Janna Kenny for patient phenotyping, and Miss Jasmine Gratton for assistance in the lab. Likewise, I'm also grateful to the Cilia Defects Lab, who have offered assistance within this thesis; Dr Elizabeth Forsythe, Miss Grace Freke, and Dr Toby Collins, in particular.

Several collaborators have supported or aided this work, including UCL Genomics, UCL Biological Mass Spectrometry Centre, the HipSci project, Dr Peter Munro and Professor Mike Cheetham's lab from the UCL Institute of Ophthalmology, and Professor Sami Kaski's lab from the Helsinki Institute for Information Technology. Furthermore, I'd like to thank Dr Valentina Di Foggia from Professor Jane Sowden's lab, and Dr Nathalie Moens from Professor Fiona Watt's lab, for teaching me how to culture iPSCs.

I'd like to thank my family for their infinite support and motivation to succeed. I am grateful to my parents, who have provided encouragement and enthusiasm in everything I have done. With thanks to Colin for reading and offering excellent guidance for my chapters. Throughout this entire process, I'm incredibly thankful to Chris, who has provided limitless love and reassurance.

Finally, many thanks to my examiners for taking the time to read this thesis.

List of abbreviations

ACMG	American College of Medical Genetics
ADPKD	Autosomal dominant polycystic kidney disease
ALMS	Alström Syndrome
ALS	Amyotrophic Lateral Sclerosis
ASB-14	Amidosulfobetaine-14
AVSD	Atrioventricular septal defect
Ax	Axoneme
BAM	Binary alignment map
BB	Basal body
BBS	Bardet-Biedl Syndrome
BCA	Bicinchoninic acid
BGI	Beijing Genomics Institute
BMI	Body mass index
BR	Broad range
BRC	Biomedical Research Centre
BSA	Bovine serum albumin
CADD	Combined annotation dependent depletion
CCT	Chaperone containing TCP-1
CDK	Cyclin dependent kinase
CF	Cystic fibrosis
CFTR	Cystic fibrosis transmembrane conductance regulator
CHF	Chronic heart failure
CI	Confidence interval
CID	Collision induced dissociation
CMT	Charcot Marie Tooth disease
CPM	Counts per million
CRC	Colorectal cancer
CRP	C reactive protein
CT	Ciliary tip
CVD	Cardiovascular disease
DE	Differentially expressed
DHH	Desert Hedgehog
DMEM	Dulbecco's modified eagle medium
DMSO	Dimethyl sulfoxide
DTE	Dithioerythritol
E8	Essential 8
EB	Embryoid body
EBM	Embryoid body media
EDTA	Ethylenediaminetraacetic acid
EFTF	Eye field transcription factor
ELISA	Enzyme linked immunosorbant assay

eQTL	Expression trait loci
ER	Endoplasmic reticulum
ES	Enrichment score
ESC	Embryonic stem cell
ESI	Electrospray ionisation
ESP	Exome sequencing project
EURO-WABB	EU rare diseases registry for Wolfram, Alström, and Bardet-Biedl Syndromes
EURORDIS	Rare Diseases Europe
EV	Extracellular vesicle
EVC	Ellis Van Creveld Syndrome
ExAC	Exome aggregate consortium
FBS	Fetal bovine serum
FC	Fold change
FDR	False detection rate
FT	Flow through
Fwk	Fetal week
GATK	Genome analysis toolkit
gDNA	Genomic DNA
GEF	Guanine nucleotide exchange factor
GnomAD	Genome aggregation database
GnRH	Gonadotropin-releasing hormone
GOSH	Great Ormond Street Hospital
GPCR	G-protein coupled receptor
GRCh	Genome Reference Consortium human
GSEA	Gene set enrichment analysis
HAUS	Homologous to Augmin subunits
HbA1c	Glycated haemoglobin
HDL	High density lipoprotein
HDMS	High definition mass spectrometry
HGMD	Human Gene Mutation Database
HGP	Human genome project
HH	Hedgehog
HipSci	Human induced pluripotent stem cell initiative
HPLC	High performance liquid chromatography
IAA	Iodoacetamide
ICH	Institute of Child Health
IF	Intermediate filaments
IFT	Intraflagellar transport
IgG	Immunoglobulin G
IGV	Integrative Genome Viewer
IHH	Indian Hedgehog
IMS	Ion mobility separation
Indel	Insertion or deletion

iPOP	Integrative personal omics profile
iPSC	Induced pluripotent stem cell
iTRAQ	Isobaric tag for relative and absolute quantitation
IVA	Ingenuity Variant Analysis
JATD	Jeune's asphyxiating thoracic dystrophy
JBTS	Joubert Syndrome
KSR	Knockout serum replacement
LC-MS	Liquid chromatography mass spectrometry
LCA	Leber Congenital Amaurosis
LN2	Liquid nitrogen
lncRNA	Long non-coding RNA
m/z	Mass to charge ratio
MAF	Minor allele frequency
MALDI	Matrix-assisted laser detection/ionisation
MEF	Mouse epithelial fibroblasts
miRNA	Micro RNA
MKKS	McKusick-Kaufman Syndrome
MKS	Meckel-Gruber Syndrome
MO	Morpholino oligonucleotide
MRI	Magnetic resonance imaging
mRNA	Messenger RNA
MS/MS	Tandem mass spectrometry
MSC	Mesenchymal stem cell
MSigDB	Molecular signatures database
MTOC	Microtubule organising centre
MZ	Monozygotic twin
NES	Normalised enrichment score
NGS	Next generating sequencing
NHBLI	National Heart, Lung, and Blood Institute
NIHR	National Institute of Health Research
NIM	Neural induction media
NMD	Nonsense mediate decay
NPHP	Nephronophthisis
NR	Neural retina
OCT	Optimal cutting temperature
OFD	Oro-Facial-Digital Syndrome
OMIM	Online Mendelian Inheritance in Man
PAGE	Polyacrylamide gel electrophoresis
PANTHER	Protein analysis through evolutionary relationships
PBST	PBS + 0.1% tween
PC	Principle component
PCA	Principle component analysis
PCM	Pericentriolar material

PCP	Planar cell polarity
PDM	Primary driving mutation
PEDF	Pigment epithelium derived factor
Penstrep	Penicillin-Streptomycin
PFA	Paraformaldehyde
PMSF	Phenylmethane sulfonyl fluoride
Progenesis QIP	Progenesis QI for Proteomics
PROVEAN	Protein variation effect analyser
PTM	Posttranslational modification
Q	Quadrupole
QC	Quality control
qPCR	Quantitative real time PCR
RA	Retinoic acid
RDM	Retinal differentiation media
RFC	Relative fold change
RFU	Relative fluorescent unit
RIN	RNA integrity number
RNA-seq	RNA sequencing
RNAi	RNA interference
RP	Retinitis pigmentosa
RPE	Retinal pigmented epithelium
RPKM	Reads per kilobase million
RQ	Relative quantitation
rRNA	Ribosomal RNA
RT	Reverse transcriptase
RT-PCR	Reverse transcriptase PCR
SAG	Smoothened agonist
SAM	Sequence alignment map
SCGSv1	SYSCILIA gold standard version 1
SDS	Sodium docecyl sulphate
SELEX	Systematic evolution of ligands by exponential enrichment
SHH	Sonic hedgehog
SIFT	Sorting intolerant from tolerant
siRNA	Small interfering RNA
SLS	Senior-Løken Syndrome
Smo	Smoothened
snRNA	Small nuclear RNA
SNV	Single nucleotide variant
SOP	Standard operating protocol
SRS	Silver Russell Syndrome
TBE	Tris/Borate/EDTA
TEM	Transmission electron microscope
TFA	Trifluoroacetic acid

ToF	Time of flight
TPR	Tetratricopeptide repeat
TRiC	TCP-1 Ring Complex
tRNA	Transfer RNA
TZ	Transition zone
UCSC	University of California, Santa Cruz
UDG	Uracil-DNA glycosylase
UPLC	Ultra performance liquid chromatography
UPR	Unfolded protein response
USH	Usher Syndrome
UTI	Urinary tract infection
UV	Ultraviolet
VPA	Valproic acid
VEO-IBD	Very Early Onset-Inflammatory Bowel Disease
WES	Whole exome sequencing
WGS	Whole genome sequencing

Chapter 1: Introduction

1.1 Eukaryotic cilia

Cilia are evolutionarily conserved microtubule-based organelles present on eukaryotic cells. These microtubules stretch from the cell body and can act as cellular signalling antennae, beat in waves to encourage movement of fluids, or have critical transport roles in highly specialised cells. The microtubule structures are robust and stable due to acetylation of tubulin monomers. There are 3 types of cilia characterised by different microtubule arrangements, which determine their function (**Figure 1.1B**).

Motile cilia are organised in a 9 + 2 arrangement, with a central microtubule pair surrounded by 9 outer microtubule doublets. Dynein arms and radial spokes protrude from regular intervals along the microtubule doublets. Hydrolysis of ATP allows the dynein arms to drive a sliding motion between the microtubule doublet relative to the central pair, causing the motile cilia to beat (King 2016). Motile cilia are found in large numbers on the apical surface of specialised cells and function in coordinated rolling waves. For example, they beat away secretions in the respiratory epithelium and middle ear, cerebrospinal fluid in the brain, or encourage movement of sperm in the fallopian tubes (Mitchison and Valente 2017).

A second form of cilia are located at the embryonic node, thus are known as nodal cilia. In contrast to motile cilia, they lack a central microtubule pair, demonstrating a 9 + 0 arrangement. With one cilium per cell, they use the retention between dynein arms and microtubules to generate a rotatory motion (Szymanska and Johnson 2012). This mediates a leftward flow of morphogens during development to establish left-right patterning (Wagner and Yost 2000; Praetorius and Spring 2005).

The final form are primary cilia, also known as sensory cilia (**figure 1.1A**). Previously considered a vestigial accessory to the cell (Webber and Lee 1975), primary cilia are now known to be a crucial cellular signalling hub in postmitotic cells. They display a 9 + 0 arrangement, but lack dynein machinery rendering them stationary. They are present in the single form on all non-dividing eukaryotic cells, with the exception of bone marrow derived cells (Praetorius and Spring 2005). They are thought to have many tissue specific roles, including protein trafficking, mechanosensation for external chemical stimuli, and acting as an interface for ligand-receptor interactions (Wolfrum and Schmitt 2000; Lee et al. 2015; Berbari et al. 2008). For the purpose of

this thesis, the anatomy and molecular function of the primary cilium shall be explored in more detail.

1.2 The core of the primary cilium

The centrosome is an organelle comprised of barrel-shaped mother and daughter centrioles, which are surrounded by an unstructured protein mass, called pericentriolar material (PCM). The centrioles act as a microtubule organising centre (MTOC) throughout the cell cycle. During the early G_1 phase, the centrioles migrate to the apical plasma membrane where the mother subsequently forms the basal body of the primary cilium. The basal body displays a "pin-wheel" 9 x 3 + 0 microtubular structure anchored by transition fibres to the plasma membrane and is the nucleation point of the 9 microtubule doublets, known collectively as the axoneme (**figure 1.1**) (Reiter et al. 2012). The transition fibres have a dual function, as they also serve as docking sites for transport particles (Wei et al. 2015). At the proximal end of the basal body, the transition zone forms in the early stages of ciliogenesis. The transition zone is characterised by Y-linking proteins that attach the axoneme to the ciliary membrane and act as a membrane diffusion barrier, referred to as the ciliary gate (Reiter et al. 2012; Garcia-Gonzalo et al. 2011). This compartmentalisation allows regulated entry and exit between the cytoplasm and the cilium, and instigates an exclusive signalling environment. The axoneme extends from the basal body via ciliogenesis. The primary cilium is topologically continuous with the cell membrane, allowing movement of membrane associated proteins from the cell body to the ciliary tip.

1.3 Intraflagellar transport

The ciliary axoneme extends perpendicular from the cell body and can vary in length, from 1µm up to 9µm (Dummer et al. 2016). The extension of the axoneme requires transportation of ~600-1000 distinct proteins from the endoplasmic reticulum (ER) and Golgi apparatus to the nascent ciliary tip (Lechtreck 2015). The polarised nature of tubulin heterodimers allows motor proteins, namely kinesin and dynein, to deliver proteins up and down the organelle in a process called intraflagellar transport (IFT). IFT is a highly conserved bi-directional movement of cargo within the ciliary membrane. As the cilium is devoid of protein synthesis apparatus, robust anterograde and retrograde transport (away and towards the cell body, respectively) of proteins by IFT machinery is required for ciliogenesis, cilium maintenance, and transfer of signalling cargo to the ciliary tip.

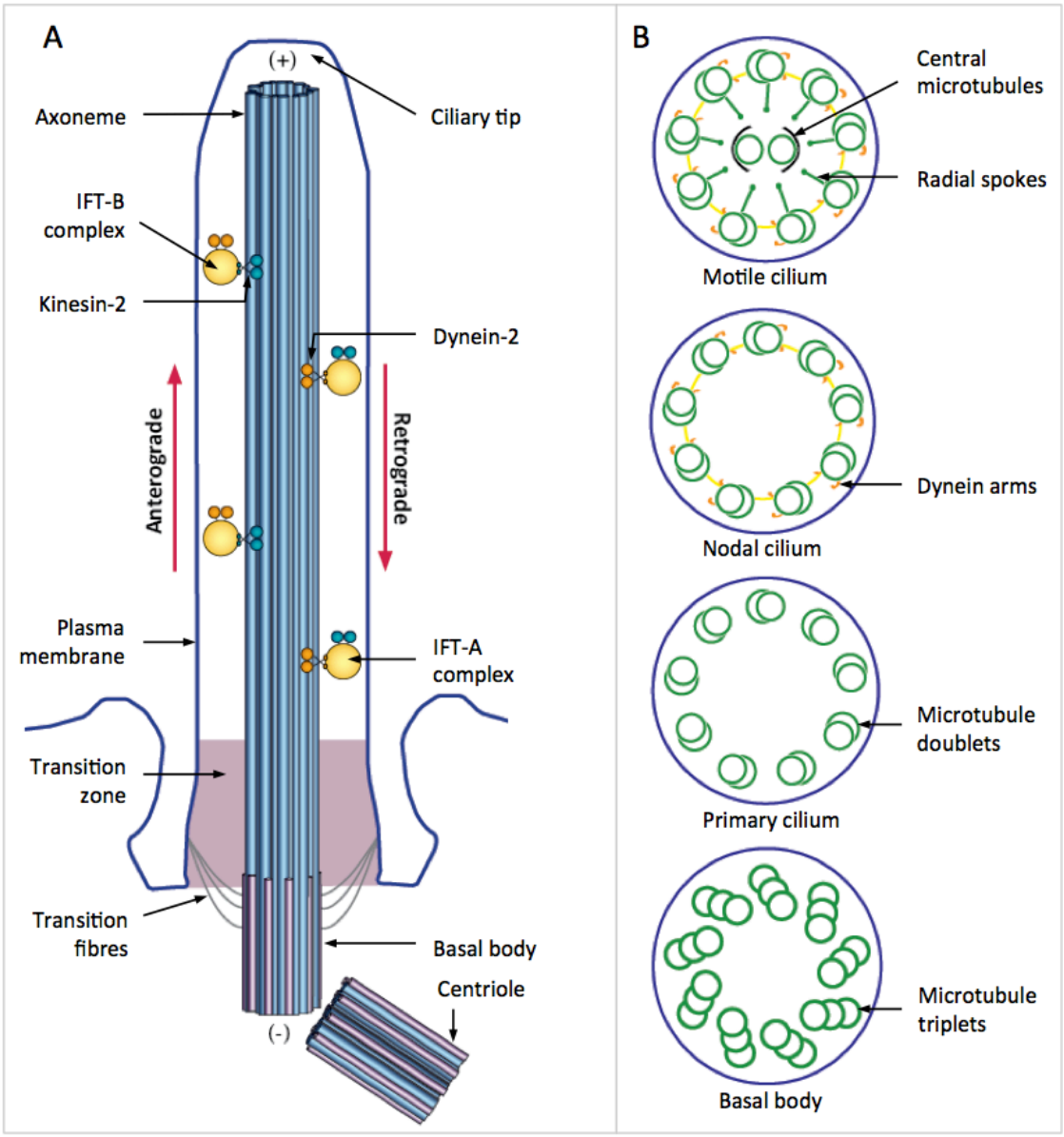

Figure 1.1 – Cross sections of cilia. A) The primary cilium is an extension of the plasma membrane that projects from the cell body. 9 microtubule doublets nucleate from the basal body, acting as the scaffold for intraflagellar transport. B) Horizontal cross section of different microtubule patterns in cilia and the basal body.

Proteins KIF3A, KIF3B, and KAP3 form the kinesin-2 motor complex, which specifically travels towards the plus-ended ciliary tip. The kinesin motor associates with IFT machinery, made up of IFT-A and IFT-B complexes (**table 1.1**), and an inactive cytoplasmic dynein motor (Pedersen et al. 2008). Anterograde protein cargo is thought to specifically bind to a loading site on IFT-B (Lechtreck 2015). At the ciliary tip, the cargo is released for use in ciliogenesis or signal transduction, and IFT machinery is turned over for transport back to the cell body. The retrograde specific IFT-A particle functions to recycle cargo and anterograde machinery to the base of the

cilium. Retrograde transport is also crucial for the conserved Sonic hedgehog (SHH) signalling pathway (Huangfu and Anderson 2005). IFT machinery turnover activates dynein-2, consisting of DYNC2H1 and DYNC2LI1, which specifically returns along the minus direction of the microtubules (Pazour et al. 1998).

It is thought that cargo entry into the cilium is mediated by the BBSome, a complex derived from 8 evolutionarily conserved proteins (BBS1, BBS2, BBS4, BBS5, BBS7, TTC8/BBS8, BBS9, BBIP10/BBS18) (Nachury et al. 2007; Loktev et al. 2008; Jin et al. 2010). All BBSome proteins have separately been implicated in causing Bardet-Biedl Syndrome. It was proposed that the BBSome was implicated in IFT after it was found to move at the same rate as the IFT machinery through the cilium (Blacque et al. 2004). Since then, it was shown to have a crucial role in IFT turnover, as inactivity of the BBSome resulted in accumulation of IFT-B complex at the ciliary tip (Wei et al. 2012). It has been proposed that BBSome proteins, BBS7 and BBS8, act as a cohesive factor required for IFT machinery assembly (Ou et al. 2005) and are required for proficient IFT function within the cilium (Wei et al. 2012). The BBSome also has a role in recruitment of ciliogenesis or signalling cargo to the base of the cilium for trafficking. Via interaction with small GTPases, such as RAB, ARF, and ARL families, the BBSome is recruited to the ciliary membrane. Here, the BBSome was shown to interact with $RAB8^{GTP}$ and RABIN8, which was proposed to promote docking and fusion of vesicles carrying cargo from the Golgi to the cilium base (Nachury et al. 2007; Loktev et al. 2008).

IFT-A	IFT-B	Kinesin-2	Dynein-2	BBSome
IFT43	IFT20	KIF3A	DYNC2H1	BBS1
IFT122A	IFT27	KIF3B	DYNC2LI1	BBS2
IFT122B	IFT46	KAP3		BBS4
IFT139	IFT52			BBS5
IFT140	IFT57			BBS7
IFT144	IFT72			TTC8/BBS8
	IFT74			BBS9
	IFT80			BBIP10/BBS18
	IFT81			
	IFT88			
	IFT172			

Table 1.1 – Proteins found in different key complexes required for IFT

1.4 Key processes enabled by primary cilia

The primary cilium is responsible for coordinating a plethora of core cellular processes. Cilia can have various tissue specific functions on different specialised cells. In essence, they act as cellular signalling antennae. Through IFT, as many as 1000 proteins are trafficked to the apical ciliary membrane (Liu et al. 2007; Gherman et al. 2006) and translation of stimuli is facilitated into downstream intracellular signalling cascades. Due to the compartmentalisation of the primary cilium, it can be ensured that signalling is well orchestrated, regulated and highly sensitive. Their responsibility is broad, regulating cellular events including polarity, migration, differentiation and tissue morphology (Basten and Giles 2013; Singla and Reiter 2006; Ishikawa and Marshall 2011).

1.4.1 G-Protein Coupled Receptors

The cilium creates an efficient appendage for G-protein coupled receptor (GPCR) localisation (figure 1.2A). GPCRs are well characterised transmembrane receptors, constructed from a tertiary structure of 7 transmembrane spanning alpha helices. Briefly, signal transduction occurs by detection of a sensory input by the GPCR N-terminal extracellular domain. This initiates an activation of the cytosolic G-protein by exchange of binding from GDP to GTP. The activation by GTP dissociates a subunit from the G-protein trimer, which prompts downstream signalling through cAMP transduction (Schou et al. 2015).

There are approximately 800 different GPCRs in humans, each activated by a vast range of agonists, including peptides, lipids, nucleotides, ions, odorants and photons (Milligan and Kostenis 2006; Schou et al. 2015). At least 25 GPCRs are known to localise to the primary cilium to date (Hilgendorf et al. 2016), the utilities of which span many cellular processes. These include phototransduction by opsins (Sun et al. 2012; Masyuk et al. 2008), olfactory function (Lowe and Gold 1991), energy homeostasis and control of food consumption (Händel et al. 1999; Loktev and Jackson 2013), neuronal processes (Omori et al. 2015), and reproductive function (Koemeter-Cox et al. 2014). It is thought that the BBSome directly interacts with components of some GPCRs (Berbari et al. 2008; Domire et al. 2011; Jin et al. 2010). Knockout studies in mice have displayed aberrant localisation and signalling of the somatostatin receptor, Sstr3 (Händel et al. 1999), melanin-concentrating hormone receptor, Mchr1 (Berbari et al. 2008) and a neuropeptide receptor, Npy2R (Loktev and Jackson 2013). Also, BBSome knockdown in murine neuronal primary cilia has demonstrated increased dopamine accumulation, though the localisation of the dopamine receptor was not affected (Domire et al. 2011).

1.4.2 Hedgehog signalling

Efficient IFT within the primary cilium is vital for key signal transduction pathways, including the hedgehog (HH) pathways. HH is required for successful mammalian embryonic development for organogenesis and development of the body axis, as well as regulation of adult stem cells in tissue maintenance and regeneration (Petrova and Joyner 2014). There are 3 known HH ligands, namely Sonic hedgehog (SHH), Indian hedgehog (IHH), or Desert hedgehog (DHH), where SHH is the best characterised and thus is more commonly studied with respect to the primary cilium. The first connection between the primary cilium and HH signalling was acquired through mutagenesis screens in mouse embryos. Disruption of two components of the IFT-B complex, *Ift172* and *Ift88*, resulted in a marked decrease in downstream *Shh* mediators, as well as aberrant *Shh*-dependent development of the neural tube (Huangfu et al. 2003). Further investigations disrupted both retrograde and anterograde complexes and determined that although Shh ligand binding was unaffected, downstream mediators were aberrantly expressed, including the HH receptor *Ptch1* and pathway effectors Gli transcription factors (Liu et al. 2005; Huangfu and Anderson 2005). These experiments suggested that the primary cilium is a structural requirement for robust HH signalling (**figure 1.2B**).

1.4.3 Mechanosensation

Despite their rigidity enforced by a strong microtubule backbone, the primary cilium detects physical stimuli by movement under fluid flow or pressure, called mechanosensation. Studies have focussed on the link between mechanosensation and cilia that protrude from epithelial cells of the renal tubular lumen (Nauli and Zhou 2004; Nauli et al. 2003). In this tissue specific role, bending of cilia by fluid flow regulates proliferation and differentiation of renal cells by activating an intracellular Ca^{2+} signalling cascade (Praetorius and Spring 2001).

The polycystin proteins, PC1 and PC2, are encoded by the *PKD1* and *PKD2* genes respectively, which are mutated in autosomal dominant polycystic kidney disease (ADPKD) (Boucher and Sandford 2004). Together, these proteins heterodimerise to form a receptor channel complex, located at the ciliary membrane of renal epithelial cells (Nauli et al. 2003). It was determined that shear stress by fluid flow triggers a PC1/PC2 dependent influx of Ca^{2+} into the primary cilium. This increase in intracellular Ca^{2+} concentration can lead to a variety of consequences, including cell morphology, differentiation, growth, and apoptosis. Furthermore, it is thought that Ca^{2+} influx switches on non-canonical Wnt signalling (Happé et al. 2011) (**figure 1.2C**).

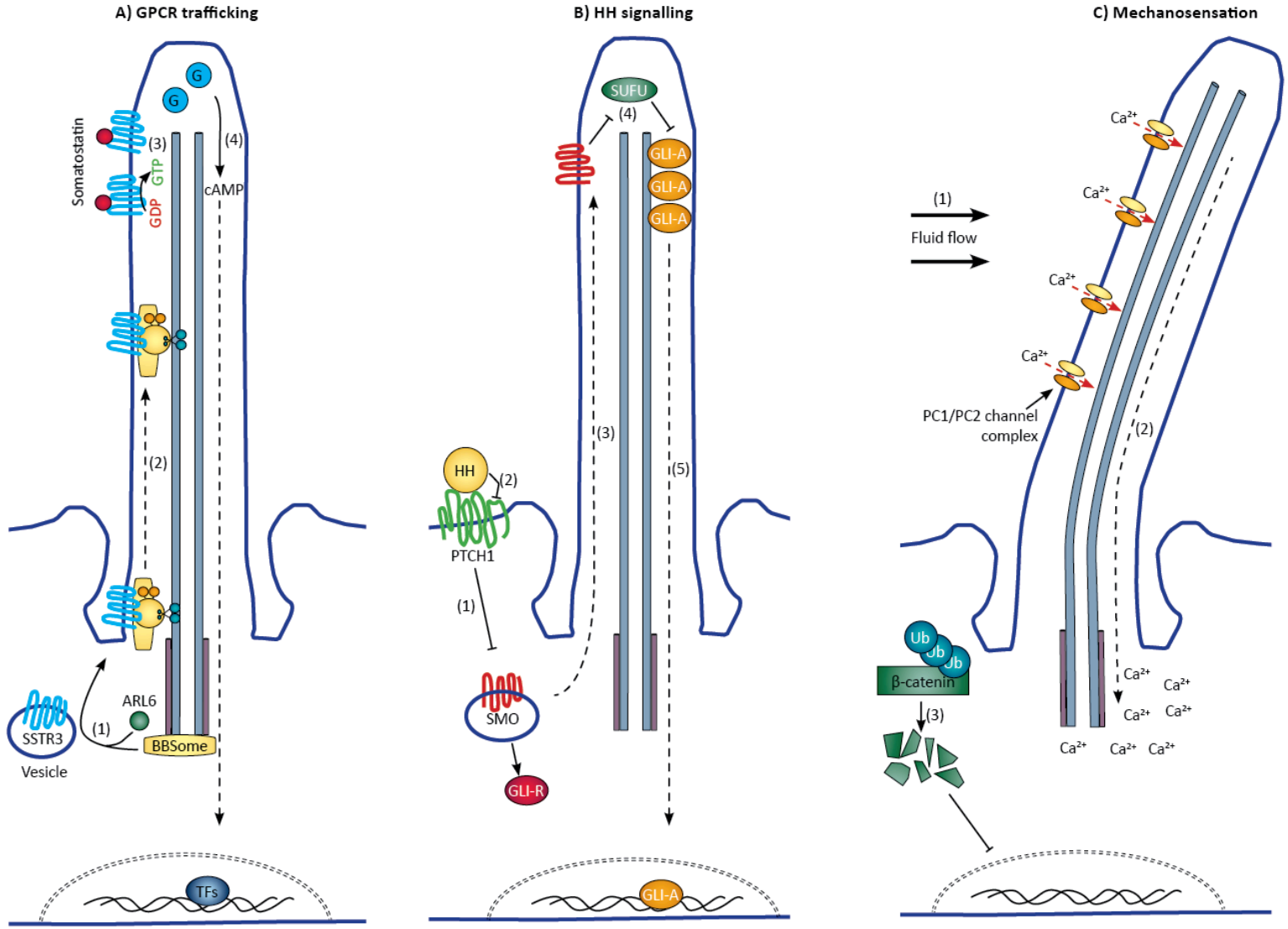

Figure 1.2 – Examples of key processes enabled by primary cilia (legend on next page)

Figure 1.2 (on previous page) – A) Cilia specific GPCR trafficking: (1) The BBSome recruits the cilia associated GPCR, such as SSTR3, to the ciliary membrane via interaction with ARL6 (BBS3). (2) The receptor is trafficked to the ciliary tip by interaction with the anterograde IFT-B complex. (3) The ligand, such as somatostatin, induces a conformational change upon binding, converting GDP to GTP and dissociating the G protein. (4) Release of G proteins activates downstream pathways, for example, through the cAMP-dependent signalling.
B) Hedgehog signalling: (1) Localisation of PTCH1 at the plasma membrane sequestered SMO in the cytoplasm, allowing the repressed GLI form (GLI-R) to be processed. (2) HH binds to PTCH1, quenching its suppression on SMO. (3) SMO is free to be trafficked to the ciliary tip, where it inhibits SUFU. (4) Suppression of SUFU allows accumulation of GLI at the ciliary tip, which is processed to the active form (GLI-A). (5) GLI-A translocates to the nucleus where it activates HH pathways.
C) Mechanosensation: (1) Detection of extracellular fluid flow induces influx of calcium through PC1/PC2 receptor channel complex. (2) Calcium translocates and accumulates in the cytosol, which activates subcellular pathways to prevent pro-proliferative signalling. (3) Canonical Wnt signalling is prevented as β-catenin is sequestered in the cytosol where it is ubiquitinated for proteasomal degradation.

1.4.4 Wnt signalling and planar cell polarity

Similar to HH signalling, Wnt signalling is an evolutionarily conserved pathway and in some circumstances, may be mediated or organised by the primary cilium. Canonical Wnt signalling, non-canonical planar cell polarity (PCP) and non-canonical Wnt/Ca^{2+} pathways have been characterised. The canonical pathway is pro-proliferative, whereas non-canonical pathways are responsible for differentiation, homeostasis, and polarisation of tissues (Wallingford and Mitchell 2011). Canonical Wnt signalling is β-catenin dependent, while non-canonical signalling occurs in the absence of β-catenin. Although the primary cilium is not essential for Wnt signalling (Eggenschwiler and Anderson 2007), it has been suggested that the cilium plays a role in restricting the canonical pathway (Gerdes et al. 2007), and orchestrates a cellular switch towards non-canonical Wnt pathways (Basten and Giles 2013; Simons et al. 2005; Ross et al. 2005).

The downstream effect of canonical Wnt signalling accumulates β-catenin in the cytosol, which subsequently translocates to the nucleus and activates pro-proliferation transcription factors. In the absence of Wnt ligand, β-catenin is sequestered in the cytosol and targeted for degradation. As a result, a cellular switch drives cellular differentiation and PCP. The primary cilium is thought to act as a closed system for non-canonical Wnt signalling, where interacting proteins are isolated to prevent the accumulation of β-catenin (Gerdes et al. 2007; Simons et al. 2005).

1.5 The nonmotile ciliopathies

Due to its pivotal role in development, cellular function and orchestration of key signalling pathways, defective primary cilia can lead to a number of disabling rare diseases, collectively termed ciliopathies. At least 73 genes have been uncovered so far, which are implicated in over 20 different ciliopathies of primary cilia (Rachel, Li, et al. 2012; Waters and Beales 2011). These numbers are likely to expand, as new genes are being diagnosed and ciliopathies proposed. For example, it has been suggested that amyotrophic lateral sclerosis (ALS) (Ma et al. 2011), Lowe syndrome (Coon et al. 2012), 3M syndrome, and Hermansky-Pudlak syndrome may be new ciliopathies (Boldt et al. 2016). The well-defined ciliopathies are Alström syndrome (ALMS), Bardet-Biedl Syndrome (BBS), Ellis van Creveld syndrome (EVC), Jeune's asphyxiating thoracic dystrophy (JATD), Joubert Syndrome (JBTS), Oro-Facial-Digital syndrome (OFD), McKusick-Kaufman Syndrome (MKKS), Meckel-Gruber Syndrome (MKS), Nephronophthisis (NPHP), Leber Congenital Amaurosis (LCA), Senior-Løken Syndrome (SLS), and Usher syndrome (USH). Nonmotile ciliopathies can be generally classified into three groups: syndromic nonskeletal, syndromic skeletal, and nonsyndromic ciliopathies (table 1.2).

1.5.1 Syndromic nonskeletal ciliopathies

Alström syndrome (ALMS) is an autosomal recessive disease, characterised by retinal dystrophy, childhood obesity and neurosensory hearing loss. Due to a strong phenotypic overlap to BBS, including secondary features of renal disease, hypogonadism in males and type II diabetes, it can often be misdiagnosed (Tobin and Beales 2009). However, it is genetically distinct from BBS as a single gene, *ALMS1,* accounts for all diagnoses (Collin et al. 2002).

Bardet-Biedl Syndrome (BBS) is an autosomal recessive disorder, characterised by retinal dystrophy, obesity, renal dysfunction, polydactyly, genital abnormalities, and learning difficulties. Currently 21 genes are thought to cause the syndrome, making it one of the most genetically heterogeneous ciliopathies. As the subject of this thesis, BBS is comprehensively described in section 1.8.

McKusick-Kaufman Syndrome (MKKS) is an autosomal recessive disorder resulting in genital abnormalities, particularly hydrometrocolpos in females, polydactyly, and rarely, features of cardiac defects. As with ALMS, there can often be diagnostic overlap with BBS in early childhood. The key distinguishing feature is lack of retinal dystrophy, learning problems, renal dysfunction, or obesity (Elise Schaefer et al. 2011). The *MKKS* gene accounts for all known cases of the

syndrome. Shortly after determining its causative role in MKKS (Stone et al. 2000), it was identified as the 6th locus to be mutated in BBS (*BBS6*) (Katsanis et al. 2000).

	Syndromic nonskeletal ciliopathies							Syndromic skeletal ciliopathies			Nonsyndromic ciliopathies	
Phenotype	ALMS	BBS	MKKS	MKS	JBTS	SLS	USH	EVC	JATD	OFD	NPHP	LCA
Retinal dystrophy	1	1			2	1	1		2	2		1
Renal dysfunction	2	1		1	2	1			2	2	1	
Polydactyly		1	1	1	2			1	1	1		
Learning difficulties		1			1					2		
Structural CNS abnormalities				1	1				2	2		
Severe skeletal abnormalities								1	1	1		
Genital abnormalities	2	1	1	2								
Obesity	1	1										
Hearing loss	1	2					1					

Table 1.2 – The clinical features of the different nonmotile ciliopathies, demonstrating the large phenotypic overlap between disorders. Primary diagnostic features are marked with 1, whereas 2 denotes secondary or less commonly associated phenotypes. Data adapted from Mockel et al. 2011 and Online Mendelian Inheritance in Man (OMIM).

Joubert Syndrome (JBTS) is an autosomal recessive disorder characterised by a hallmark "molar tooth sign" from cranial magnetic resonance imaging (MRI) indicative of brain stem and cerebellar malformations. Additional neurological symptoms include ataxia, hypoplasia of the cerebellar vermis, developmental delay, irregular breathing, and oculomotor apraxia (Tobin and Beales 2009). Secondary features have some overlap with other ciliopathies, such as retinal dystrophy, renal dysfunction and polydactyly (**table 1.2**). JBTS is highly heterogeneous with at least 17 genes accounting for diagnoses (**table 1.3**). ~50% of all cases arise from loss of function mutations (nonsense, splice-site disrupting or frameshift variants) in *CEP290* (Brancati et al. 2007), a gene mutated in 6 other ciliopathies.

Meckel-Gruber Syndrome (MKS) is a severe autosomal recessive disorder, which commonly results in embryonic or neonatal lethality due to cilia dysfunction during development. Cases are characterised by severe renal failure, encephalocoele or other structural CNS malformations, hepatic abnormalities and polydactyly. Some cases report a characteristic sloping forehead,

bowing of the limbs, and underdeveloped genitalia (Salonen 1984). With both syndromes displaying severe CNS defects, there is clinical and genetic overlap between MKS and JBTS; 8 out of 11 MKS genes are shared with JBTS (table 1.3). The first identified MKS gene, *MKS1*, was implicated as the 13th identified BBS gene (*BBS13*) (Leitch et al. 2008).

Senior-Løken Syndrome (SLS) is an autosomal recessive disease characterised by the combinatory features of two distinct ciliopathies, cystic kidneys of NPHP and retinal dystrophy of LCA. As a result, SLS has notable genetic overlap with NPHP and LCA (Omran et al. 2002) (table 1.3).

Usher syndrome (USH) is an autosomal recessive condition resulting in early onset retinal dystrophy and congenital sensorineural hearing defects. In Usher syndrome type I, caused by mutations in *MYO7A*, *USH1C*, *CDH23*, *PCDH15,* or *USH1G*, unintelligible speech and constant vestibular dysfunction are also observed. It is a relatively recent addition to the ciliopathy spectrum (Sorusch et al. 2014). Usher proteins are thought to have tissue specific roles, both mediating trafficking in the connecting cilium of photoreceptors, and acting as structural supports in stereocilia of cochlear hair cells (Cosgrove and Zallocchi 2014).

1.5.2 Skeletal ciliopathies

Jeune's asphyxiating thoracic dystrophy (JATD) is a severe autosomal recessive ciliopathy, characterised by short ribs, narrowed thoracic cage, and shortening of bones. Polydactyly is commonly an additional feature, as well as secondary phenotypes including renal, liver, or retinal dysfunction. Mortality commonly occurs within the first few years of life, owing to respiratory constriction from pulmonary hypoplasia (Tobin and Beales 2009). JATD is a genetically heterogeneous disorder, where genes mutated are involved in intraflagellar transport (table 1.3). The skeletal abnormalities indicative of JATD are thought to arise from aberrant SHH signalling caused by defective IFT (Haycraft et al. 2007).

Ellis van Creveld syndrome (EVC) is an autosomal recessive skeletal dysplasia, displaying phenotypic overlap with JATD. Features include short limbs, polydactyly, short ribs, and developmental delay, while ectodermal and cardiac defects distinguish the syndrome from JATD (Waters and Beales 2011). Two genes, *EVC* and *EVC2*, have been uncovered to cause the syndrome, which both localise to the primary cilium. The molecular distinction between EVC and JATD may be attributed to defects in Indian hedgehog signalling, rather than SHH disrupted in JATD (Ruiz-Perez et al. 2007).

Oro-Facial-Digital syndrome (OFD) is an X-linked dominant disorder, characterised by developmental malformations of the oral cavity, face, and digits. Commonly these features are accompanied by polycystic kidneys, structural CNS abnormalities, and learning difficulties (Waters and Beales 2011). A single gene, *OFD1*, has been associated with the disorder, also mutated in an X-linked recessive form of JBTS (Coene et al. 2009).

1.5.3 Nonsyndromic ciliopathies

Leber Congenital Amaurosis (LCA) is an autosomal recessive retinopathy, characterised by severe retinal degeneration, pigmentation (retinitis pigmentosa (RP)), and nystagmus. Degeneration usually presents within the first year of life, but patients have been reported to have profound retinal dysfunction from birth (Chung and Traboulsi 2009). LCA is genetically heterogeneous, affecting proteins localising only to the basal body or transition zone (table 1.3). There is some genetic overlap with BBS, NPHP, and SLS, which may be explained by mutations affecting LCA protein function in a photoreceptor specific manner (Parfitt et al. 2016).

Nephronophthisis (NPHP) is an autosomal recessive nephropathy, characterised by early onset renal cysts and tubulointerstitial fibrosis. Additional renal features may include polyuria, polydipsia, secondary enuresis, and anaemia. In ~10% of cases, extrarenal symptoms have been described, but these are associated with diagnostic features of other ciliopathies such as retinal dystrophy (SLS), skeletal dysplasia (JATD), or hypoplasia of the cerebellar vermis (JBTS) (Parisi et al. 2004; Tory et al. 2007; Hoefele et al. 2005; E Schaefer et al. 2011). NPHP is genetically heterogeneous, with 8 out of 13 proteins localising the transition zone of the primary cilium and a distinct lack of causative basal body proteins (table 1.3). There is genetic overlap with other ciliopathies with nephropathy features, such as BBS, SLS, MKS and JBTS.

1.6 A spectrum of overlapping heterogeneous disorders

The ciliopathies are often described as a spectrum of disorders, as disruption of a single organelle exhibits a range of severities and phenotypes. Ciliopathies span from severe multisystem disorders resulting in embryonic lethality (MKKS) to nonsyndromic retinal degeneration that occurs late in life (LCA) (Azari et al. 2006; Estrada-Cuzcano et al. 2012). To complicate the model further, there is profound clinical heterogeneity within each disorder, which may be driven by secondary modifying mutations (Ramsbottom et al. 2015; Kousi and Katsanis 2015; Badano et al. 2006). The confines of each disease can be challenging to diagnose, as often intellectual disability,

retinal defects, digit malformations, and renal or hepatic dysfunction can present in all ciliopathies (Gerdes et al. 2009).

Genes	BBS	JATD	JBTS	MKKS	MKS	NPHP	LCA	SLS	OFD	Location
BBS1	✓									BB
BBS2	✓						✓			
ARL6 (BBS3)	✓						✓			
BBS4	✓									
BBS5	✓									
MKKS (BBS6)	✓			✓						
BBS7	✓									
TTC8 (BBS8)	✓						✓			
BBS9	✓									
BBS10	✓									
BBS12	✓									
MKS1 (BBS13)	✓		✓		✓					
LZTFL1 (BBS17)	✓									
BBIP10 (BBS18)	✓									
C8ORF37 (BBS21)	✓						✓			
LCA5							✓			
RP2							✓			
OFD1			✓						✓	
TMEM216			✓		✓					
ARL13B			✓							
AHI1			✓							
TRIM32 (BBS11)	✓									IF
CEP290 (BBS14)	✓		✓		✓	✓	✓	✓		TZ
WDPCP (BBS15)	✓				✓					
SDCCAG8 (BBS16)	✓					✓		✓		
IQCB1						✓	✓	✓		
RPGRIP1L			✓		✓	✓	✓			
RPGR							✓	✓		
RP1							✓			
RPGRIP1							✓			
NPHP4						✓		✓		
NEK8						✓				
NPHP3					✓	✓				
B9D2					✓					
NPHP1			✓			✓		✓		
CC2D2A			✓		✓					
B9D1			✓		✓					
TCTN2			✓		✓					
TCTN3			✓						✓	
TCTN1			✓							

TMEM67			✓		✓	✓				Ax
INPP5E			✓							
INVS						✓				
IFT27 (BBS19)	✓									IFT
IFT172 (BBS20)	✓									
TTC21B		✓	✓			✓				
WDR19		✓				✓		✓		
WDR35		✓								
DYNC2H1		✓								
IFT80		✓								
KIF7			✓							CT
GLIS2						✓				

Table 1.3 – The genetic heterogeneity of nonmotile ciliopathies, both within a single disorder and across different ciliopathies. Genes are listed by localisation within primary cilium, where BB = basal body, IF = intermediate filaments, TZ = transition zone, Ax=Axoneme, IFT= intraflagellar transport machinery, CT = ciliary tip. Table adapted from Rachel et al. 2012 and localisation taken from SYSCILIA gold standard version 1 (SCGSV1) database.

The spectrum also exists at a molecular level. ALMS and MKKS are molecularly defined by mutations in a single gene, *ALMS1* and *MKKS* (*BBS6*), respectively. On the other hand, BBS is caused by pathogenic variants across at least 21 genes (**table 1.3**). A single gene is also capable of manifesting multiple syndromes, for example, *CEP290* (BBS, JBTS, MKS, NPHP, LCA or SLS), *RPGRIP1L* (JBTS, MKS, NPHP or LCA) and *TTC21B* (JATD, JBTS or NPHP). In these cases, the way the genetic variant affects the protein function plays a role in determining the phenotypic effect. For example, null mutations in *MKS1* resulting in nonsense mediated decay (NMD) cause MKS, whereas missense variants in *MKS1* exhibit BBS (Leitch et al. 2008). Furthermore, LCA commonly results from a deep intronic variant in *CEP290*, which produces a cryptic exon as a result of aberrant splicing. This variant is thought to specifically affect the retina due to a residual unaffected *CEP290* transcript rescuing other cell types (Drivas et al. 2015; Parfitt et al. 2016).

Although primary cilia are ubiquitously expressed, each ciliopathy affects a combination of varying tissues, suggesting tissue specific roles for genes mutated in respective disorders. The mechanism that separates the molecular differences between disorders is not fully understood. There is a general trend that certain ciliopathies are attributed to mutation of proteins localising to the same ciliary compartment (Shaheen et al. 2016). For example, JATD is caused by defects in IFT proteins, whereas JBTS and MKS proteins primarily localise to the transition zone or axoneme.

Furthermore, unlike BBS proteins, NPHP and SLS proteins do not localise to the basal body (table 1.3).

1.7 The Ciliome

The cilium is a compartmentalised organelle that is temporally and spatially distinct from other cellular interactions. As a result, the proteins involved in cilia assembly, function, maintenance, and disassembly are thought to work in an intricate and relatively closed network. The interacting compendium of proteins has been named the ciliome (Inglis et al. 2006). A curated database of cilia and basal body associated proteins was first assembled in 2006, derived from comparative genomic, proteomic, microarray and promoter studies from a variety of model organisms (Gherman et al. 2006). The first edition of the ciliome consisted of over 1000 proteins, however estimates have been considered at 1200 to 2500 ciliary-related genes (van Dam et al. 2017; Inglis et al. 2006; Lai et al. 2011).

A number of silencing experiments using RNA interference (RNAi) in ciliated cells has identified novel ciliary genes, complexes, functions and interactions (J. Kim et al. 2010; Lai et al. 2011; Wheway et al. 2015). A study by Kim and colleagues (2010) validated 49 putative genes that modulate ciliogenesis, as well as identification of regulatory proteins linking new pathways, endocytic recycling and actin dynamics, with ciliogenesis. There were a number of proteins that overlapped with cell cycle entities, which verified the suspected relationship cilia have with cell cycle, migration and polarisation (Christensen et al. 2008). An interesting finding from Lai et al. (2011) was that knockdown of different constituents from the same complex did not necessarily exhibit the same phenotype. For example, silencing of *Ift46* resulted in shortened cilia, whereas knockdown of *Ift88* prevented ciliation, despite both proteins functioning in the IFT-B complex. The difference could be explained by additional or specific roles a protein has within or outside of their respective complex.

Large-scale studies have implemented whole genome screens or systems biology approaches using proteomics techniques. Wheway et al. (2015) instigated a whole genome screen using small interfering RNA (siRNA) to detect key cilia genes in murine inner medullary collecting duct (IMCD3) cells and human telomerase reverse transcriptase retinal pigmented epithelial (TERT-RPE1) cells. They identified 112 ciliogenesis genes including previously unrelated components of the ubiproteome, splicing factors and GPCRs. This study determined a novel interaction between members of the proteasome and cilia, where knock down of these subunits resulted in increased

cilia numbers. This highlighted the prominence of robust proteostasis in mediating ciliary function (Wheway et al. 2015). The close network of ciliary proteins was further displayed by Boldt and colleagues (2016) using a powerful technique of tandem affinity purification experiments coupled to mass spectrometry. Using 217 baits for well characterised ciliary proteins, they were able to model a unique interacting landscape consisting of 1319 proteins in 52 complexes and 4905 interactions. Their work supported previous links to vesicle transport, signalling, ubiquitination and cytoskeletal regulation (Mick et al. 2015; Hernandez-Hernandez et al. 2013; Christensen et al. 2008; J. Kim et al. 2010). 544 non-ciliary proteins were identified, supporting the belief that gold standard ciliary proteins have interconnecting roles outside of the cilium (Boldt et al. 2016; Bizet et al. 2015).

Within the Cilia consortium, the aim has not only been to recognise novel identifications but also create a gold standard set of cilia interacting genes. The SYSCILIA gold standard v1 (SCGSv1) consists of 303 genes, 66 of which are specific to motile cilia, and are comprised from 27 experimental ciliary studies (van Dam et al. 2013). The SCGSv1 dataset was stringently curated from genes with a definite ciliary role that have been experimentally identified and validated. The latest published ciliome is the CiliaCarta database, consisting of benchmarked and statistically integrated data from genome-wide studies, literature and annotations (van Dam et al. 2017). This more extensive dataset contains a total of 836 genes. Of note, these large-scale screening and curation studies are consistently identifying genes with unknown ciliary function, suggesting that there are still biological processes yet to be uncovered.

The studies described above have augmented our knowledge of how the components of the ciliome interact, revealing interconnecting complexes and new relationships with different cellular processes. They have also created a valuable resource for dissecting the protein network in the context of ciliopathies, and allowing prediction of novel candidates for disease, as well as generating possible targets for therapies. This highly coordinated and interconnected network of proteins helps explain why ciliopathies are phenotypically similar with various overlapping clinical features. It also provides an explanation for how mutations in interacting proteins can manifest the same disease or different diseases arise from variants in the same gene. An example of the latter is *CEP290*, a gene mutated in 6 distinct ciliopathies (BBS, JBTS, MKS, NPHP, LCA, and SLS; table 1.3) (Coppieters et al. 2010). A theory for its multifaceted involvement is that CEP290 directly interacts with different complexes linked with each ciliopathy (Zaghloul and Katsanis 2009). For example, CEP290 interacts with RPGRIP1 and RPGR resulting in LCA (Rachel, Li, et al.

2012); proteins of the transition zone including CC2D2A and TMEM67, which cause MKS and JBTS (Gorden et al. 2008; Garcia-Gonzalo et al. 2011); components of the BBSome and BBS6 causing BBS (Barbelanne et al. 2015; Rachel, May-Simera, et al. 2012); and IQCB1 resulting in NPHP and SLS (Barbelanne et al. 2013) (**figure 1.3**).

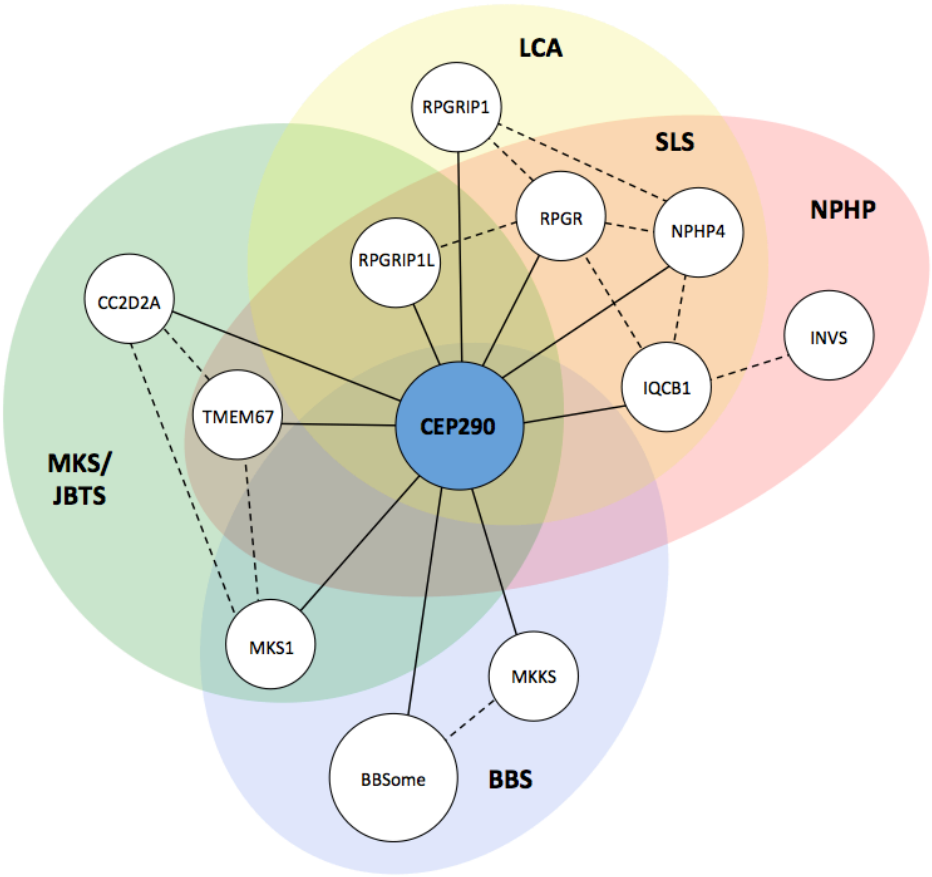

Figure 1.3 – The multifaceted interactions of CEP290, which may explain its role in different ciliopathies. Solid lines refer to direct interactions between CEP290 and associating proteins. Dotted lines refer to interactions not directly involving CEP290. Proteins are also grouped by disorders they are implicated in, where BBS = Bardet-Biedl Syndrome, LCA = Leber congenital amaurosis, SLS = Senior-Løken Syndrome, NPHP = Nephronophthisis, MKS = Meckel-Gruber Syndrome, JBTS = Joubert Syndrome. Image adapted from Rachel et al. 2012.

1.8 Bardet-Biedl Syndrome

Bardet-Biedl Syndrome (BBS) is a pleiotropic genetic disorder that exhibits both genetic and phenotypic heterogeneity. It is a rare disease, estimated to affect 1 in 160,000 individuals in Northern Europe (Waters and Beales 2011). The prevalence is greater in isolated communities such as Kuwait (1 in 13,500) (Farag and Teebi 2008) or Newfoundland (1 in 17,000) (Green et al.

1989), where there is an increased level of consanguinity or a smaller gene pool. BBS was first described in the early 20th century. Georges Bardet reported the clinical features of retinal deterioration, polydactyly, obesity and hypogonadism in 2 siblings (Bardet 1920). Independently, Arthur Biedl published findings for siblings displaying the same clinical features as Bardet's, with additional mental retardation (Biedl 1922). These 5 features became the diagnostic criteria for Bardet-Biedl syndrome. It wasn't until some 50 years later that renal dysfunction was reported (Bauman and Hogan 1973), and was later suggested to be an additional cardinal feature of BBS (Harnett et al. 1988; Green et al. 1989; Churchill et al. 1981).

Throughout early clinical reports on BBS, a large amount of phenotypic heterogeneity was described (Green et al. 1989; Klein and Ammann 1969), often making it challenging to correctly diagnose patients. For example, in a population-based survey by Klein and Ammann (1969), 45.6% of patients were reported to display a 'complete' BBS phenotype, whereas others manifested an 'incomplete' (4/5 cardinal features, exclusive of renal dysfunction), 'abortive' (2/5 cardinal features) or undetermined condition. Therefore, a set of diagnostic criteria was set out where 4/5 traits were required for a BBS diagnosis (Schachat and Maumenee 1982). Moreover, there were often difficulties in diagnosing patients at a young age due to the progressive nature of some symptoms. The age of onset for retinal dysfunction and renal abnormalities can vary considerably. Polydactyly is the only dysmorphic cardinal feature seen at birth, which is not uncommon or distinct to BBS. To facilitate an earlier diagnosis age, a number of secondary features have been incorporated into the diagnostic criteria. These features include developmental delay, speech disorders, behavioural traits, diabetes mellitus, strabismus/ cataracts, orodental abnormalities, cardiovascular anomalies, syndactyly/brachydactyly, and fibrocystic liver disease (Waters and Beales 2011; Beales et al. 1999). Therefore, Beales et al (1999) suggested updated criteria, where at least 4 cardinal features, or a combination of 3 cardinal plus 2 secondary features are required for BBS.

1.8.1 Identification of BBS loci

Initially, the genetic model for BBS was not clear. Bergsma and Brown (1975) discussed whether BBS could be inherited through a polygenic model, as it was not understood how a multisystem disorder could arise from a single gene. On the other hand, BBS features were clearly distinct and relatively consistent, so it was suspected that the BBS gene would map to a single locus (Katsanis, Lupski, et al. 2001). The model was complicated further when, upon linkage analysis of affected families, many loci were identified. *BBS2* was the first BBS locus uncovered, which was mapped

to chromosome 16 in a large consanguineous Bedouin family (Kwitek-Black et al. 1993). This was shortly followed by *BBS1* to chromosome 11 (Leppert et al. 1994), *BBS3* to chromosome 3 (Sheffield et al. 1994) and *BBS4* to chromosome 15 (Carmi et al. 1995). Further genetic analysis of distantly related Newfoundland families determined that there was evidence for more BBS loci (Young, Woods, et al. 1999), revealing the identification of *BBS5* (Young, Penney, et al. 1999) and *BBS6* (Katsanis et al. 2000).

From 2003 onwards there was a steady identification of novel BBS loci by linkage analysis (*BBS7*) (Badano et al. 2003), homozygosity mapping (*BBS8-BBS16*) (Ansley et al. 2003; Nishimura et al. 2005; Stoetzel et al. 2006; Leitch et al. 2008; S. K. Kim et al. 2010; Metzler 2011) and more recently whole exome and whole genome sequencing (*BBS17-BBS20,* and *BBS21*, respectively) (Marion, Stutzmann, et al. 2012; Scheidecker et al. 2014; Aldahmesh et al. 2014; Schaefer et al. 2016; Heon et al. 2016). Due to the advent of next generation sequencing (NGS) technologies, the diagnosis rate has increased from 80% to ~90% determined cases in the last half a decade (Prof P. Beales, personal communication).

The distribution of BBS gene variants can vary amongst different ethnic populations. The majority of pathogenic variants are located in *BBS1* and *BBS10*, which are found at an incidence of 23% and 20%, respectively (Mykytyn et al. 2002; Forsythe and Beales 2013) (table 1.4). Within European populations, *BBS1* p.M390R is the most common variant, where an arginine residue replaces a methionine at the 390th amino acid in BBS1. This variant accounts for ~80% of all *BBS1* findings and is thought to have risen from an ancient haplotype (Mykytyn et al. 2003; Beales et al. 2003). The incidence of other genes is rare, where some genes, such as *BBS15/WDPCP*, have only been seen in a single family (S. K. Kim et al. 2010). Variants in *BBS4*, *BBS5* and *BBS8* are more commonly seen in Middle Eastern families (Forsythe and Beales 2013).

BBS gene name	Gene Alias	Loci	Frequency	Known Function	Localisation	References
BBS1		11q13.2	23%	Part of the BBSome complex	BB	(Leppert et al. 1994; Mykytyn et al. 2002; Jin et al. 2010)
BBS2		16q21	8%	Part of the BBSome complex	BB	(Kwitek-Black et al. 1993; Jin et al. 2010)
BBS3	*ARL6*	3q11.2	2%	GTPase that functions with BBSome to coat proteins for IFT	BB, TZ	(Sheffield et al. 1994; Fan et al. 2004; Wiens et al. 2010; Jin et al. 2010)
BBS4		15q24.1	2%	Part of the BBSome complex	BB	(Carmi et al. 1995; Jin et al. 2010)
BBS5		2q31.1	<1%	Part of the BBSome complex	BB	(Young, Penney, et al. 1999; Jin et al. 2010)
BBS6	*MKKS*	20p12.2	6%	Forms chaperonin complex for BBSome assembly	BB	(Katsanis et al. 2000; Seo et al. 2010)
BBS7		4q27	2%	Part of the BBSome complex	BB	(Badano et al. 2003; Jin et al. 2010)
BBS8	*TTC8*	14q31.3	1%	Part of the BBSome complex	BB	(Ansley et al. 2003; Jin et al. 2010)
BBS9		7p14.3	6%	Part of the BBSome complex	BB	(Nishimura et al. 2005; Jin et al. 2010)
BBS10		12q21.2	20%	Forms chaperonin complex for BBSome assembly	BB	(Stoetzel et al. 2006; Seo et al. 2010)
BBS11	*TRIM32*	9q33.1	<1%	E3 ubiquitin ligase	IF	(Chiang et al. 2006)
BBS12		4q27	5%	Forms chaperonin complex for BBSome assembly	BB	(Stoetzel et al. 2007; Seo et al. 2010)
BBS13	*MKS1*	17q22	5%	Regulates ciliogenesis and centriole migration	BB	(Tammachote et al. 2009; Leitch et al. 2008)
BBS14	*CEP290*	12q21.32	1%	Required for recruitment/targeting of key centrosomal proteins. Involved in regulation of BBSome	BB, centrosome	(Leitch et al. 2008; Barbelanne et al. 2015)
BBS15	*WDPCP*	2p15	<1%	Necessary for ciliogenesis	Axoneme	(S. K. Kim et al. 2010)
BBS16	*SDCCAG8*	1q43	<1%	Unknown	BB, centriole, TZ	(Metzler 2011; Insolera et al. 2014)
BBS17	*LZTFL1*	3p21.31	<1%	Regulates BBSome trafficking and HH signalling	BB, cytosol	(Marion, Stutzmann, et al. 2012; Seo et al. 2011)
BBS18	*BBIP10*	10q25.2	<1%	Part of the BBSome complex. Necessary for microtubule maintenance	BB	(Scheidecker et al. 2014; Jin et al. 2010)
BBS19	*IFT27*	22q12.3	<1%	Anterograde IFT	IFT-B	(Aldahmesh et al. 2014; Lucker et al. 2005)
BBS20	*IFT172*	2p23.3	<1%	Anterograde IFT	IFT-B	(Schaefer et al. 2016; Lucker et al. 2005)
BBS21	*C8ORF37*	8q22.1	<1%	Unknown	BB	(Heon et al. 2016; Khan et al. 2016)

Table 1.4 – Summary of causative BBS genes, including frequency of gene in the BBS population and cellular localisation. BB=basal body, TZ=transition zone, IF= intermediate filaments.

1.8.2 BBS causality and localisation

Despite successful cloning of *BBS1-BBS5* at the turn of the century, the aetiology of BBS still remained elusive (Mykytyn et al. 2002; Mykytyn et al. 2001; Nishimura et al. 2001). The breakthrough occurred in 2003 with the cloning and expression experiments of *BBS8* in ciliated cells and *Caenorhabditis elegans* (Ansley et al. 2003). *TTC8* was identified as *BBS8* after screening for structural protein homology with BBS4, both of which contained conserved tetratricopeptide repeats (TPRs). TPRs are structural motifs that are common in complexing proteins, and are known to be associated with flagella and cilia (Xu et al. 2015). Ansley and colleagues (2003) noticed the structural similarity with the prokaryotic protein PilF, required for multimerisation and localisation of the bacterial pilus (Koo et al. 2008). They therefore speculated that BBS8 might localise to mammalian cilia. The investigation determined that BBS8 was a centrosomal protein, mimicking localisation of y-tubulin at MTOCs, and was tightly associated with the basal body in ciliated IMCD3 cells. These findings were mimicked with the *C. elegans* othologues of *BBS1, BBS2, BBS7* and *BBS8*, demonstrating that BBS was likely caused by a defective basal body.

1.8.3 Complexes, roles and interactions

The majority of BBS proteins are known to function in complexes and networks, which can help explain why mutations disrupting single proteins in the complex can result in the same clinical phenotype. Through investigation of structural homology, it was observed that certain BBS proteins shared tertiary similarity and interaction motifs. BBS1/BBS2/BBS7 are formed from β-propellers and BBS4/BBS8 both have conserved TPR domains, raising the possibility that these proteins share functional likeness and interact within complexes. Unexpectedly, it was found that at least 6 proteins associated with BBS4 through tandem affinity purification experiments, namely BBS1, BBS2, BBS5, BBS7, BBS8 and BBS9 (Nachury et al. 2007). Subsequently, an additional protein, BBIP10 (BBS18) was found to associate, forming the fully assembled BBSome complex (Loktev et al. 2008). It has been established that this core complex is required for recruitment of cargo to the ciliary membrane, through interaction with RABIN8 and RAB8, and may have a stabilising role in IFT (Wei et al. 2012; Jin et al. 2010) (**figure 1.4**). Although BBSome proteins assemble into a unit, they are known to have specific roles either independently or within the complex. BBS1 interacts directly with ARL6, which recruits the BBSome to the ciliary membrane (Jin et al. 2010; Mourão et al. 2014). BBS4 is required for anchoring of the BBSome to the basal body, due to direct interaction with the centriolar satellite protein, PCM1 (Kim, 2004). BBS5 is thought to have a specific role in photoreceptors due to light-dependent phosphorylation and

interaction with phototransduction protein, Arrestin1 (Smith, 2013). External to the BBSome, BBIP10 (BBS18) was shown to be required for polarisation, stabilisation and acetylation of microtubules (Loktev et al. 2008).

It is thought that the BBSome is not required for ciliogenesis, as knockdown of most members do not affect ciliary assembly, with the exception of BBS1 and BBIP10 (BBS18) (Loktev et al. 2008; Jin et al. 2010). Furthermore, murine models are viable and express cilia when *Arl6, Bbs2, BBS4*, and *Bbs7* are separately knocked out, which is not the case for *Bbs1* (Mykytyn et al. 2004; Nishimura et al. 2004; Zhang et al. 2011; Zhang et al. 2013). This suggests that BBS1 has an independent essential role in cilium assembly. Moreover, the importance of *BBS1* is reflected in human genetics, where homozygous null mutations have been identified in all subunits of the BBSome, apart from *BBS1* (Jin et al. 2010). Disease causing mutations in BBS1 most commonly present with the p.M390R missense variant on at least one allele (Mykytyn et al. 2002).

A second complex was discovered featuring three BBS proteins, MKKS (BBS6), BBS10 and BBS12, with homology to the TCP-1 Ring Complex (TRiC) chaperonin family (also called chaperone containing TCP-1 (CCT)). As mutations in these chaperonin-like genes result in the same phenotype as the BBSome genes, a close interaction was expected between the two complexes. Seo and colleagues (2010) demonstrated that the 3 chaperonin-like proteins were required for BBSome assembly and formed a complex with 6 other TRiC/CCT chaperones that localised to the centrosome. BBS6, BBS10 or BBS12 did not exhibit typical ATPase activity required for protein folding by chaperone proteins, but were necessary to stabilise BBS2 and BBS7 and coordinate the formation of the BBSome (**figure 1.4**) (Seo et al. 2010). Thus, their role may be to mediate the association of BBSome proteins with the 6 TRiC/CCT chaperones.

By contrast, the ciliary function of the outstanding BBS proteins remains elusive. *BBS11* was identified as the RING finger E3 ubiquitin ligase, *TRIM32* (Chiang et al. 2006). BBS11 has been shown to be associated with SHH repressor, GLIS2, (Ramachandran et al. 2014) and other ubiquitin ligases were shown to localise to primary cilia, indicating a potential role in ciliary function (Boldt et al. 2016). However, little else is known about its BBS-specific role. Mutation of *LZTFL1/BBS17* was described in a consanguineous family exhibiting atypical BBS, featuring obesity, hypogonadism, cognitive impairment, situs inversus and insertional polydactyly (Marion, Stutzmann, et al. 2012). BBS17 interacts with BBS2, BBS7 and BBS9 of the BBSome and has also been implicated in trafficking of Smoothened (SMO), a SHH signalling transducer, to the ciliary

membrane (Seo et al. 2011). *BBS19* and *BBS20* are the only BBS genes that localise within an IFT complex. It has been determined that BBS19 functions outside of the IFT-B complex assembly to assist with BBSome assembly and exit from the cilia (Liew et al. 2014; Eguether et al. 2014).

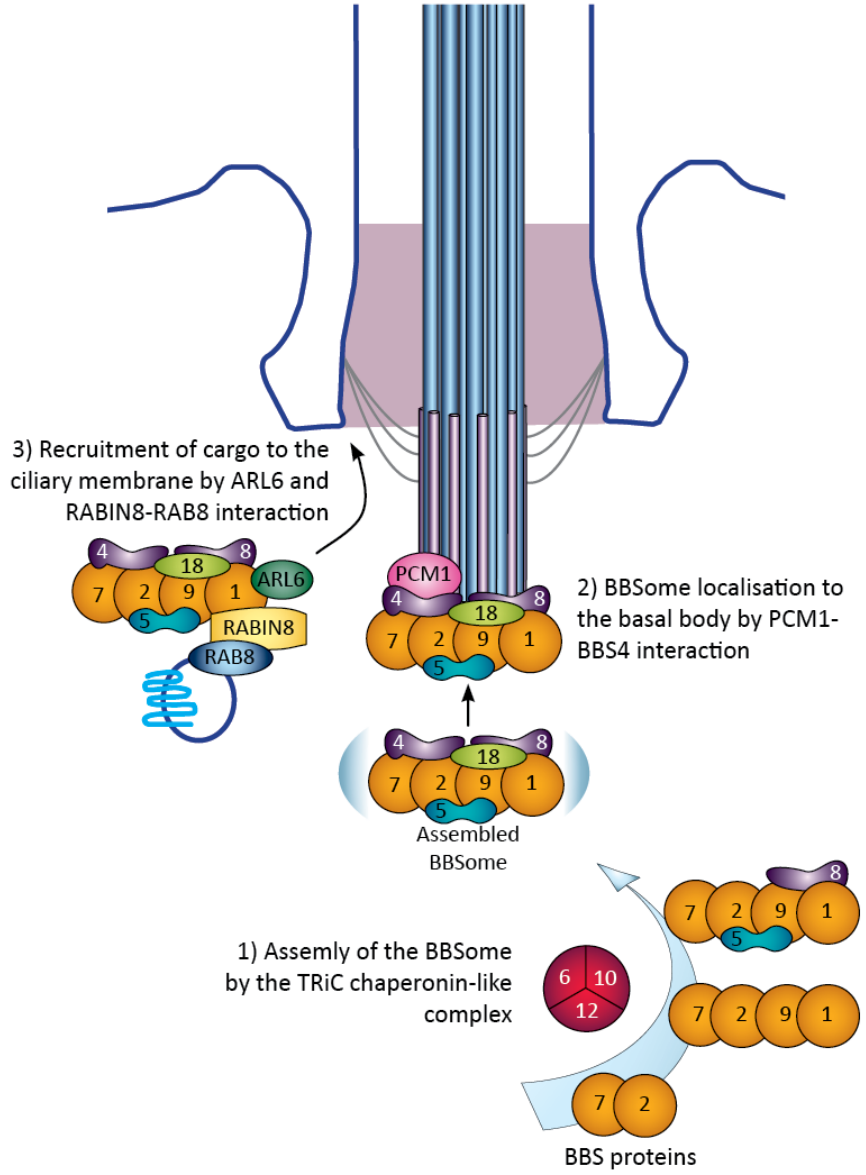

Figure 1.4 – Molecular assembly and interactions of the BBSome. 1) The components of the BBSome (BBS1, BBS2, BBS4, BBS5, BBS7, BBS8, BBS9, and BBS18) are assembled sequentially by the TRiC chaperonin-like complex (BBS6, BBS10, and BBS12), thought to be initiated by interaction with BBS7 and BBS2 first (Seo et al. 2010). 2) Once assembled, the BBSome is targeted to the basal body due to an interaction between BBS4 and PCM1. 3) The BBSome functions through interactions with ARL6/BBS3, RABIN8 and RAB8 to recruit protein cargo to the ciliary membrane for intraflagellar transport.

1.8.4 Molecular basis of BBS

Through manipulation of BBS genes in experiments utilising ciliated cells and model organisms, our knowledge of BBS has augmented in recent years. Although the precise aetiology differentiating all 21 genes is not entirely understood, here I will outline the molecular basis of the cardinal BBS features.

1.8.4.1 Retinal dystrophy

Retinal dystrophy is the most prevalent BBS feature, predicted to be present in 93% of cases (Forsythe and Beales 2013). Patients are thought to have functional eyesight at birth, but progressive visual decline begins early, typically leading to registered blindness before adulthood (Beales et al. 1999). Retinal dystrophy in BBS is usually diagnosed as rod-cone dystrophy, describing the deterioration of rod photoreceptor first, followed by decline of cone photoreceptors. Cone-rod dystrophy is less commonly described (Mockel et al. 2011). Rod photoreceptors are responsible for light sensitivity; thus, the first retinal symptom of rod-cone dystrophy tends to be nyctalopia (night blindness). Subsequent loss of visual acuity and colour vision follows with global retinal degeneration. In a survey of 109 BBS patients by Beales et al. (1999), the mean age of onset for night blindness and subsequent registration of blindness was 8.5 years and 15.5 years, respectively. Secondary ocular features can include cataracts, strabismus, nystagmus, myopia, astigmatism and macular degeneration (Green et al. 1989). The pathobiology of retinal dystrophy in BBS is reviewed in detail in chapter 6.

1.8.4.2 Obesity

Obesity is a result of energy imbalance caused by a greater calorific intake than metabolic expenditure. Childhood truncal obesity has been reported to affect 72-92% of BBS patients (Forsythe and Beales 2013), though other surveys suggest 72% are overweight and 52% are obese (Beales et al. 1999). Obesity is diagnosed by the World Health Organisation (WHO) through assessment of body mass index (BMI), calculated by dividing body weight by height squared (kg/m^2). The BMI healthy weight range is 18.5-24.9 kg/m^2, whereas 25-29.9 kg/m^2 is classed as overweight and >30 kg/m^2 is obese. Although there is a high prevalence of obesity in BBS patients, some argue that it is a non-specific phenotype in today's Western society [1], even more so in subjects likely to be inactive due to visual decline (Pietiläinen et al. 2008). Regardless, obesity is

[1] A 2008 WHO survey revealed that 26% of American men and 29% of American women were reported as obese (World Health Organisation 2008)

associated with morbidity and mortality in patients and can lead to secondary features of cardiovascular disease, cancer and type II diabetes.

BBS-related obesity is molecularly distinct from nonsyndromic obesity. Studies have determined that obesity results from hyperphagia triggered by leptin resistance and defective satiety control in hypothalamic neurons. *BBS4* null mice exhibited hyperphagia, consuming between 20-30% more than their wild type littermates, and displayed elevated circulating leptin (Eichers et al. 2006). Leptin sensitivity could not be restored by administration of leptin or calorie restriction suggesting that resistance was not a secondary symptom of obesity, but intrinsic to BBS (Seo et al 2009). Leptin resistance was confirmed in a study of 50 BBS patients, who had significantly greater intraabdominal fat mass, insulin resistance and blood pressure compared to 100 BMI, age, and sex matched controls (Feuillan et al. 2011). This finding contributed to the hypothesis that BBS-dependent obesity is driven by leptin resistance caused by aberrant presentation of GPCRs responsible for control of satiety, such as MCHR1 and the leptin receptor, LEP-R, on the cilia of hypothalamic neurons (Loktev and Jackson 2013; Seo et al. 2009; Berbari et al. 2008). In support of this, ablation of BBS proteins in neurons was shown to significantly reduce expression of *POMC* transcripts, a gene activated downstream of leptin binding, which signals through STAT3 (Rahmouni et al. 2008). Recently, the BBSome was shown to interact via BBS1 with LEP-R, which was necessary for trafficking of the leptin receptor to the plasma membrane of hypothalamic cells (Guo et al. 2016; Seo et al. 2009). This provided a possible mechanism that leptin resistance occurs due to defective trafficking of the leptin receptor in neural cilia.

Other studies have suggested that adiposity is not exclusively driven by leptin signalling defects, but BBS patients may have an increased propensity for adipogenesis due to ciliary defects (Marion et al. 2009; Marion, Mockel, et al. 2012). Marion and colleagues (2009) determined that preadipocytes were transiently ciliated during adipogenesis, and ablation of *BBS10* or *BBS12* resulted in upregulation of proadipogenic pathways. It was suggested that the increased propensity for adipogenesis was driven by disrupted Wnt and HH signalling, as these pathways are potent adipogenic inhibitors (Ross et al. 2000; Kopinke et al. 2017).

1.8.4.3 Polydactyly

Polydactyly is the growth of extra digits and can affect a single, or up to all 4, limbs. BBS patients have an incidence of 63-81% polydactyly, where 4-limbs were reported in 21% of cases, 9% hand polydactyly and 21% foot polydactyly (Beales et al 1999). It most commonly has a postaxial

presentation, where extra digits form on the ulnar side of the hand or fibula side of the foot. They may manifest as fully developed digits or subtly as a nubbin. Polydactyly is not distinct to ciliopathies and can present in 1 in 1400 live Caucasian births, with up to 10 times greater prevalence in African populations (Aucourt et al. 2012). As a result, extra digits are routinely surgically removed after birth, which highlights the importance of maintaining a detailed clinical history to correctly diagnose patients. Other common limb abnormalities include syndactyly (fused digits), brachydactyly (shortening of digits) and tapered fingers (Green et al. 1989).

Abnormal digit development results from defective SHH signalling during limb bud formation. *Shh* null mutations in mice lack any digits, whereas loss of function of *Gli3* induces polydactyly (Robert and Lallemand 2006). *Gli3* contains both activator and repressor domains, which are activated through proteolytic cleavage. Disruption of IFT in mice led to polydactyly of all four limbs, which suggested that IFT was involved in the processing of Gli3 from an activator to repressor form (Liu et al. 2005; Huangfu and Anderson 2005). Interestingly, BBS mutant mice do not display polydactyly, making the exact role of BBS proteins in limb development elusive.

1.8.4.4 Genital abnormalities

Genital anomalies have been reported to affect between 59-98% of BBS patients (Forsythe and Beales, 2013). The primary malformation affecting males is hypogonadism, including micropenis, maldescent of testes and hypospadias. Abnormalities in females are generally less common, but can include hydrometrocolpos, vaginal atresia, vaginal agenesis and polycystic ovaries (Deveault et al. 2011). Reports of delayed puberty and irregular menstruation cycles are also common (Beales et al. 1999).

The mechanism causing genital abnormalities in BBS remains uncertain. The developmental defects may be a product of aberrant SHH signalling, as *Shh* is required for development of a genital tubercle and androgen-independent morphogenesis in mice (Lin et al. 2009). Anomalies may be attributed to improper hormone signalling. The GCPR KISSR1 is known to localise to primary cilia of gonadotropin-releasing hormone (GnRH) neurons in the hypothalamus (Hilgendorf et al. 2016). Signalling by Kisspeptin, the ligand for KISSR1, is involved in puberty onset and adult reproductive function, and mutations in *KISSR1* can result in hypogonadotropic hypogonadism (de Roux et al. 2003). Ablation of cilia on GnRH neurons was shown to reduce kisspeptin-mediated neuronal activity (Koemeter-Cox et al. 2014).

1.8.4.5 Renal dysfunction

Renal anomalies are estimated to affect 53% of patients (Forsythe and Beales 2013). Renal abnormalities of BBS are defined as structural malformations, including renal cysts, fetal lobulation, fibrosis, calyceal clubbing/blunting, dysplastic kidneys, renal calculi, or vesicoureteric reflux (Harnett et al. 1988). Other symptoms related to the urorenal system, such as recurrent urinary tract infections, polyuria and polydipsia are classed as secondary BBS features. The age of onset of renal abnormalities can span from late childhood onwards. It is associated with mortality and morbidity, but transplantation for treatment of chronic kidney failure can successfully prolong life (Sharifian et al. 2007).

The renal malformations observed in ciliopathies are thought to result from dysfunctional mechanosensation in the renal tubule, caused by aberrant trafficking of PC1 and PC2. As seen in ADPKD, failure to produce a functional PC1/PC2 channel complex resulted in abnormal cell morphology and polarity, which led to cyst formation (Boucher and Sandford 2004; Happé et al. 2011; Nauli et al. 2003). Mechanosensation is required for Ca^{2+} influx into the cell, which activates Ca^{2+}-dependent Wnt signalling. It has been proposed that interruption of ciliary function switches renal cells towards the canonical Wnt signalling pathway, elevating β-catenin levels (Simons et al. 2005). Cyst formation occurs as a result of β-catenin induced overproliferation (Lin et al. 2003; Saadi-Kheddouci et al. 2001). A direct interaction between the BBSome (via BBS1) and PC1 has been established, which is required for trafficking of PC1 to the ciliary membrane (Su et al. 2014).

1.8.4.6 Learning difficulties

Learning difficulties are estimated to affect 61% of patients. Other cognitive features include developmental delay, behavioural issues, mental disability and speech deficit (Beales et al. 1999). ~50% of children attend special schools and can exhibit changeable behaviour with outbursts of frustration (Forsythe and Beales, 2013). There are reports that ~50% of cases have neocortical and hippocampal volume loss, associated with impaired cognition, sensory perception and memory (Baker et al. 2011; Guemez-Gamboa et al. 2014).

The links between the cilium and neurodevelopment are thought to be related to SHH or other key signalling pathways. Ablation of primary cilia in murine neurons resulted in decreased hippocampal proliferation and delayed spatial learning, linked to loss of Shh signalling (Amador-Arjona et al. 2011; Tong et al. 2014). Furthermore, there may be a link between somatostatin signalling and learning difficulties. Reduced somatostatin signalling is associated with neurological

and cognitive dysfunction and the receptor SSTR3 is known to localise to primary cilia of hippocampal neurons (Händel et al. 1999). Berbari et al. (2008) demonstrated that Sstr3 localisation was dependent on the BBSome, as *Bbs2*$^{-/-}$ and *Bbs4*$^{-/-}$ mice failed to express Sstr3 on hippocampal neurons.

1.8.5 Genotype-Phenotype Correlation in BBS

As over 20 genes have been described to manifest BBS, it is perhaps not surprising that phenotypic heterogeneity is a common hallmark described across cohorts (Deveault et al. 2011). There has been some attempt at categorising genes to establish a genotype-phenotype correlation for BBS, particularly with ocular features (Héon et al. 2005; Daniels et al. 2012; Azari et al. 2006). For example, Heon and colleagues (2005) defined ocular phenotypes of 34 patients with retinal dystrophy carrying mutations in *BBS2*, *BBS3* and *BBS4*. They determined that myopia was present in *BBS3* and *BBS4* patients, but not with mutations in *BBS2*. Correlations have also been demonstrated for renal disease, where 5 families with mutations in *SDCCAG8* (*BBS16*) exhibited high penetrance early onset renal failure and simultaneous absence of polydactyly (E Schaefer et al. 2011). Furthermore, Imhoff et al. (2011) conducted an investigation with 33 BBS patients of a French cohort to uncover genotype-phenotype correlations. They proposed a link between severity of renal phenotype with different BBS complexes affected. 70% of patients with mutations affecting the chaperone-like BBS genes, *BBS6, BBS10,* and *BBS12* had renal abnormalities, compared to 15% of patients with variants affecting BBSome genes (Imhoff et al. 2011). These research studies were typically limited by small cohort sizes and utilisation of patients from the same ethnic background. Larger detailed studies, such as the evaluation of 105 cases by Deveault et al (2010), did not identify any genotype-phenotype correlation, except that patients with variants in *BBS1* tended to have later onset and milder severity.

Limited evidence of genotype-phenotype correlation in BBS is even apparent between patients with a genetic diagnosis in the same gene, or with the same disease-causing variant (Estrada-Cuzcano et al. 2012; Beales et al. 2003). This phenomenon extends to BBS cases within the same family, where there have been reports of siblings with clinical features of other ciliopathies alongside patients with BBS (Zaki et al. 2011). It is widely considered that this phenotypic expressivity is accounted for by secondary genetic modifiers. These may present at a second site within the disease-causing loci, or in another gene that functions in the same complexes/networks as the disease-causing loci (Zaghloul and Katsanis 2009; Badano et al. 2006;

Beales et al. 2003; Katsanis, Lupski, et al. 2001; Katsanis, Ansley, et al. 2001). This topic is discussed in detail as the subject of chapter 3.

1.8.6 Aetiology of *BBS1* p.M390R

The variant p.M390R in *BBS1* is the most common variant in patients of European descent, and arguably the most complex. Many studies have characterised it as the loci that manifests a less severe form of BBS (Deveault et al. 2011; Daniels et al. 2012; Estrada-Cuzcano et al. 2012), where patients have been reported to have reduced factors of obesity and cardiovascular disease than patients with mutations in *BBS10* (Feuillan et al. 2011; Forsythe et al. 2015). This may be attributed to the characteristics of the variant more than the loci. The most common variant found in *BBS10* is a frameshift mutation resulting in a truncated protein (p.C91Lfs*5), whereas p.M390R is a single base change that retains the full-length protein. Evaluations of retinal dystrophy have also highlighted the vast spectrum of severity that is typical in patients with *BBS1* p.M390R mutations. Patients carrying this variant have been reported to manifest all cardinal features of BBS; retinal dystrophy paired with one other feature, such as obesity or polydactyly; and nonsyndromic rod-cone dystrophy (Cannon et al. 2008; Azari et al. 2006; Estrada-Cuzcano et al. 2012). Other patients only presented with retinal degeneration into their fifth decade of life, or were genotyped as unaffected relatives or control subjects who were later found to be positive for the variant (Azari et al. 2006; Estrada-Cuzcano et al. 2012; Beales et al. 2003).

Genetic variants that drastically alter protein structure, such as nonsense, frameshift or mutations affecting the splice junctions, induce a loss of function response, as resulting proteins will either be dysfunctional or degraded by NMD. In the case of missense variants, such as p.M390R, the full-length protein is translated and the aetiology is not so clear. *BBS1* p.M390R is known to be a hypomorphic variant (Zaghloul et al. 2010), where protein exhibits reduced performance as function is partially lost. The disease mechanism of the variant has seldom been studied. Mourão and colleagues (2010) published the first study investigating the protein structure of BBS1 and its interaction with ARL6, essential for recruitment of the BBSome to the ciliary membrane (Jin et al. 2010), in *Chlamydomonas reinhardtii.* The authors purified the N-terminal of BBS1, which bound with high affinity near to the GTP binding pocket of ARL6-GTP. They ascertained that the binding interface between BBS1 and ARL6 is highly conserved across different species, where p.M390R lies in a conserved β-propeller domain close to the binding site. Structural assays determined that the p.M390R protein was suitably folded, but the β-propeller was partially unstructured, reducing the affinity of BBS1 for ARL6-GTP (Mourão et al. 2014). This

would suggest that BBS is caused by insufficient recruitment of the BBSome to the ciliary membrane by ARL6 in patients with *BBS1* p.M390R variants. In addition to this conclusion, it is important to note that BBS1 has direct interactions with many other ciliary proteins and cargo (Seo et al. 2009; Loktev and Jackson 2013; Jin et al. 2010; Zhang et al. 2012). The direct effect of p.M390R on these interactions has not yet been elucidated.

1.9 In-depth Phenotyping of Rare Diseases

1.9.1 High definition, in-depth phenotyping at Great Ormond Street Hospital

In 2013, the Centre for Translational Omics at UCL Great Ormond Street Institute of Child Health (UCL GOS ICH) launched an innovative project; High definition, In-depth phenotyping at Great Ormond Street Hospital, coined the HIGH-5 project. Funded by the National Institute of Health Research (NIHR) Great Ormond Street Hospital (GOSH) Biomedical Research Centre (BRC), the project aimed to use next generation sequencing (NGS) technologies to elucidate the pathobiology of rare diseases, identify genetic modifiers or clinical biomarkers, and personalise patient care. Patients were recruited into cohorts from 7 different genetic or multifactorial diseases. 15 patients with the common *BBS1* c.1169T>G/ p.M390R mutation were enrolled into a BBS cohort, which attained additional funding support from the NIHR Rare Diseases Translational Research Collaboration (TRC).

The present study operated as a pilot project to ascertain suitable methods for sample collection, processing and storage. Furthermore, it provided an opportunity to trial novel state-of-the-art technologies and determine analytical methods for investigation of large datasets. The HIGH-5 project was possible due to the collaboration of researchers from different backgrounds, including clinicians, bioinformaticians, geneticists, computer scientists, and analytical scientists.

1.9.2 Study of rare diseases

A rare disease, sometimes called an orphan disease, is a disorder that affects a small percentage of a population. The definition of a rare disease is not universal. According to Rare Diseases Europe (EURORDIS), a disease is classified as rare if it affects less than 1 in 2000 individuals, whereas in the United States a rare disease affects fewer than 200,000 Americans at any one time (Azie and Vincent 2012). It is predicted that in Europe, at least 30 million individuals may be suffering from any of 6000 rare diseases (EURORDIS 2017), where rare monogenic disorders affect 1 in 50 (Boycott et al. 2017). Therefore, the burden of rare diseases is high and cannot be overlooked.

Recently, awareness by advocacy groups and research funding for rare diseases has increased (Boycott et al. 2017). This is partly due to overlapping features of rare diseases in common disorders, where mechanistic insight into disrupted pathways may reveal therapeutic targets for more common diseases. Ciliopathies are a prime example, where collectively phenotypes are rare. However, features such as obesity, retinal dystrophy, hearing loss, and renal failure, are individually more common in the general population. In addition, rare disease research augments our scientific knowledge of biological processes and mechanisms, which otherwise may be unfathomable. On the other hand, research of rare diseases poses some challenges. Due to the small population of patients within a given geographical area, research can often be limited by sample size. Furthermore, high levels of genetic and phenotypic variability within rare diseases can complicate patient stratification and study design.

1.10 The age of next generation sequencing and omic analysis

Next generation sequencing (NGS) refers to high throughput nucleic acid sequencing, such as whole exome sequencing (WES), whole genome sequencing (WGS), and RNA sequencing (RNA-seq). In the time since the Human Genome Project (HGP), there has been an exponential increase in novel sequencing approaches and bioinformatic tools to aid discovery and clinical genetics (Ashley 2016). It is thought that ~50% of all rare genetic diseases now have a molecular diagnosis, which is predicted to increase to near completion over the next decade with further development of NGS technologies (Boycott et al. 2013). The advent of NGS is aided by the decreasing cost of sequencing. This has dramatically fallen in the last two decades, as the first genome by the HGP cost $100 million, and now an individual's genome can be sequenced for <$1500 today (Wetterstrand 2016). As more efficient genome sequencing technologies are being developed by biotech companies, this price will likely decrease to less than $1000 in the near future, making it an affordable research and diagnostic tool (Lupski et al. 2011).

Development of high throughput sequencing and profiling technologies have not only been limited to assessment of nucleic acids. It is now possible to characterise an individual's protein or metabolite composition, epigenetic marks through sequencing of DNA's methylation patterns, or microbiota population by sequencing DNA of gut bacteria (Topol 2014). Collectively, these fields of research have been coined omics, encompassing genomics, transcriptomics, proteomics, metabolomics, epigenomics, and microbiomics. Although individually each technology has its

advantage, a multi-omic integration of datasets can provide enhanced biological insight to aid interpretation, and offer enrichment of data for a more comprehensive analysis.

1.10.1 Precision medicine

As sequencing has become more affordable, large-scale population sequencing projects have been completed, including the 1000 genomes project, the NHLBI Exome Sequencing Project (ESP), and the Broad Institute's Exome Aggregate Consortium (ExAC) and Genome Aggregation Database (GnomAD) (Lek et al. 2016; 1000 Genomes Project Consortium et al. 2015). Each project has demonstrated that a vast amount of genetic variation exists between different populations, as well as between individuals. On average, a human genome was found to contain ~3.5 million single nucleotide variants (SNVs) that differed from the reference genome, where the number of rare variants was much higher than anticipated (Conrad et al. 2010; Lupski et al. 2011). It is well understood that this variation may affect an individual's susceptibility to disease or influence their response to pharmacological treatment (Wray et al. 2007; Goldstein et al. 2003). Now with NGS and omic technologies, a new era of medicine is possible, where patients can be treated with respect to their genetic and environmental backgrounds. This has been referred to as 'personalised', 'individualised', or 'precision' medicine (Ashley 2016; Topol 2014).

Precision medicine is defined as "tailoring of medical treatment to the individual characteristics of each patient" (US National Research Council 2011). This involves stratification of patients by genetic diagnosis or molecular phenotype, rather than generalisation of patients within a single disease type. This personalised approach aims to aid disease prediction, molecular diagnosis, and development of targeted therapies with fewer adverse effects. The utility of precision medicine in rare diseases is exemplified by treatment of cystic fibrosis (CF). CF is caused by mutations in an epithelial ion transporter encoded by *CFTR*. Different sites of variation within the gene results in different loss of function outcomes. For example, a deletion of p.F508 results in protein misfolding and subsequent proteasomal degradation, whereas a missense mutation at p.G551D leads to defective channel opening at the epithelial membrane (Ashley 2016; Brodlie et al. 2015). The small molecule Ivacaftor was developed as a CF treatment, which targets CFTR at the membrane, keeping the channel in an active, open state. This treatment would only be beneficial in patients with CFTR at the cell surface, and thus would not be useful in patients with p.508del variants (Ramsey et al. 2011; Brodlie et al. 2015). This demonstrates that individuals with the same phenotype may require different therapeutic interventions, and thus highlights the importance of understanding the subject's genetic background.

1.10.2 Integrative personal omics profile study: Proof of principle

One of the first applications of human multi-omic data was completed by Michael Synder's lab, where Synder himself was the subject of investigation (Chen et al. 2012). The integrative personal omics profile (iPOP) study followed Synder over a period of 2 years, serially sampling blood components at 20 different time points for a combination of WGS, RNA-seq, proteomic and metabolomic mass spectrometry, and antibody profiling. Interestingly, Synder contracted two viral infections over the sampling period, allowing assessment of healthy and diseased states.

The authors used WGS to identify genetic risk factors for disease, where they detected variants associated with increased risk of basal cell carcinoma, coronary artery disease, hypertriglyceridemia, and type II diabetes (Chen et al. 2012). They subsequently utilised this knowledge by recording measurements, such as fasting plasma glucose and glycated haemoglobin (HbA1c), to monitor the possibility of type II diabetes. Despite Synder presenting a lack of physical risk factors (such as high BMI), they detected a spike in glucose levels, indicative of type II diabetes onset, in response to infection by a respiratory syncytial virus. As a result of this information, he was able to alter his lifestyle by increasing exercise and changing his diet to bring his glucose levels back to normal (Chen et al. 2012). Further to the iPOP study, Synder shared his knowledge with family members who were similarly at risk, encouraging them to also change their lifestyle choices (Topol 2014). The study also demonstrated the benefits of a multi-omic strategy, where datasets were integrated to present a comprehensive analysis of both diseased and healthy states. Characterisation of the transcriptome and proteome allowed them to identify biological pathways, some of which were only detectable through a combination of both datasets (Chen et al. 2012).

1.11 Introduction to omics techniques

1.11.1 Whole genome and whole exome sequencing

The human genome consists of ~3 billion bases, of which ~1.5% is known to be protein-coding (Topol 2014). There are different commercially available technologies for DNA sequencing practices, which utilise different sequencing chemistries. Within this thesis, Illumina technology (Illumina Inc, San Diego, CA, USA) was applied, which uses a sequencing by synthesis approach.

The main methodological differences between WGS and WES is that WES isolates DNA that aligns with exonic regions of the genome prior to sequencing. This is completed using specific exon

capture baits. The argument for favouring WES over WGS used to be lower sequencing costs, where it simply wasn't realistic for research groups to justify the cost of WGS. As it is cheaper, WES is typically sequenced to a deeper coverage (90X instead of 30X for WGS), which increases the accuracy of variant calling. However, as discussed previously, the cost of WGS has reduced drastically, thus many more studies are preferentially using WGS over WES (Caulfield et al. 2017). Furthermore, as WES only sequences exons, other areas of DNA that may harbour deleterious variants are overlooked; for example, regulatory elements in intergenic regions, including transcription factor binding sites or enhancers (Gerstein et al. 2012); or intronic regions, which may harbour mutations that induce aberrant splicing (Cummings et al. 2016). On the other hand, the connotation that WGS sequences the entirety of the genome is misleading. Up to 4% of the genome, equating to ~900 genes, is not yet possible to sequence as they lie in telomeres or centromeres (Topol 2014).

1.11.1.1 Types of human genetic variation

Single nucleotide variation (SNV) is the most prevalent class of variation, mostly lying under the term single nucleotide polymorphism (SNP). A polymorphism is defined as a common variant that is present in greater than 1% of the population. A great proportion of SNPs are not thought to contribute to disease or phenotypic variation, and thus have a neutral impact on the genome (Frazer et al. 2009). It is because of their neutral behaviour that they have accumulated to a high frequency in the population, as highly deleterious variants would be removed by natural selection. However, some types of SNV are deleterious. The nature of the SNV depends on where the site of variation occurs (i.e. within a protein-coding gene), as well as the consequence of the base substitution.

Non-synonymous nucleotide changes in protein-coding genes that effect the protein function are more likely to be deleterious. There are different classes of nucleotide variants that are involved in human disease (table 1.5). Nonsense or frameshift mutations are more likely to be deleterious as they lead to a truncated protein sequence. If the variant occurs early within a transcript, the transcript is likely to undergo NMD, which prevents the mRNA being translated. If a truncating mutation occurs on both alleles, the result is a null phenotype as the protein is not expressed. On the other hand, the pathogenicity of missense variation is more difficult to predict, as the transcript with a missense mutation is still transcribed.

Type of variation	Description
Missense	Nucleotide substitution that forms the codon of a different amino acid
Nonsense	Nucleotide substitution that forms a stop codon, inducing early termination of translation
In-frame	Deletion or insertion of one or more codons, resulting in loss or gain of amino acids but does not affect the reading frame
Frameshift	Deletion or insertion of one or more nucleotides that interrupts the reading frame. May result in early or late termination of translation
Splice site loss	Disruption of the nucleotides that form the splice acceptor (AG) or donor (GU) sites. This can result in the loss or gain of translated mRNA.
Splice site gain	Creation of an acceptor or donor splice site with greater affinity for the splicing machinery than the genuine splice site. This can result in the loss or gain of translated mRNA.

Table 1.5 – Types of human genetic variation that affect protein coding genes

1.11.1.2 Classification of pathogenicity

With over 3 million variants in a single human genome, it is challenging to determine whether a variant is likely to have a deleterious or phenotypic effect. When reporting potentially disease-causing mutations, correct classification of variants is critical to prevent publication of false positive findings. Databases of population frequency data can aid classification of rare, private or common variants. Rare or private variants are more likely to be disease causing, as natural selection is expected to eradicate deleterious disease variants before they reach a high allele frequency (Lupski et al. 2011). Rare variants are typically defined as having a minor allele frequency (MAF) <1%, whereas private mutations are novel, in the sense that they have never been detected in prior sequencing experiments from the same population (Gao and Keinan 2014). Within this project, ExAC and GnomAD databases have been utilised (Lek et al. 2016). ExAC contains over 10 million variants identified by exome sequencing data obtained from 60,706 unrelated individuals. On the other hand, GnomAD database contains data obtained from 15,496 genome sequences, in addition to 123,136 exomes. As the patients within this study were all of European descent, population data from the non-Finnish European population within ExAC and GnomAD were also used, containing 33,370 and 63,369 sequences, respectively (Lek et al. 2016).

A rare SNV is not necessarily pathogenic. Functional prediction algorithms have been developed to help establish the pathogenicity of a variant based on amino acid conservation, molecular interactions of the tertiary structure. Within this thesis, Sorting Intolerant From Tolerant (SIFT) (Ng and Henikoff 2003), PolyPhen-2 (Adzhubei et al. 2010), Protein Variation Effect Analyser (PROVEAN) (Choi et al. 2012; Choi and Chan 2015), and Combined Annotation-Dependent Depletion (CADD) (Kircher et al. 2014) algorithms were used.

SIFT utilises protein sequence homology to ascertain whether an amino acid substitution will result in a phenotypic change. This method assumes that the change of a conserved amino acid will be deleterious. The algorithm will also take the amino acid properties into account, for example, if a conserved polar amino acid is substituted with another polar residue, the change will be predicted to be tolerated (Ng and Henikoff 2003).

PolyPhen-2 utilises a more sophisticated approach for prediction of damaging variants. It uses both protein sequence and structural based prediction features. The structural capability allows elucidation of whether the amino acid change may affect intermolecular interactions, or the interface with complexing ligands or proteins. However, the structural features of PolyPhen-2 can only be utilised if there is a known or homologous 3D structure for the protein of interest (Adzhubei et al. 2010).

PROVEAN is not restricted to missense variants, but can also classify in-frame insertions and deletions (indels). Like SIFT, PROVEAN uses protein homology to ascertain the deleterious nature of variants. In contrast to other sequence homology based algorithms, PROVEAN also assesses the neighbourhood flanking sequences, allowing evaluation of whether the amino acid lies within conserved protein domains. This aimed to reduce the number of false positives reported (Choi et al. 2012).

CADD is a sophisticated scoring system, which allows classification of coding and non-coding SNVs and indels. The algorithm has classified the deleteriousness of millions of observed and simulated variants in the human genome. Deleteriousness is evaluated by comparison of different metrics, including conservation scores, consideration of functional motifs and gene expression information, and knowledge of protein sequence/structure. CADD works by ranking the deleteriousness of an input variant in relation to all possible 8.6 billion substitutions in the human reference genome. The result is a score that ranges from 1 to 99, where higher figures relate to a more deleterious prediction. A recommended cut-off for evaluation of deleterious variants is 15-20 (Kircher et al. 2014).

Of note, functional prediction algorithms should be interpreted with caution as false positive rates are high; for example, SIFT has a 20% false positive error rate (Ng and Henikoff 2003).

Algorithms can also present discordant predictions. Therefore, it is advisable to use the strength of multiple prediction methods.

1.11.2 Transcriptomics by RNA-sequencing

Whereas an organism's DNA sequence is highly stable and relatively static, tangible cellular processes are far from it. RNA takes the form of DNA's messenger, where the transcription of DNA into messenger RNA (mRNA) defines the cell identity. In contrast to DNA, RNA is highly dynamic and reflects a snapshot of a cellular physiological state. The whole transcriptome refers to the total RNA of a given tissue or cell type. Although there are different types of RNA molecule, mRNA transcripts were utilised within this thesis. Before the advent of NGS technologies, transcriptome studies implemented microarray or quantitative real-time PCR (qPCR). These techniques are targeted and use specific probes/primers, thereby requiring existing knowledge of the genome. By contrast, RNA-seq allows the complete cataloguing and quantitation of mRNA transcripts of a given sample within a single reproducible, high-throughput assay (Wang et al. 2009). Transcript quantitation is determined as a measure of read depth (Nagalakshmi et al. 2010), which provides a direct and simple read-out for gene expression levels.

RNA-seq can be used to answer many different biological questions that other quantitative RNA arrays cannot remedy. It allows determination of transcriptional architecture by providing information about transcriptional start sites, the location of 5' and 3' ends, alternative splicing patterns, and post transcriptional modifications (Ozsolak and Milos 2011). Recently, Cummings et al. (2017) demonstrated the utility of RNA-seq for diagnosis of Duchenne muscular dystrophy patients manifesting variants that affected splicing. They identified both coding and non-coding pathogenic variants that resulted in aberrant splicing, including two heterozygous synonymous variants in different patients, where the single base change created a donor splice site within an exon, but did not change the amino acid sequence (Cummings et al. 2017). This study highlights the need for sequencing beyond the genome, as the variants within this study would otherwise not have been possible to resolve with DNA sequencing methods alone.

As the whole mRNA transcriptome of a tissue is sequenced, RNA-seq also provides a vast amount of cellular and biological information, allowing investigation of how gene expression is affected by genetic variation. Differences in gene expression between biological groups can be established by identification of differentially expressed genes, which enables quantification of different conditions, such as response to treatment or disease.

1.11.3 Proteomics

The term proteomics originated in the 1990s to describe the study and detection of an organism's protein composition, as well as analysis of their structure and function (James 1997). In contrast to nucleic acids, proteins are notoriously difficult to study, as each can differ by size, molecular charge, posttranslational modifications (PTMs), or structural conformation. Indeed, from 23,000 genes in the human genome, there are an estimated 100,000 protein variants, including splice variants, cleaved polypetides and post-translationally modified proteins (Wilhelm et al. 2014). In recent years, NGS technologies have favoured nucleic acids, which takes advantage of innate chemical properties, such as hybridisation. However, although the central dogma dictates that DNA makes RNA makes protein, in reality this is not an equivalent process, as correlation of abundances between RNA and protein have been reported as only ~40% similar (de Sousa Abreu et al. 2009). As the functional product of a cell, reliable protein sequencing techniques are required that can sensitively detect and quantify an organism's proteome. Protein composition is a direct indication of a time-specific snap shot of a cell. By investigation of the proteome, there is an opportunity to gain information that is immediately instructive of a tissue's pathobiology that the genome and transcriptome cannot offer.

Proteomic techniques can either be targeted or untargeted, depending on the biological questions that need to be answered. An untargeted strategy is more challenging as resolution and assay sensitivity is reduced for complex mixtures. On the other hand, targeted proteomics is limited to identification of optimised proteins. The techniques that have been used in this project are outlined below.

1.11.3.1 Mass spectrometry

Although engineered for both targeted and untargeted capabilities, within this thesis, mass spectrometry has been used only for untargeted analysis. Over the last 20 years, there have been new technologies released to enhance the sensitivity, reliability, and simplicity of data generation (Kuharev et al. 2015; Cravatt et al. 2007; Crutchfield et al. 2016). Mass spectrometry has evolved into a powerful analytical method where tissues, cell types and even whole organisms have had their proteomes identified (Nagaraj et al. 2011; de Godoy et al. 2008; Wilhelm et al. 2014). Concurrently, the study of proteins by mass spectrometry can produce variable results, influenced by factors such as sample preparation, undersampling, and storage; thus, judicious study design and expertise is required to complete reliable experiments.

The mass spectrometer detects compounds through at least two distinct characteristics; mass and charge. The mass spectrometer can be easily coupled to other separation techniques, such as chromatography, to distinguish between compounds by chemical properties (Hawkridge and Muddiman 2009). At present, most mass spectrometers do not have the resolution to dependably recognise large proteins. Therefore, cleavage of proteins into peptides using the proteolytic enzyme trypsin is common prior to mass spectrometry in an approach called 'bottom up' proteomics. Trypsin is particularly suited to mass spectrometry since it specifically cleaves a protein on the C-terminal side of an arginine or lysine residue, which aids peptide identification (Meissner and Mann 2014). The value of any profiling technique lies with accurate quantitation. This can be a challenge for mass spectrometry, as it is not intrinsically quantitative, due to different ionisation efficiencies of different analytes (Zhu et al. 2010). To overcome this, many quantitation techniques have been developed, which are grouped into two classes, label-free and isotopically labelled mass spectrometry (Schulze and Usadel 2010).

1.11.3.2 Targeted proteomic techniques

For sensitive and reliable mass spectrometry, readily ionisable and low complexity samples are required. A targeted proteomic strategy provides increased assurance of sensitivity where mass spectrometry can fall short. Antibody based approaches are currently the gold standard for individual protein detection, as a well-derived antibody has high affinity and specificity to its target (Lollo et al. 2014). Western blotting is a simple and widely used method for identification of a protein from a sample. Similarly, enzyme-linked immunosorbent assay (ELISA) uses antibodies to specifically bind to a protein, but can quantify this analyte through addition of an enzyme-linked secondary antibody that chemically emits a fluorescent signal. Recently, ELISAs have been developed into multiplexed assays where many analytes can be tested in a sample (Elshal and McCoy 2006). However, the prosperity of this technique depends on antibody availability for each analyte, and critically developing antibodies that do not cross-react.

Within this thesis a novel targeted proteomics technology was trialled, called SOMAscan (SOMAlogic Inc, Boulder, CO, USA). This was inspired by the specificity of the ELISA, with the added advantage that it can be multiplexed to analyse over 1000 proteins (Gold et al. 2010). This utilises aptamers, compounds derived from nucleic acid structures that can specifically recognise protein epitopes (Lollo et al. 2014). Over several years of development, SOMAscan was established to contain over 1300 aptamers with high specificity, sensitivity, and no risk of cross-

reactivity commonly seen with antibody-dependent assays (Gold et al. 2010). This technology was tested and evaluated in chapter 4.

1.12 Pathway analysis

Complete sequencing of RNA is an effective and sensitive tool for quantitation of the transcriptome. However, with reports that up to three quarters of the human genome is capable of being transcribed (Djebali et al. 2012), this sequencing technique produces a vast wealth of data that is not trivial to interpret. Although identification of differentially expressed transcripts or proteins can reduce this data into statistically significant changes, the resulting list of genes/protein provides limited biological insight. It is well known that gene products do not function as discrete components, but in large, often overlapping, biological processes. Pathway analysis provides a method of classifying findings from RNA-seq or proteomic profiling experiments into statistically driven, informative groups. Not only does this reduce data complexity, but it also allows identification of pathways that are different between patient and control groups, which may link to disease pathology. Multiple knowledge based tools are available for pathway analysis. Within this thesis, gene set enrichment analysis (GSEA), Protein Analysis Through Evolutionary Relationships (PANTHER) overrepresentation testing, and PANTHER statistical enrichment testing were implemented.

1.12.1 Gene Set Enrichment Analysis

GSEA was developed to help address common issues related to working with large transcriptome datasets. Statistical analysis to find differentially expressed genes can result in two scenarios; firstly, one may be inundated with thousands of significant genes, with indeterminate biological meaning; or secondly, correction for multiple testing hypothesis may result in too few genes for further analysis. Moreover, cellular systems act in a balance, where an upregulation of a single gene may impact other pathways. Single gene analysis would be incapable of providing the total cellular processes within the experiment. Finally, repetition of experiments by different research groups can result in little overlap at a gene level (Fortunel et al. 2003). Investigation of pathways or networks affected may provide greater understanding that is more reproducible (Subramanian et al. 2005).

The effectiveness of GSEA relies on robust reference gene sets. There are over 10,000 gene sets specifically developed for use with GSEA in the Molecular Signatures Database (MSigDB) (Liberzon et al. 2011). Within this project, the Hallmark gene sets have been applied. The Hallmark database

was developed by MSigDB to reduce redundancy and heterogeneity between the exhaustive gene sets already present in MSigDB (Liberzon et al. 2015). The resulting database features 50 reviewed and manually annotated gene sets, containing 7343 genes in total. These extensively cover cellular pathways involved in biological processes, molecular functions and cellular component, such as proliferation, signalling, immunology, metabolism, development, and DNA damage (Liberzon et al. 2015).

1.12.2 PANTHER Statistical Overrepresentation Test

The PANTHER classification system is an open access database containing manually curated and computationally generated gene sets (Mi et al. 2013; Mi et al. 2017). The database is comprised of gene ontology (GO) terms (Gene Ontology Consortium 2015), spanning terms associated with molecular function, biological processes, and cellular components. In addition to these, there are gene sets associated with pathways, including their own PANTHER pathways, and the Reactome pathways (Croft et al. 2011). As a result, PANTHER encompasses 57,271 annotations in 13,387 genes (Mi et al. 2017).

Statistical overrepresentation analysis is a relatively simple test for identification of pathways within a set of genes/proteins with a given threshold. The input list contains genes/proteins that are separated by up or downregulation and usually a cut off of fold change or a statistical value, such as p-value (Khatri et al. 2012). The input list is then compared against the PANTHER database of annotated genes and categorised accordingly. A 2x2 statistical test is used to ascertain whether the number of genes are statistically overrepresented or underrepresented for a given pathway compared to what is expected for a random sample of genes of the same size (Goeman and Bühlmann 2007). The output is the fold enrichment between the observed number of genes within a pathway and the expected number from the 2x2 test. This method is fairly crude as it does not take independent changes into account, and thus each gene has equal weighting in the analysis. Furthermore, it relies on manually selected genes according to selection criteria, disregarding the remaining identifications of the experiment (Khatri et al. 2012).

1.12.3 PANTHER Statistical Enrichment Test

The final pathway analysis used in this study was the PANTHER statistical enrichment test. It works similarly to GSEA, where the entire list of genes identified by sequencing are tested. Genes are ordered by fold change and plotted to show the distribution of fold change over the whole input set. Next, genes are categorised into pathways according to gene annotations within the

PANTHER database. Pathway classification for PANTHER enrichment analysis utilises the Reactome pathways knowledgebase (Croft et al. 2011), a database of hierarchically organised processes, containing 8701 annotated genes (Fabregat et al. 2016). The fold change of all genes within a given pathway are then compared with the total gene set to determine whether they are statistically positively or negatively enriched (figure 1.6). The benefit of statistical enrichment over the overrepresentation test is that all genes are considered and up and downregulated genes are analysed simultaneously, allowing ascertainment of the coordinated differential regulation of a pathway.

1.12.4 Pathway topology mapping

Identification of gene products that function in pathways only provides limited information about the actual interactions that exist in a tissue. Pathways do not exist as separate linear entities, but in fact are highly dynamic and cross-talking. Furthermore, the same protein may function in different pathways. Classification of protein-protein interactions within a biological pathway or protein set provides valuable information about protein complexes and interconnecting networks. Within this study, STRING database was used to identify how functional gene products may interact. STRING functions by assessment of both direct physical interactions and indirect functional associations (Szklarczyk et al. 2015). Interactions stored within the STRING database exist from 5 sources. They may be known from experimental interactions, found in primary manually curated bases, determined through text mining, known interactions through co-expression analysis, or interactions based on homology in one or more different organisms (Szklarczyk et al. 2015).

1.13 Induced pluripotent stem cells

The final chapter of this thesis utilises a relatively novel phenomenon that has transformed the way researchers can use cell culture to understand disease pathology; induced pluripotent stem cells (iPSCs). Stem cells are embryonic cells capable of self-renewal and differentiation into any of the three germ layers. Human embryonic stem cells (ESCs) have been used to model development and disease since they were first isolated by Thomson and colleagues in 1998 (Thomson et al. 1998). However, due to the ethical impediments associated with ESCs, researchers attempted to reverse differentiated cells through addition of transcription factors absolutely required for stem cell pluripotency. In 2006, this was achieved by Takahashi and Yamanaka through retroviral infection of the transcription factors OCT4, SOX2, KLF4, and c-MYC in murine embryonic fibroblasts (Takahashi and Yamanaka 2006). This combination of genes was

affectionately termed the Yamanaka cocktail, and was subsequently used to reprogram adult fibroblasts (Takahashi et al. 2007). These iPSCs were established to be analogous to ESCs in morphology, gene expression, and their ability to differentiate into the three germ layers (Takahashi et al. 2007; Yoshida and Yamanaka 2010). iPSCs have been presented as an indispensable tool to help uncover disease mechanisms from patient cells, bypassing the dependency on animal models. iPSCs can be derived from patients and differentiated into cells that are inaccessible by biopsy, such as neurons (Wang et al. 2015; Kim et al. 2014). This has allowed researchers to effectively study disease development, determine molecular mechanisms of disease, or trial drug and gene therapy targets (Yoshida and Yamanaka 2010; Sterneckert et al. 2014).

iPSCs used in this thesis were derived from fibroblasts and reprogrammed by the human induced pluripotent stem cell initiative (HipSci) as the Wellcome Trust Sanger Institute, Cambridge, UK (Kilpinen et al. 2017). This initiative has reprogrammed nearly 500 cell lines from healthy donors and 86 lines from patients with rare genetic diseases, of which BBS constitutes 50 lines (Streeter et al. 2017). Each line has been comprehensively assayed, creating phenomic, genomic, transcriptomic, proteomic, and epigenomic information, the majority of which is open access and available via their data portal (www.hipsci.org). iPSCs reprogrammed by HipSci from 3 BBS patients and 2 healthy controls have been used as a model to investigate retinal dystrophy, presented in chapter 6.

1.14 Objectives of this thesis

This project functioned alongside the pilot initiative for the NIHR GOSH BRC funded HIGH-5 project, which had three general aims with the goal of personalising patient care. These were to elucidate the pathobiology of rare diseases through NGS technologies, determine genetic or regulatory factors that may attribute to heterogeneity of rare diseases, and identify novel biomarkers that may help to predict or pre-empt disease onset. Experiments within this thesis have been designed to address these aims with respect to the rare disorder, Bardet-Biedl Syndrome.

15 BBS patients were recruited for this study, where all individuals shared the same underlying diagnosed variant, *BBS1* c.1169T>G/ p.M390R, inherited either as a homozygous or compound heterozygous SNV. Recruitment criteria required patients to consent to use of biological material

for research purposes, namely a skin biopsy, whole blood, and urine. Each type of biological material was probed by different technologies to address the aims of this project (**figure 1.7**).

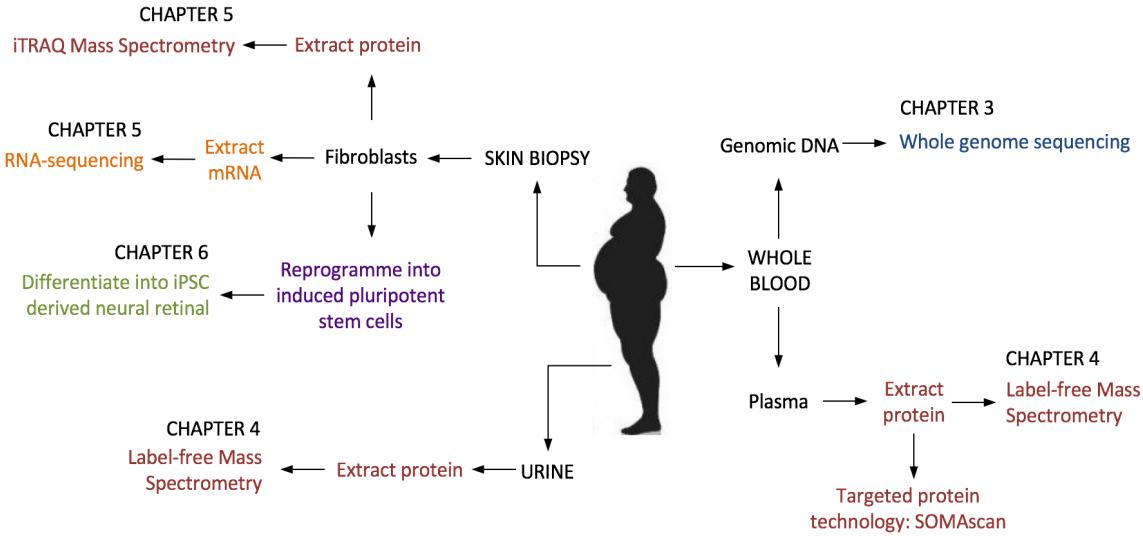

Figure 1.5 – Overall study outline that displays how biological materials, such as skin biopsy, urine, or whole blood, were investigated by different technologies for each thesis chapter.

Genomic DNA and plasma were extracted from whole blood, which were used for whole genome sequencing and proteomic profiling, respectively. Protein was also isolated from urine for proteomic label-free mass spectrometry. Patient-derived dermal fibroblast cells were obtained from skin biopsies, which were used for transcriptomic and proteomic profiling. These cells are responsible for extracellular matrix organisation in the dermis, as well as regulation of skin physiology (Sorrell and Caplan 2004). Furthermore, they can be readily cultured from a small biopsy obtained by a minimally invasive procedure, making them a suitable cell type for a pilot project, such as this one. Fibroblasts have the additional advantage that they can be reprogrammed into iPSCs. In addition to BBS patient-derived material, plasma, urine, fibroblasts, and iPSCs were obtained from clinically healthy control individuals.

Ultimately, within this thesis, genomic, transcriptomic, and proteomic technologies were implemented in a multi-omic strategy to investigate the following objectives:

- Determine whether interfamilial heterogeneity, which presents as differences in disease onset or severity, is a consequence of genetic modifying alleles that interact with BBS1 in the primary cilium. Do these variants contribute to enhance the mutational burden of BBS, when compared to control individuals?

- Utilise proteomic profiling technologies to discover molecular biomarkers that may be used to measure disease progression or detect secondary features of BBS from urine and plasma biofluids.

- Implement a multi-omic strategy via integration of transcriptomic and proteomic profiling to uncover differentially regulated pathways between BBS and control fibroblasts that may elucidate BBS pathology.

- Develop a tissue specific model from BBS patient-derived iPSCs for the investigation of retinal dystrophy *in vitro*.

Chapter 2: Materials, Methods, and Patients

2.1 Sample Collection

2.1.1 Ethics

Samples were obtained from consented patients attending the Bardet-Biedl Syndrome multidisciplinary clinic at Guy's Hospital of Guy's and St Thomas' NHS Foundation Trust, London. Recruitment to the study required signed consent for use of all samples, including a skin biopsy, whole blood and urine. Ethical consent to participate in the research study was granted by The National Research Ethics Service as part of the EU rare diseases registry for Wolfram, Alström, and Bardet-Biedl Syndromes (EURO-WABB) study, supported by the European Commission under the Health Programme Framework (Farmer et al. 2013). Upon collection of patient samples, they were anonymised and appropriately processed.

2.1.2 Plasma isolation of whole blood

Whole blood samples from 15 patients and 6 controls were obtained in the fasted state and were processed on the same day. Up to 5ml of blood was collected in an ethylenediaminetetraacetic acid (EDTA) vacutainer (Sarstedt) and mixed well to ensure chelation. 1ml aliquots of whole blood were dispensed in clean labelled microcentrifuge tubes and centrifuged at 5000*g* for 10 minutes at 4°C. The plasma was isolated as the supernatant, which was stored in 500µl aliquots at -80°C.

2.1.3 Skin biopsy

A skin biopsy was taken using a 5mm biopsy punch. The biopsy was taken from the back of the upper arm, which was sterilised with an alcohol wipe and Lignocaine local anaesthetic was administered. The biopsies were taken by Professor Phil Beales or Dr Elizabeth Forsythe at the BBS clinic, Guy's Hospital, London. Each biopsy was cut in half and separately placed into sterile Dulbecco's Modified Eagle Medium (DMEM) (Gibco) supplemented with 10% fetal bovine serum (FBS) (Gibco) and 1% Penicillin-Streptomycin (10,000U/ml) (Pen-Strep) (Gibco). One half of the biopsy was couriered to the Wellcome Trust Sanger Institute, Cambridge, UK to be reprogrammed into induced pluripotent stem cells (iPSC), whereas the other was kept for fibroblast culture at the Institute of Child Health, UCL.

2.1.4 Urine collection

At least 10ml of fasted midstream urine was collected from 15 patients and 6 controls. On the day of collection, these were aliquoted into 2ml volumes and stored at -80 °C.

2.2 Fibroblast culture

2.2.1 Culturing fibroblasts from a skin biopsy

All cell culture work was carried out in a class II tissue culture hood and aseptic technique was implemented throughout the procedure.

2.2.1.1 Plating the skin biopsy

Using a scalpel, the biopsy was cut through the epidermis into 4-8 approximately equal segments. Each piece was compressed, skin side down, onto a T25 tissue culture flask. To facilitate adherence, the tissue was left at room temperature for 10 minutes before 2ml of DMEM, 10% FBS, 1% Pen-Strep was added. The biopsies were incubated for 4 hours at 37°C, 5% CO_2, after which an additional 3ml media was added. Flasks were kept at 37°C, 5% CO_2, monitored for growth daily and media changed every 3-4 days. Fibroblasts started to appear between 7-14 days and were passaged when colonies surrounding the tissue were 80-90% confluent.

2.2.1.2 Passaging fibroblasts

To passage fibroblasts in a T25 flask, the spent media was removed and cells washed with 5ml 1x Dulbecco's phosphate-buffered saline (DPBS). To detach cells, 2.5ml Trypsin-EDTA, 0.05% (Gibco) was added and incubated for up to 5 minutes at 37°C or until cells has dissociated. Once ~80% of cells had detached, 3ml of DMEM, 10% FBS, 1% Pen-Strep was added to inhibit the proteolytic activity of trypsin. The trypsin-DMEM mix was transferred to a 15ml falcon and centrifuged at 1000*g* for 5 minutes, room temperature. The supernatant was discarded, pellet resuspended in DMEM medium and split at a ratio of 1:3. **Table 2.1** shows seeding densities for each flask type used.

Flask/ plate	Surface area (cm^2)	Seeding density	Cells at confluency	Volume of growth medium (ml)	Volume of trypsin-EDTA, 0.05%	Number of cryovials per flask
T25	25	0.7×10^6	2.5×10^6	5	2.5	1
T75	75	2.1×10^6	7.5×10^6	15	5	3
T175	175	4.6×10^6	1.75×10^7	35	10	7
6-well	9	0.3×10^6	1.2×10^6	2	1	

Table 2.1 – Seeding densities for fibroblasts in different flasks or plates used

2.2.1.3 Freezing and thawing fibroblasts

After trypsinisation and centrifugation (as above), the cell pellet from a T25 flask was resuspended in 1ml freezing media, consisting of 90% supplemented DMEM and 10% dimethyl sulfoxide (DMSO) (Sigma-Aldrich). Cells were transferred to a cryovial and initially stored in a Mr. Frosty freezing container (ThermoFisher) filled with isopropanol at -80°C. This allowed gentle cryopreservation before long-term storage in liquid nitrogen (LN_2). Each cryovial of cells was thawed in a water bath at 37°C and resuspended with 4ml of DMEM, 10% FBS, 1% Pen-Strep. After centrifugation for 5 minutes at 1000*g*, the supernatant was discarded. The pellet was resuspended in 5ml supplemented DMEM and transferred to a T25 flask prior to incubation at 37°C, 5% CO_2.

2.2.2 Treated and untreated fibroblasts for RNA extraction

Fibroblasts from 15 patients and 6 control subjects were plated in 2 wells of a 6-well plate at a density of 0.3 x 10^6 cells. When cells reached 90% confluency in both wells, they were separately processed for either serum starved (treated) or serum fed (untreated).

The untreated sample was washed with three times 1x DPBS. On the final wash, the cells were dissociated from the culture dish using a scraper and collected in a microcentrifuge tube. A pellet was formed by centrifugation at 1000*g* for 5 minutes at 4°C and supernatant discarded. QIAzol lysis reagent (QIAGEN) was added to the pellet and homogenised using a 23G needle and 1ml syringe. The samples were either processed for RNA extracted immediately or stored at -80°C for processing later. Simultaneously, the second well was treated with serum-depleted media to induce ciliogenesis. The well was washed three times with DPBS and 2ml of DMEM, 0.5% FBS, 1% Pen-Strep was added. After 48 hours, the cells were washed and collected by scraping, as outlined above.

2.2.3 Treated and untreated fibroblasts for protein extraction

As a great yield of protein was required for mass spectrometry, fibroblasts from 15 patients and 6 control subjects were grown in 4x T175 flasks. Fibroblasts were seeded in each T175 flask at a density of 3.2 x 10^6 cells. When 90% confluency was attained, 2 flasks were washed in ice cold DPBS three times. After the final wash, 5ml of DPBS was added to each flask and cells were collected using a cell scraper and transferred to a 15ml falcon. This step was repeated with an additional 5ml DPBS and centrifuged at 4°C for 5 minutes at 1000*g*. The supernatant was discarded and the pellet stored at -80°C until required for further processing. Simultaneously, the

remaining 2 flasks were washed three times and treated with serum depleted media (DMEM, 0.5% FBS, 1% Pen-Strep) for 48 hours. Cells were collected by washing and scraping, as described above.

2.3 Nucleic acid extraction

2.3.1 Isolation of DNA

Genomic DNA (gDNA) from whole blood was collected in EDTA vacutainers (Sarstedt) and extracted by the laboratories of the North-East Thames Regional Genetics Service, Great Ormond Street Hospital, Barclay House, London. DNA quality, purity and quantity were assessed in-house by 1% agarose gel electrophoresis, NanoDrop 1000 (ThermoFisher) spectrophotometry and Qubit 2.0 (ThermoFisher) fluorometry assay, respectively.

gDNA from 5 iPSC samples was extracted for Sanger Sequencing using the QIAamp DNA mini kit (QIAGEN). Extraction was undertaken according to the manufacturer's instructions. 20µl of QIAGEN protease was mixed with 200µl of sample and 200µl AL buffer, and incubated at 56°C for 10 minutes to lyse the cells. 200µl of 100% ethanol was added and pulse-vortexed for 15 seconds. The entire mixture was passed through a QIAamp mini spin column by centrifuging at 6000*g* for 1 minute. The flow-through was discarded and the column placed in a fresh collection tube, where 500µl of buffer AW1 was added and centrifuged at 6000 *g* for 1 minute. The column was passed in another fresh collection tube, 500µl of buffer AW2 was added and centrifuged at maximum speed for 3 minutes. The flow-through was discarded and column transferred to a 1.5ml microcentrifuge tube. 30µl of elution buffer AE was added and incubated at room temperature for 5 minutes. Finally, DNA was isolated by 6000*g* centrifugation for 1 minute before quantitation by Qubit 2.0 fluorometry assay.

2.3.2 RNA extraction

RNA was extracted from fibroblasts and differentiating optic cup samples using the miRNeasy mini kit (QIAGEN) according to manufacturer's instructions. Surfaces and equipment were cleaned with RNase AWAY decontamination reagent (ThermoFisher) before extraction. Homogenised cells in QIAzol were thawed at room temperature and were shaken vigorously for 15 seconds with 140µl of chloroform in a fume hood. After incubation at room temperature for 3 minutes, samples were centrifuged for 15 minutes at 12,000*g* at 4°C to generate three layers. The upper aqueous phase was transferred to a 1.5ml RNase free microcentrifuge tube, combined with 600µl 100% ethanol and mixed thoroughly by pipetting. Each sample was passed through an

RNeasy Mini spin column by centrifugation at 10,000*g* for 15 seconds at room temperature. The column was retained and 350µl RW1 buffer washed the column after centrifugation at 10,000*g* for 15 seconds. RNase-free DNase I (QIAGEN) mix was added to the column to remove any contaminating DNA and incubated for 15 minutes at room temperature. The column was washed again with 350µl RW1 buffer by centrifugation at 10,000*g* for 15 seconds. Next, x2 500µl RPE buffer was added, the first centrifuged for 15 seconds and the second for 2 minutes, both at 10,000*g*. Finally, 30µl of RNase free-water was added to the column to elute RNA into an RNase-free microcentrifuge tube. Eluted RNA was quantified and stored at -80°C.

2.3.3 Assessment of nucleic acid quantity and purity

DNA and RNA samples were quantified by the Qubit 2.0 fluorometer using the Qubit dsDNA broad range (BR) or RNA BR assay kit (ThermoFisher), respectively. The protocol for each nucleic acid type is identical, except the reagents differ between kits. To produce the Qubit working solution, Qubit BR fluorescent reagent was diluted 1:200 in Qubit BR buffer and vortexed well. Two standards were prepared by diluting 10µl of each standard in 190µl working solution. 1µl of each sample was diluted in 199µl working solution. 500µl thin-walled PCR tubes (ThermoFisher) were used and samples were left to incubate at room temperature for 2 minutes. The concentration of each sample was calculated by the fluorometer obtained from the standard curve.

Within the NanoDrop 1000 (ThermoFisher) operating software, the nucleic acid module, followed by either DNA or RNA was selected. The NanoDrop was blanked with ddH_2O water and calibrated with elution buffer prior to analysis of 1µl of sample. The A_{260}/A_{280} ratio output was recorded to assess the purity of each sample. A ratio of ~1.8 or ~2.0 was considered pure enough for DNA and RNA, respectively.

2.3.4 gDNA agarose gel electrophoresis

A 1% gel was made by dissolving 0.5g of UltraPure Agarose (Invitrogen) in 50ml 1x Tris/Borate/EDTA (TBE) buffer (BioRad) in a microwave. Once cooled to ~50°C, 1µl Ethidium Bromide (Sigma-Aldrich) was added as a fluorescent tag and mixed well. The solution was poured into a casting tray with combs. Once the gel was set, the comb was removed to reveal wells and transferred to the electrophoresis tank containing 1x TBE buffer. 1µl of genomic DNA was mixed with 4µl 5x Green GoTaq Flexi buffer (Promega) and 5µl of each sample was loaded to each well on the gel. 4µl of Quick-Load 1kb DNA ladder (New England Biosciences) was added to the first

well as a marker. The gel was run at 110 volts for 30 minutes. Bands were visualised using the Gel Doc XR+ ultraviolet (UV) imaging system (Bio-Rad).

2.4. Protein extraction

2.4.1 Protein extraction from urine

Protein was isolated from the urine of 15 patient and 6 control subjects for liquid chromatography mass spectrometry (LC-MS). Urine was thawed on ice, vortexed well and centrifuged at 13,000 rpm for 15 minutes. 1.5ml of supernatant was mixed with 3ml of ddH_2O and added to Amicon Ultra-4 centrifugal filter units with 3kDa molecular weight cut off (Merck Millipore). Urine was filtered by ultracentrifugation at 4,000 rpm for 50 minutes. The protein concentrate was recovered. This ultrafiltration technique was utilised to concentrate the protein, as well as remove low molecular weight metabolites, ions and detergents. The filter was washed with 100µl of 50mM ammonium bicarbonate and added to the concentrate. Protein was quantified by Pierce bicinchoninic acid (BCA) assay (ThermoFisher).

2.4.2 Protein enrichment by ProteoMiner kit

Protein was isolated from the plasma of 2 patient subjects, BBS010 and BBS011, for LC-MS. To remove the most abundant plasma proteins from the samples, protein enrichment was implemented using the ProteoMiner protein enrichment small-capacity kit (Bio-Rad). This was followed according to manufacturer's instructions.

Plasma was thawed on ice, vortexed well and centrifuged for 10 minutes at 10,000 rpm. 200µl of plasma was applied to 20µl of primed beads on the ProteoMiner column and incubated on a rotational shaker for 2 hours at room temperature. The column was centrifuged at 1,000*g* for 60 seconds and flow-through was discarded (one sample flow-through was retained to run on SDS PAGE). 200µl wash buffer was added and incubated for 5 minutes at room temperature on a rotator to ensure proper washing. The wash was removed by centrifugation at 1,000*g* for 60 seconds. The wash step was repeated twice more (a flow-through from each wash of one sample was retained for SDS-PAGE). The column was incubated with 200µl deionised water for 1 minute and centrifuged at 1,000*g* for 60 seconds. To elute the enriched protein, 20µl of elution reagent was incubated for 15 minutes with occasional vortexing. The protein was eluted by centrifugation into a microcentrifuge tube at 1,000*g* for 60 seconds. The elution step was repeated twice more and pooled together. Protein was quantified by BCA assay.

2.4.3 Homogenisation of fibroblasts

Cell pellets from serum starved and serum fed fibroblasts of 15 patient and 6 control subjects were thawed on ice, resuspended in 500µl 50 mM Ammonium bicarbonate pH 7.8, 2% Amidosulfobetaine (ASB)-14 and added to 0.5ml homogenisation tubes with 1.4mm ceramic beads (Bertin Technologies). Samples were bead homogenised in the Minilys bench top homogeniser (Bertin Technologies) for 3x 60 second at 3000 rpm, placed on ice between cycles. The homogenate was transferred to a fresh microcentrifuge tube and centrifuged at 13,000 rpm for 10 minutes at 4°C. The supernatant was kept and quantified by BCA assay.

2.4.4 Protein quantitation by BCA assay

Protein was quantified using the Pierce BCA assay kit (ThermoFisher), which was completed according to manufacturer's instructions in a 96-well microplate suitable for absorbance assays. Bovine serum albumin (BSA) standards were generated by serial dilution to create a standard curve at concentrations of 2000, 1500, 1000, 750, 500, 250, 125 ng/µl. Working solution was created by dilution of BCA reagent B (containing 4% cupric sulphate) with BCA reagent A (containing sodium carbonate, sodium bicarbonate, bicinchonicic acid and sodium tartrate in 0.1M sodium hydroxide) at a 50:1 ratio. 2µl of standard or sample was added to each well in triplicate, followed by 200µl of working solution. The plate was covered and incubated at 37°C for 1 hour. Protein concentration was measured at 555nm and 595nm by an Infinite 200 PRO NanoQuant plate reader (Tecan). The absorbance reading for each standard was used to generate a standard curve in Microsoft Excel. The curve was used to calculate the concentration of each sample and the volume of protein required for downstream analysis. In-house mass spectrometry required 50µg from plasma, 25µg from urine and 50µg from fibroblasts. An additional 300µg from fibroblasts was sent to BGI for isobaric tag for relative and absolute quantitation (iTRAQ) mass spectrometry, and 20µg from plasma was used for SOMAscan assay (SOMAlogic).

2.5 Protein preparation for unlabelled liquid chromatography mass spectrometry

2.5.1 Acetone protein precipitation

Protein extracted from plasma and urine for in-house mass spectrometry were acetone precipitated to desalt and help purify the samples. The correct quantity of protein calculated from the BCA assay standard curve was aliquoted to new microcentrifuge tubes. Ice cold acetone was added to the protein at a ratio of 3:1, vortexed well and incubated overnight at -20°C. The

remaining protein was stored at -80°C. The next day, samples were centrifuged at 13,000 rpm for 10 minutes at 4°C and supernatant discarded. 100µl of ddH_2O was added to the pellet, vortexed and dried overnight using the Modulyo freeze-drier (Edwards). Dried pellets were further processed for mass spectrometry.

2.5.2 Digest and clean-up of protein from urine

Freeze-dried samples were reconstituted in 20µl of 100mM Tris, pH 7.8, containing 8 M urea and 2% ASB-14 and incubated on a shaker at 900 rpm for 1 hour at room temperature. 1.5µl dithioerythritol (DTE) (Sigma-Aldrich) (3mg DTE in 100µl 50mM Ammonium bicarbonate, pH 7.8) was added and incubated on a shaker for 1 hour at 900 rpm. 3µl of Iodoacetamide (IAA) (Sigma-Aldrich) (36mg/ml in 50 mM Ammonium bicarbonate, pH 7.8) was added and incubated on a shaker at room temperature for 30 minutes at 900 rpm in the dark. 11.5µl Trypsin Gold (Promega, mass spectrometry grade) and 155µl of ddH_2O was mixed well with each sample and incubated overnight at 37°C.

To remove salts and buffers before mass spectrometry, digests were cleaned up using reverse phase C18 columns. Each column was first washed with 1ml 50% acetonitrile containing 0.1% trifluoroacetic acid (TFA), then primed with 2x 1ml 0.1% TFA. Each sample was diluted 1:1 with 0.2% TFA, added to a primed column and allowed to slowly drip through. The sample was collected and reapplied to the column. Residual salts were washed away by addition of 2x 1ml of 3% acetonitrile containing 0.1% TFA. Cleaned peptides were collected from the column by addition of 2x 250µl of 50% acetonitrile containing 0.1% TFA. The peptides were freeze-dried overnight and stored at -80°C until required for mass spectrometry.

2.5.3 Protein preparation from plasma

2.5.3.1 SDS-PAGE

Protein enriched from 2 patient plasma samples were separated via sodium docecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) prior to enzymatic digestion into peptides. In addition, 3 flow-through collections retained from enrichment via ProteoMiner were separated by SDS-PAGE.

Freeze-dried samples after acetone precipitation were reconstituted in 10µl of 1x SDS-Tris (Sigma-Aldrich) and incubated at room temperature for 1 hour. 1.5µl DTE (3mg DTE in 100µl 1x SDS-Tris) was added and incubated on a shaker for 1 hour at 900 rpm. 3µl of IAA (36 mg/ml in

50mM Tris, pH 7.8) was added and incubated on a shaker at room temperature for 45 minutes at 900 rpm in the dark. A further 15.5µl of 1x SDS-Tris was added and incubated at 95°C for 5 minutes. The protein electrophoresis gel tank (Bio-Rad) was prepared and filled with 1x Tris/Glycine/SDS running buffer (ThermoFisher). A precast 4-15% Mini-PROTEAN TGX stain-free gel (Bio-Rad) was placed into the buffer chamber, where 5µl of colour prestained protein standard (New England Biosciences) was added to the first well. 30µl (~50 µg) of each sample was added to the remaining wells. The gel was run at 200V for 30 minutes. Subsequently, the gel was removed and placed face down in Coomassie blue stain (40% methanol, 7.5% acetic acid, 0.1% Coomassie in ddH_2O) and incubated on a shaker for 30 minutes. The Coomassie blue was removed and gel washed three times in destainer (40% methanol, 7.5% acetic acid in ddH_2O) for 1 hour each, the final wash was left overnight.

Samples for mass spectrometry were cut into bands and profiled individually. Each sample was cut into 8 pieces, where stronger bands were cut into smaller segments so that each piece contained approximately ~6.25µg of protein. Each band was placed in a separate microcentrifuge tube and washed in 200µl of 50mM ammonium bicarbonate, pH 7.8 three times, vortexed well between each wash. After the final wash, 500µl 100% acetonitrile was added to dehydrate the gel pieces and incubated on a shaker for 30 minutes at 900 rpm. The acetonitrile was removed and samples dehydrated completely in the Vacufuge concentrator (Eppendorf) on V-AL setting for 45 minutes at 45°C.

2.5.3.2 In-gel trypsin digest and clean up

60µl of Trypsin Gold (mass spectrometry grade) was added to each band and incubated at room temperature until the gel bands had expanded (approximately 30 minutes). 50mM ammonium bicarbonate, pH 7.8 was added to each band until covered, mixed well and incubated at 37°C overnight.

To clean up the peptide mixture, 200µl of 1% formic acid was added to each band, vortexed well and incubated on the shaker for 20 minutes at 900 rpm. The solution was transferred to a fresh microcentrifuge tube and 2x 300µl 50% acetonitrile, 1% formic acid was added to each band, shaking for 20 minutes and pooling each wash. Finally, 200µl of 1% formic acid was added directly to the washed peptide solution, vortexed well and freeze-dried overnight. The spent bands were discarded and freeze-dried peptide samples were stored at -80°C until required for mass spectrometry.

2.5.4 Protein preparation for SOMAscan

Extracted protein from the plasma of 15 patients and 6 controls was prepared for SOMAscan targeted proteomic profiling (SOMAlogic). 20 ng of protein was diluted with 10µl/ml HALT protease inhibitor cocktail (ThermoFisher) in SOMA AB buffer (SOMAlogic). Samples were sent to UCL genomics, UCL Great Ormond Street Institute of Child Health, London, UK for processing.

2.5.5 Protein preparation for iTRAQ mass spectrometry

42 samples consisting of 300 µg of protein each was sent to Beijing Genomics Institute (BGI), Hong Kong, China for iTRAQ mass spectrometry. Protein was extracted by homogenisation from 21 starved (treated) and 21 fed (untreated) fibroblasts of 15 patients and 6 controls. Protein processing was completed by BGI. Before processing, samples were pooled into 14 groups of 3 (5x 3 treated patients, 2x 3 treated controls, 5x 3 untreated patients, 2x 3 untreated controls). Treated and untreated samples formed 2 distinct iTRAQ experiments. A bridge sample was included in both for normalisation and comparison across the runs.

Samples were acetone precipitated at -20 °C overnight, dried, and reconstituted in lysis buffer, consisting of 2 mM EDTA and 1mM phenylmethane sulfonyl fluoride (PMSF), a serine protease inhibitor. Samples were reduced with 10 mM DTT and alkylated with 55 mM IAA. Protein was quantified using a Bradford assay, where absorbance was measured at 595 nm. Sample concentrations were calculated from a standard curve of known BSA concentrations. Samples were checked for integrity by running 30 µg of protein on polyacrylamide gel electrophoresis. 100 µg of protein was then digested with 10 µg Trypsin Gold at 37 °C for 12 hours. Samples were dried, reconstituted in 0.5 M Triethylammonium bicarbonate buffer and iTRAQ labelled using the iTRAQ reagent 8-plex kit (SCIEX) according to manufacturer's instructions. Labelled samples were incubated for 2 hours and 8-plex tagged samples pooled into 1 run.

Peptides were cleaned and fractionated before mass spectrometry using C18 reverse-phase chromatography. Digested samples were diluted in 2ml 5% acetonitrile, pH 9.8 and loaded on a 4.6x 250 mm Gemini C18 column containing 5 µm particles (Phenomenex). Peptides were eluted with the Shimadzu LC-20AB high performance liquid chromatography (HPLC) pump system at a flow rate of 1ml/min with a gradient of 5% buffer B (95% acetonitrile, pH 9.8) for 10 minutes, 5-35% buffer B for 40 minutes and 35-95% buffer B for 1 minute. The system was maintained in

95% buffer B for 3 minutes and equilibrated with 5% buffer B for 11 minutes. Fractions were collected every minute, after which peptides were pooled as 20 fractions and vacuum-dried.

2.6 Sequencing and profiling

2.6.1 Whole genome sequencing

15 patient and 30 control DNA samples were prepared to 60µg concentrations and distributed to BGI, Hong Kong. BGI quantified and checked DNA integrity before commencing library preparation using the TruSeq DNA PCR-free Preparation Kit (Illumina), followed according to manufacturer's instructions. The samples were sequenced on the Illumina HiSeq X Ten sequencer to 30X coverage, generating 90Gb of raw data per sample. BGI returned sequenced samples as fastq files, which were processed in-house.

2.6.2 Whole exome sequencing

16 patient and 25 control DNA samples were sent to PerkinElmer Inc, Connecticut, USA for sequencing. The samples were prepared using Agilent SureSelect exome version 3 50Mb library preparation kit and sequenced to 90X coverage on the HiSeq 1000 sequencer. Perkin Elmer returned sequenced samples as fastq files, which was processed in-house.

2.6.3 RNA sequencing

RNA extracted from starved (treated) and fed (untreated) fibroblasts of 15 patients and 6 controls was sequenced at UCL Genomics, UCL Institute of Child Health, London, UK. RNA integrity and quantitation was assessed using the 4200 TapeStation Instrument (Agilent Technologies). 250ng RNA was taken for RNA-seq library preparation. Sample libraries were prepared to 43 bp using the NEBNext® UltraTM Directional RNA Library Prep for Illumina kit. These were sequenced with paired ends using the Illumina NextSeq 500 sequencer at a read depth of 16 million reads per sample. RNA-seq data was returned as fastq files for bioinformatics processing.

2.6.4 Unlabelled tandem mass spectrometry

Peptides were reconstituted in 3% acetonitrile, 0.1% TFA and centrifuged for 5 minutes at 13,000 rpm at room temperature. 20µl was transferred to quadrupole time of flight (QToF) Trueview vial (Waters) for profiling. Protein extracted from urine was profiled using the Synapt G2 Si QToF tandem mass spectrometer (Waters), whereas the plasma fractions were profiled on the Micromass matrix assisted laser desorption/ionisation (MALDI) QToF Premier tandem mass

spectrometer (Waters). Elution buffers A and B were prepared containing 95% H_2O, 5% DMSO, 0.1% formic acid and 95% acetonitrile, 5% DMSO, 0.1% formic acid, respectively. Unlabelled mass spectrometry was performed at the UCL Biological Mass Spectrometry Centre, Great Ormond Street Institute of Child Health, London, UK, assisted by Miss Jasmine Gratton.

2.6.4.1 Synapt G2 Si QToF

For urine samples, 1µl of each sample were pooled together to make a quality control sample (referred to as the QC). The QC was injected as the first and last sample, as well as evenly dispersed throughout the run. A total of 4 QCs were run and all other samples were randomised before injection.

0.4 µg of urine peptides were injected per sample onto a 20mm NanoACQUITY ultra-performance liquid chromatography (UPLC) Symmetry C18 Trap column with 5 µm beads (Waters). The samples were loaded and eluted under the conditions in **table 2.2i** into the Synapt G2 Si QToF fitted with a ZSpray source (Waters) and ion mobility separation (IMS). Data was acquired at a resolution of 20,000, capillary voltage of 2.5-3kV with IMS gas flow at 90ml/min, IMS velocity of 650 m/s and IMS wave height of 40V. MassLynx software version 4 (Waters) was used for data acquisition and transformed data for analysis.

2.6.4.2 MALDI QToF Premier

2.5µg of plasma peptides for each fraction was injected onto the 20mm NanoACQUITY UPLC Symmetry C18 Trap column with 5µm beads. Fractions were loaded and eluted under the conditions in **table 2.2ii** from the column into the QToF Premier tandem mass spectrometer fitted with a MALDI source. Data was acquired at a MALDI laser firing rate of 200, resolution of 8,500 and capillary voltage of 2.9kV. MassLynx software version 4 (Waters) was used for data acquisition and transformed data for analysis.

2.6.5 iTRAQ TripleTOF mass spectrometry

20 fractions for each iTRAQ run were resuspended in 5% acetonitrile, 0.1% formic acid and centrifuged at 20,000*g* for 10 minutes. Supernatant was loaded at a concentration of 0.5µg/µl to a LC-20AD nanoHPLC (Shimadzu) onto a 20mm C18 trap column. The samples were loaded and eluted under the conditions in **table 2.2iii**. Peptide spectra were acquired by the TripleTOF 5600 system (AB SCIEX) fitted with a Nanospray III source (AB SCIEX) and a pulled quartz tip emitter (New Objectives). Data was acquired using Analyst 1.6 software (AB SCIEX) under the following

conditions: ion spray 2.5kV, 30 psi curtain gas, 15 psi nebulizer gas, 30,000 resolution and interface heater temperature of 150°C. Raw data was converted into .mgf files for bioinformatics analysis, which was completed by BGI, Hong Kong.

i) Synapt G2 Si QToF (Waters) conditions [1]

Stage	Time (min)	Flow rate (µl/min)	% buffer A	% buffer B
Trapping	1	15	99	1
Elution	Initial	0.3	97	3
	40	0.3	60	40
	42	0.3	15	85
	45	0.3	97	3

ii) Premier MALDI QToF (Waters) conditions [1]

Stage	Time (min)	Flow rate (µl/min)	% buffer A	% buffer B
Trapping	1	10	97	3
Elution	Initial	0.3	97	3
	35	0.3	45	55
	36	0.3	1	99
	41	0.3	97	3

iii) TripleTOF 5600 (AB SCIEX) conditions [2]

Stage	Time (min)	Flow rate (µl/min)	% buffer A	% buffer B
Trapping	4	8	95	5
Elution	Initial	0.3	95	5
	41	0.3	65	35
	46	0.3	20	80
	49	0.3	95	5

Table 2.2 – Chromatography trapping (sample loading) and elution conditions before ionisation into Synapt G2 Si (i), QToF Premier (ii) and TripleTOF (iii) mass spectrometers. [1] Buffer A = 95% H2O, 5% DMSO, 0.1% Formic acid, buffer B = 95% acetonitrile, 5% DMSO, 0.1% Formic acid. [2] Buffer A = 95% H2O, 0.1% Formic acid, buffer B = 95% acetonitrile, 0.1% Formic acid.

2.7 Data processing

2.7.1 Whole genome and whole exome sequencing

Fastq files were checked for quality using fastQC (Babraham Bioinformatics), which was run using Java. Reads were then aligned to the reference genome using Burrows-Wheeler Aligner (BWA) (Li and Durbin 2009). WGS and WES reads were aligned to Genome Reference Consortium human (GRCh) genome build 38 and 37, respectively. Alignment created Sequence Alignment Map (SAM) files, which were converted into BAM files (a binary version of SAM) using SAMtools (Li et al.

2009). Picard Tools (Broad Institute) was used to mark duplicate reads in WES data, before all samples were realigned and recalibrated by Genome Analysis Tool Kit (GATK) (McKenna et al. 2010). Genotypes were called using Haplotype caller, which is built in to GATK, and further quality controls were implemented. Finally, a variant call file (vcf) for each sample was generated, which were used for variant analysis.

All WGS files were processed by Dr Federico Minneci using a pipeline devised by Dr Chela James and Miss Nital Jani (GOSgene Bioinformatics Team, UCL GOS Institute of Child Health, London). WES files were processed by Dr Chela James.

2.7.2 RNA sequencing

Fastq files were quality checked using fastQC before read alignment to GRCh38 by TopHat and Bowtie (Trapnell et al. 2009). TopHat outputted BAM files, which were checked again for quality by RSeQC (Wang et al. 2012), before read filtering to mapping quality ≥4 by SAMtools. Only genes that had more than 10 reads in 4 or more samples were retained for further analysis. From filtered BAM files, reads were counted using featureCounts (Liao et al. 2014), which were inputted into differential gene expression analysis. RNA-sequencing processing was completed by Dr Georg Otto (GOSgene Bioinformatics Team, UCL GOS Institute of Child Health, London).

2.7.3 Protein identification and quantitation of unlabelled MS data

Protein identification and quantitation was completed using Progenesis QI for proteomics (QIP) (Waters). Peptide spectra from samples within the same run were uploaded into a single Progenesis QIP experiment, where spectra were aligned to a reference spectra to generate a score >80%. Ions with a retention time <10 minutes and >48 minutes were removed. Peptides with a charge >10 were removed. Peptides were identified from raw MS/MS spectra using an MS^E search, which were screened against the Human Swiss-Prot database, canonical sequences only (from January 2015, containing 20,198 sequences). Proteins with no unique peptide identifications or a peptide length <5 were omitted. Proteins were quantified by ion abundance quantification in Progenesis QIP. Samples were separated into biological groups to compared between patients and controls. Plasma fractions were processed per fraction creating 8 individual experiments, which were combined after quantitation. Statistical significance was ascertained using an ANOVA test and false discovery rates (FDR) were generated to account for multiple testing. Differences with a FDR<0.05 were considered statistically significant.

2.7.4 Protein identification and quantitation of iTRAQ data

Mascot version 2.3.02 was used to identify proteins from raw MS/MS spectra. Uniprot, *Homo sapiens* was used as the input database, containing 134,169 sequences. At least one unique peptide was required for a successful identification. Proteins were quantified using IQuant software (Wen et al. 2014). Statistical analysis for differentially expressed proteins between biological groups was also completed by BGI, Hong Kong. Protein identifications, quantitation and significantly regulated proteins were returned as .csv files for further analysis and interpretation.

2.7.5 Differential Gene Expression

Expression analysis was implemented using R version 3.3.2 and the R package, edgeR (Robinson et al. 2010). Statistical significance for differentially expressed genes was determined by fitting a negative binomial model to read counts of each gene, followed by the generalised linear model (GLM) likelihood ratio test (McCarthy et al. 2012). Differentially expressed gene tests were conducted by disease, treatment, and disease x treatment interaction for all multi-omic datasets. Disease comparisons consisted of BBS vs. control gene expression for serum fed fibroblasts and BBS vs. control gene expression for serum starved fibroblasts. Comparison by treatment involved comparison of gene expression for starved vs. fed BBS fibroblasts and starved vs. fed control fibroblasts. A covariate 'sex' adjustment was used to correct for gene/protein expression differences caused by sex. Differential expression ratio and log2 fold change (FC) for each comparison was calculated. Differential expression analysis for RNA-seq was carried out by Dr Georg Otto (GOSgene Bioinformatics Team, UCL GOS Institute of Child Health, London). Genes were considered significantly differentially expressed with an FDR<0.05.

2.7.6 Pathway analysis

Differentially expressed genes from RNA, as well as proteins from UPLC-MS/MS and iTRAQ-MS were further analysed using Protein Analysis Through Evolutionary Relationships (PANTHER) database, version 11 (Mi et al. 2017). Statistical overrepresentation analysis and statistical enrichment test on test sets were completed using multiple comparison Bonferroni correction, with an adjusted p-value statistical significance threshold of <0.05. Protein-protein interaction plots were generated using STRING version 10.5 (Szklarczyk et al. 2015).

2.7.7 Gene Set Enrichment Analysis

GSEA is based on a comparison of a test dataset against a known gene set of biologically curated genes (Mootha et al. 2003). The enrichment score (ES) is calculated by the GSEA algorithm

'walking' down list and increasing the running ES sum when a test gene corresponds to the gene set, and decreases if it is not in the list. A positively enriched gene set will obtain a positive ES, whereas a negatively correlated gene set will be negative (**figure 2.1**). GSEA creates a normalised enrichment score (NES) by taking into account the original size of the expression dataset, so that GSEA can be reliably compared across different experiments (Subramanian et al. 2005). Statistical analysis is also encompassed within the GSEA framework, where false discovery rate (FDR) is calculated to control for false positive findings. The benefit of GSEA over pathway overrepresentation analysis is that GSEA uses the entire expression dataset, and does not purely rely on differentially expressed findings (Irizarry et al. 2009).

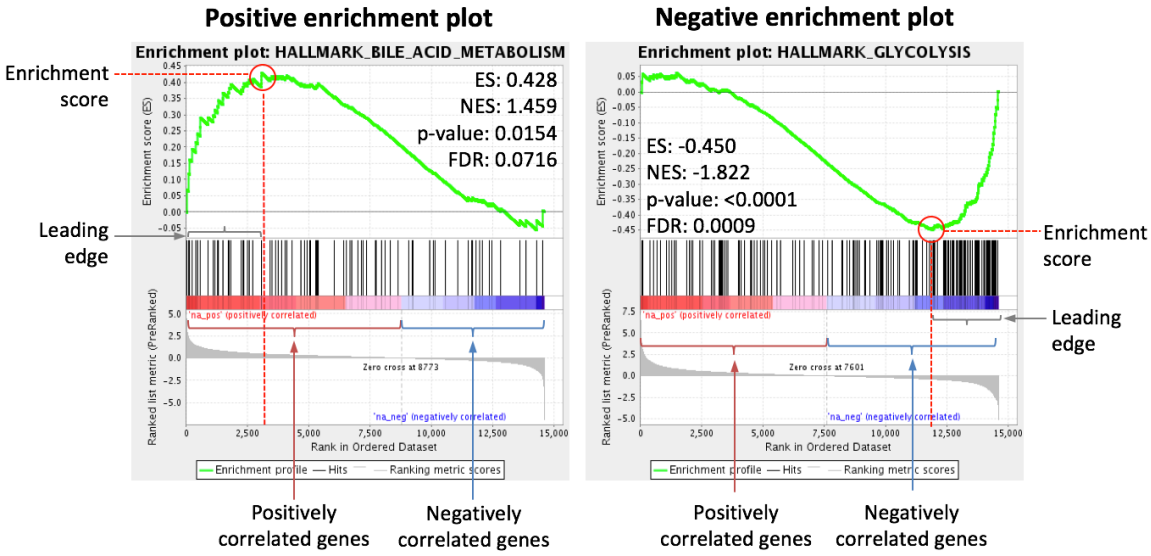

Figure 2.1 – Graphical representation of gene set enrichment analysis for positively (left) and negatively (right) enriched gene sets. The top of the plot shows the running enrichment score (ES) as the analysis progresses down the ranked dataset. The peak of the graph is the ES for the dataset, where peaks at the front of the plot are positively enriched (positive ES) and peaks at the end of the plot are negative enriched (negative ES). The horizontal black lines in the middle of the plot represents each gene analysed, and indicates whether it is positively or negatively correlated with the gene set. The leading edge is the group of genes that contribute most to the ES.

Gene set enrichment analysis (GSEA) (Subramanian et al. 2005; Mootha et al. 2003) was implemented for all comparisons with RNA-seq and iTRAQ data using the Molecular Signature Database (MSigDB) Hallmark database of 50 gene sets, version 6 (Liberzon et al. 2015). Analysis was completed using the GSEA preranked tool, and were ranked according to log2 fold change between biological groups. NES, FDR values, and GSEA profiles were generated for the top 20 positive and negative significantly enriched gene sets. Gene sets were considered to be

significantly enriched with an FDR<0.25 (as recommended by GSEA (Subramanian et al. 2005)). GSEA desktop application software version 3 was provided by the Broad Institute. Transcriptome and proteome GSEA was completed by Dr Georg Otto and Dr Xueting Wang, respectively (GOSgene Bioinformatics Team, UCL GOS Institute of Child Health, London).

2.8 Identification of modifying alleles from NGS data

2.8.1 Identification of modifying alleles

A 4-tier categorisation strategy was developed to keep genetic variants that may modify BBS disease expressivity. Variants from 15 BBS patients were filtered from WGS vcfs uploaded into Ingenuity Variant Analysis (IVA) software (QIAGEN). Calls were initially filtered to reduce the number of candidate variants. First, only variants in exonic regions were considered for analysis. Next, a genetic control filter was added that removed variants seen in 30 control genomes (20 healthy and 10 very early onset-inflammatory bowel disease (VEO-IBD) subjects). A confidence filter removed variants with call quality <20 and calls found within the top 5% most exonically variable 100base windows in healthy public genomes. Finally, a panel of genes known to cause primary ciliopathies was added. These qualifying variants were separated into 4 tiers, starting with the most stringent criteria, which gradually relaxed. At each stage of analysis, either the population frequency based on ExAC European frequency data (Tier 1&2, MAF<1%; Tier 3&4, MAF<5%) or the prediction of deleterious calls by PROVEAN or PolyPhen-2 was manipulated. To help assign functionality when PolyPhen-2 and PROVEAN were discordantly called, a CADD score >20 filter was added, where variants <20 were classified as benign. To qualify as deleterious, variants had to be called by 2 out of 3 functional prediction algorithms. The quality of variants was checked by visualising BAM files on Integrative Genome Viewer (IGV) software.

2.8.2 Burden testing

Mutational burden analysis was completed using two data sets generated from whole genome and whole exome sequencing. The subjects sequenced by WGS consisted of 14 patients (BBS011 was excluded as a monozygotic twin) and 30 unrelated European control individuals. WES cases included 16 unrelated patients, 10 of which had a definitive molecular diagnosis. The WES control group consisted of 25 unrelated individuals. All control subjects were either clinically healthy or presented a disease unconnected to any ciliopathy.

Qualifying variants were filtered using Ingenuity Variant Analysis (IVA). First, variants from WGS samples were prefiltered to only include variants in exonic regions. The confidence filter was set

as described previously. In WES subjects, variants were removed with a read depth <10. Variants with MAF<1% of total ExAC population frequency data were kept. Calls were kept if predicted pathogenic or likely pathogenic according to computed American College of Medical Genetics (ACMG) guidelines (Richards et al. 2015) or disease-associated according to the Human Gene Mutation Database (HGMD) (Stenson et al. 2017) or a CADD score >20. Variants that exhibited a frameshift, in-frame indel, start/stop codon loss, splice site loss, or missense change predicted deleterious by SIFT or PolyPhen-2 functional prediction software, were also retained. Finally, variants were kept if found within a gene list; either genes known to nonmotile ciliopathies, primary cilia genes within the SCGSv1 or a control gene panel associated with Charcot-Marie-Tooth disease (supplemental data, table A.1, table A.2, table A.3).

All qualifying variants were exported to Microsoft Excel for analysis. The number of alleles was counted within patient and control groups, where heterozygous and homozygous variants were assigned a count of 1 and 2, respectively. The number of qualifying variants were calculated for WGS and WES separately and then combined. A separate analysis was performed where primary driving mutations (PDMs) were removed from each patient with a definitive molecular diagnosis (WGS = 28 PDM alleles, WES = 20 PDM alleles). The mean number of alleles per individual was calculated for each experiment and compared between patient and control groups using a Fisher's Exact test (GraphPad Software). Tests were considered to be statistically significant with a p-value<0.05.

The number of WGS variants between patient and control groups was assessed for technical abnormalities. The number of variants was counted for each patient or control individual at each stage of filtering in IVA and the mean calculated in Microsoft Excel. Statistical differences between patient and control means were obtained by Student's T test (GraphPad Software). Tests were considered to be statistically significant with a p-value<0.05.

2.9 Induced pluripotent stem cell culture

2.9.1 Reprogramming of fibroblasts into induced pluripotent stem cells

Patient skin biopsies were sent to the Wellcome Trust Sanger Institute to be reprogrammed into induced pluripotent stem cells (iPSCs) by the Human induced pluripotent stem cell initiative (HipSci) project. Fibroblasts were cultured from the biopsy by the same method as described previously. Fibroblasts were transduced using Cytotune 2.0 Sendai virus (ThermoFisher) containing the Yamanaka cocktail of *OCT4*, *SOX2*, *c-MYC* and *KLF4* (Takahashi and Yamanaka

2006) and maintained until iPSC colonies appeared. iPSC colonies were picked and grown on a feeder-free matrix of Vitronectin XF for 3-4 weeks before shipping.

iPSC samples underwent rigorous quality control checks, including regular tests for mycoplasma contamination. iPSC pellets were assayed for chromosomal or unexpected genetic abnormalities and required a high Pluritest score (Müller et al. 2011) before banking.

2.9.2 Maintenance of iPSC colonies

iPSC colonies were maintained on a Vitronectin XF (StemCell technologies) matrix and Essential 8 (E8) medium (Life technologies). Vitronectin plates were prepared by diluting 40µl Vitronectin XF in 960µl DPBS for each well of a 6-well plate. Plates were stored at 4°C for up to 3 days and incubated at room temperature for 1 hour before use. Before cells were added to vitronectin coated wells, the vitronectin/DPBS mix was aspirated and air dried for up to 1 minute.

E8 media was prepared by diluting 2ml E8 supplement in 98ml filtered E8 basal media, which was used within 10 days of preparation. E8 media was stored at 4°C and equilibrated to room temperature for 1 hour before use.

2.9.2.1 Thawing iPSC colonies

To increase cell recovery during thawing, 10µM Y-27632 ROCK inhibitor (Sigma-Aldrich) in E8 at a concentration of 1:1000 was used. A cryovial of iPSC colonies was thawed at 37°C and immediately added dropwise to the E8/Y-27632 mix, careful to not disrupt the colonies. The cells were centrifuged at 200*g* for 3 minutes. The supernatant was discarded and pellet gently resuspended in 2ml of E8/Y-27632. The 2ml suspension was added to a vitronectin treated well of a 6-well plate and incubated at 37°C, 5% CO_2. Media was changed daily with fresh E8 media and passaged every 5-7 days once colonies were ~80% confluent.

2.9.2.2 Passaging iPSC colonies

iPSC colonies were dissociated using 0.5mM EDTA by diluting UltraPure 0.5M EDTA (Life Technologies) with DPBS at a 1:1000 dilution. Spent medium was aspirated and cells gently washed with 2ml DPBS. 1ml of 0.5mM EDTA was added to the colonies and incubated at room temperature for 4 minutes. EDTA was aspirated and colonies were gently collected in a falcon with 4x 1ml E8 media without resuspension by pipetting. Colonies were left to settle for 2-3 minutes and supernatant carefully aspirated. The pellet was resuspended in E8 media, ensuring

distribution of equally sized colonies and was typically split at a ratio of 1:6 in a new vitronectin coated plate. Plates were kept at 37°C, 5% CO_2. Colonies were carefully monitored and media changed daily.

2.9.2.3 Freezing iPSC colonies

iPSC colonies from 1 well of a 6-well plate were frozen in 1ml freezing media consisting of 90% knock-out serum replacement (KSR) (ThermoFisher) and 10% DMSO (Sigma-Aldrich). Both freezing media and Mr. Frostie freezing container filled with isopropanol were kept at 4°C until use. After detachment by EDTA and centrifugation, the iPSC pellet was resuspended in freezing media and transferred to a cryovial. These were transferred to a pre-chilled Mr. Frosty (kept at 4°C) for storage at -80°C for 24 hours, before being moved to LN_2 for long term storage.

2.9.3 Differentiation into 3D optic cups

2.9.3.1 Experimental outline

The experimental design for differentiation was developed to maximise the number of differentiations for each line. This was because differentiation of separate cell lines, and also within differentiation attempts for the same line, were sometimes inconsistent. To counteract this, a new differentiation was initiated upon every passaging event, thereby increasing the chance that the differentiation was successful (figure 2.2).

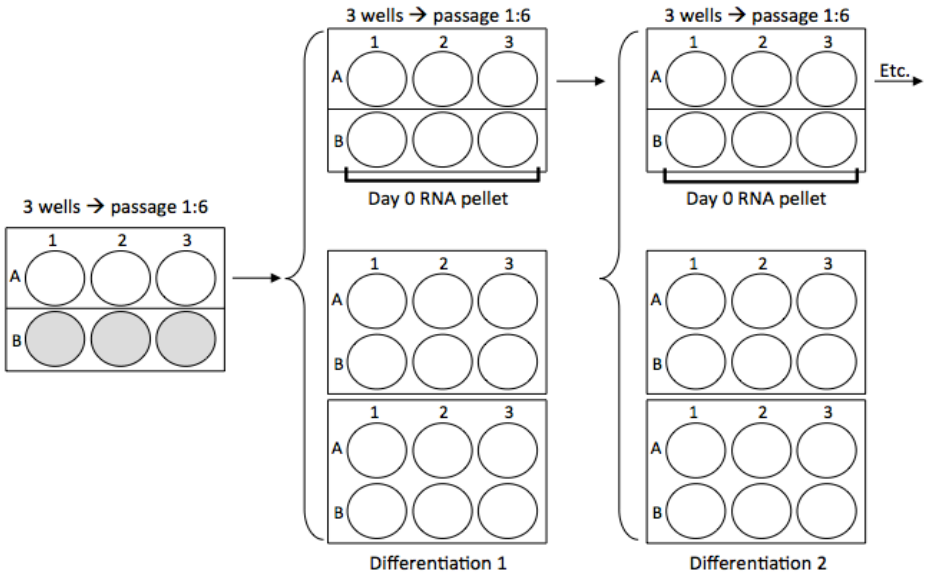

Figure 2.2 – An outline of the design adopted for maintenance and differentiation of iPSCs. This shows 3 6-well plates were maintained between every passage. When ~80% confluent, differentiation was initiated for 2 plates and the other plate was passaged or harvested for RNA extraction.

This also provided the opportunity to assess for variability between each differentiation within a single line. 5-6 differentiations were initiated for each line. The strategy also took into consideration regular sampling, which was required to track the process of differentiation. Each differentiation started with 2 confluent 6-well plates.

2.9.3.2 Embryoid body (EB) formation

Differentiation was instigated on 2x 6-well plates of iPSCs with good pluripotent morphology and low rates of spontaneous differentiation. The first day of differentiation was referred to as day 0. For this study, EBs were generated using a static suspension method.

Embryoid body media	Day 0 - 4	
Reagent	**Volume**	**Manufacturer**
DMEM/F12 (1:1) with L-glutamine	80ml	ThermoFisher
Knock out serum replacement	20ml	ThermoFisher
MEM non-essential amino acids	1ml	ThermoFisher
50 mM 2-mercaptoethanol	0.7µl	ThermoFisher
10 µM Y-27632 ROCK inhibitor on day 0	1:1000	Sigma-Aldrich

Table 2.3 – The recipe for embryoid body media.

Embryoid body media (EBM) was prepared by filter sterilising the components in **table 2.3**. Colonies were dissociated with EDTA as described previously. After centrifugation at 200*g* for 1 minute, the supernatant was removed and the total pellet from 12 wells was resuspended in 24ml EBM and 12ml each was transferred into 2x T25 flasks. Flasks were incubated at 37°C, 5% CO_2 and EB media changed daily until day 4.

2.9.3.3 Changing to neural induction media and laminin plating

On day 4 of differentiation, media was changed from EBM to neural induction media (NIM). NIM was prepared by filter sterilising all components in **table 2.4**, except N2 supplement, which was added after filtering. Media was changed daily.

On day 6 of differentiation, EBs were transferred to laminin coated 6-well plates. These were prepared on day 5 by dilution of 40µl laminin (Sigma-Aldrich, stock concentration 1mg/ml) in 1160µl DMEM (ThermoFisher). 200µl of laminin/DMEM was spread in the centre of each well of 2x 6-well plates and incubated overnight at 37°C.

Neural Induction Media (NIM)	Day 4 - 16	
Reagent	Volume	Manufacturer
DMEM/F12 (1:1) with L-glutamine	100ml	ThermoFisher
MEM non-essential amino acids	1ml	ThermoFisher
N2 Supplement	1ml	ThermoFisher
Heparin	100µl (2mg/ml stock)	Sigma-Aldrich

Table 2.4 – The recipe for neural induction media.

On day 6, the laminin was aspirated from each well and replaced with 2ml NIM. EBs from 2x T25 flasks were collected in 1x 15ml falcon and left to settle to the bottom of the falcon. Spent media was carefully aspirated, retaining the last ~3ml of media containing the EBs. 200µl of resuspended EBs were evenly distributed to each well, giving ~70-100 EBs per well. Plates were incubated for 48 hours at 37°C to allow EBs to adhere to laminin, and media was changed every 2 days thereafter.

2.9.3.4 Changing to retinal differentiation media

Retinal differentiation media (RDM) was prepared by filter sterilising all components in table 2.5 except B27, which was added after sterilisation. 3ml of RDM replaced NIM at day 16 of differentiation, where spent media was replenished every other day.

Retinal Differentiation Media (RDM)	Day 16 - 42	
Reagent	Volume	Manufacturer
DMEM, high glucose, L-glutamine	70ml	ThermoFisher
F12 Nutrient mixture (Ham)	30ml	ThermoFisher
B27 (without vitamin A)	2ml	ThermoFisher
MEM non-essential amino acids	1ml	ThermoFisher

Table 2.5 –The recipe for retinal differentiation media.

2.9.3.5 Mechanical lifting of neural retina domains

On day 28 of differentiation, identifiable neural retina (NR) domains were manually lifted to grow in suspension. Spent RDM was aspirated for 1 well, washed with 2ml DPBS and replaced with 1ml RDM. Lifting was aided by the in-hood microscope EVOS XL core cell imaging system (ThermoFisher). NR domains were identifiable by their distinctive horseshoe shape. These were lifted using a 10µl pipette tip and collected in a 15ml falcon. Care was taken to ensure that NR domains were undamaged and few other cells were lifted. This process of lifting domains in 1ml

RDM was completed up to three times for each well. Collected NR domains were left to settle and supernatant was removed. NR domains from 2 plates were evenly distributed across 2x fresh 6-well plates to grow in suspension.

Plates were regularly examined over the differentiation period, where cultures with poor morphology were removed. It was observed that over time a layer of cells would grow on the surface of the plate, where some optic cups would also attach. Therefore, optic cups were collected and transferred to fresh plates every 10 days, and any adherent cups would be manually lifted.

2.9.3.6 Addition of exogenous factors to retinal differentiation media

At day 42, FBS was added to RDM to promote cell viability. RDM + FBS was prepared by filter sterilising all components in table 2.6 except B27, which was added after sterilisation.

Retinal Differentiation Media + FBS (RDM+FBS)	Day 42 - 70	
Reagent	Volume	Manufacturer
DMEM, high glucose, L-glutamine	65ml	ThermoFisher
F12 Nutrient mixture (Ham)	25ml	ThermoFisher
B27 (without vitamin A)	2ml	ThermoFisher
Foetal bovine serum (FBS)	10ml	ThermoFisher
Taurine (100µM)	1ml	Sigma-Aldrich

Table 2.6 – The recipe for retinal differentiation media with extra supplementation from FBS and taurine.

At day 70, 1µM retinoic acid (RA) was added to RDM + FBS to encourage photoreceptor maturation. 1µl of 10mM RA stock was freshly added to each 10ml of RDM + FBS media. At day 98, the concentration of RA was lowered to 0.5µM, thus 0.5µl of RA stock was added to every 10ml media.

2.9.3.7 Sample collection

Where possible, samples from each differentiation were collected at day 0, 6, 16, 37, 60, 98, 147 and 198. When samples were collected, it was aimed to take a representative sample of the whole differentiation. The optic cups that had the best morphology were retained for long term culture.

2.10 Polymerase Chain Reaction

2.10.1 Reverse Transcription from RNA to cDNA

cDNA was generated by the SuperScript II Reverse Transcriptase (RT) kit (ThermoFisher). 500µg of RNA was incubated at 60°C for 5 minutes with 1µl of Oligo(dT)$_{12-18}$ (ThermoFisher), 1µl dNTP mix (10mM each) (ThermoFisher) and RNase free water to make a total of 12 µl. 4µl 5x First-strand buffer, 2µl 0.1 M DTT and 1µl RNaseOUT (40 units/µl) (ThermoFisher) were added, gently mixed and incubated for 2 minutes at 42°C. 1µl of SuperScript II RT was mixed and incubated for 50 minutes at 42°C, before a final incubation at 70°C for 15 minutes to inactive the reaction. To test for DNA contamination, negative controls for each sample were included where RT was omitted. cDNA samples were stored at -20°C until required.

2.10.2 Designing primers

Primers used for products run on agarose gel were design in Primer3Plus (Untergasser et al. 2007) and verified for specificity using University of California Santa Cruz (UCSC) genomes in-silico PCR (UCSC Genome Browser 2017). Primers binding to cDNA were designed to span exon-exon junctions where possible. Primer pairs were synthesised by Sigma-Aldrich and tested on positive and negative controls before experiments were carried out.

2.10.3 Polymerase chain reaction

For all polymerase chain reactions (PCRs) the GoTaq G2 Flexi kit (Promega) was used. In each PCR tube or well of a 96-well plate, the following were mixed: 17.75µl nuclease free water, 6µl GoTaq buffer, 2.4µl 25 mM magnesium chloride, 0.6µl dNTPs (10 mM each), 0.3µl of each primer (100µM), 5% DMSO, 0.15µl GoTaq polymerase and 1µl of gDNA or cDNA (25ng). PCR mixes were placed in the T100 Thermo cycler (Bio-Rad) where most experiments were run using the following touchdown programme: first, samples were denatured at 95°C for 3 minutes. This was followed by 10 amplification cycles, where samples fluctuated between a 30 second denaturing step at 95°C and a 30 second annealing step, which started at 64°C and decreased by 1°C with every cycle down to 54°C. Samples were incubated for 1 minute at 72°C, before the elongation cycle began, featuring 24 cycles between 95°C and 54°C. Finally, samples were incubated at 72°C for 8 minutes for the final extension. For primers specific to *PAX6*, a touchdown programme cycling from 60°C to 50°C was used.

For each set of primers (**table 2.7**) for a PCR experiment, a negative control using RNase free water instead of DNA was prepared. For RT-PCR experiments using cDNA, primers designed to

target *GAPDH* were used as a positive control to test that cDNA production was successful and gDNA contamination was absent.

2.10.4 Primers

Gene	Primer sequence	Product size (bp)		Gene Purpose
		cDNA	gDNA	
GAPDH F	ATGTTCGTCATGGGTGTGAA	89	181	Housekeeping gene
GAPDH R	GGTGCTAAGCAGTTGGTGGT			
BBS1 ex.12 F	ATCCCCTGTCTTGCTTTCCT	NA	225	Disease gene
BBS1 ex.12 R	TGCTTCATTTCCACCTCCTC			
VSX2 F	ATTCAACGAAGCCCACTACCCAGA	229	14542	Neural retina marker
VSX2 R	ATCCTTGGCTGACTTGAGGATGGA			
CRX F	TATTCTGTCAACGCCTTGGCCCTA	253	4877	Photoreceptor specific
CRX R	TGCATTTAGCCCTCCGGTTCTTGA			
SOX2 F	CCCCCGGCGGCAATAGCA	448	448	Neural induction marker
SOX2 R	TCGGCGCCGGGGAGATACAT			
OCT4 F	GTGACAGAGACAGGGGGAAA	196	196	Pluripotency marker
OCT4 R	AGCGATCAAGCAGCGACTAT			
PAX6 F	CGGAGTGAATCAGCTCGGTG	259 (-5a), 301 (+5a)	1186	Eye field marker
PAX6 R	CCGCTTATACTGGGCTATTTTGC			
NANOG F	CAAAGGCAAACAACCCACTT	158	158	Pluripotency marker
NANOG R	TCTGCTGGAGGCTGAGGTAT			
RX F	GAATCTCGAAATCTCAGCCC	279	279	Eye field marker
RX R	CTTCACTAATTTGCTCAGGAC			
LHX2 F	CAAGATCTCGGACCGCTACT	284	1242	Eye field marker
LHX2 R	CCGTGGTCAGCATCTTGTTA			
MITF F	CTCGAAAACCCCACCAAGTA	123	123	Retinal pigmented epithelium (RPE) marker
MITF R	CTGGTTTGGACATGGCAAG			
RCVRN F	TCCAGCTTCTGGTTGGTCTT	328	328	Photoreceptor specific
RCVRN R	CTGCCCACTCTTCCTCACTC			

Table 2.7 – Forward (F) and reverse (R) primer sequences including the expected product size from a cDNA or gDNA template.

2.10.5 2% agarose gel electrophoresis

A 2% gel was made by dissolving 4g of UltraPure Agarose (Invitrogen) with 200ml 1x Tris/Borate/EDTA (TBE) buffer (BioRad) in a microwave. Once cooled to ~50°C, 4µl Ethidium Bromide (Sigma-Aldrich) was added as a fluorescent tag and mixed well. The solution was poured

into a casting tray with combs. Once the gel was set, the comb was removed to reveal wells and transferred to the electrophoresis tank containing 1x TBE buffer. 5µl of cDNA PCR product was mixed with 5µl 5x Green GoTaq Flexi buffer (Promega) and 10µl of each sample was loaded to each well on the gel. 4µl of HyperLadder 50bp (Bioline) DNA molecular weight marker was added to the first well as a marker. The gel was run at 110 volts for 45 minutes. Bands were visualised using the UVP BioDoc-It Imaging system (Analytik Jena).

2.10.6 Quantitative PCR

For quantitative PCR (qPCR), TaqMan gene expression primer and probes were utilised (Applied Biosystems) and experiments were performed in triplicate. All TaqMan gene expression assays contained primers spanning exon-exon junctions and probes conjugated with 5' FAM reporter and 3' MGB-NFQ quencher (**table 2.8**). TaqMan master mix contained AmpliTaq Gold DNA polymerase, Uracil-DNA glycosylase (UDG), ROX passive reference and optimised buffer components. In each well of a MicroAmp Optical 96-well reaction plate (Applied Biosciences), 1µl of 20x TaqMan gene expression assay, 10µl 2x TaqMan gene expression master mix, 4µl cDNA template (total of 100 ng) and 5µl of RNase free water were combined. The plate was sealed before being placed in the StepOnePlus qPCR system (Applied Biosystems) and run on a standard qPCR cycle.

The thermal cycling conditions were as follows: samples were incubated for 2 minutes at 50°C to allow UDG activity, then for 10 minutes at 95 °C to quench UDG activity and activate AmpliTaq Gold. This was followed by thermo cycling between a 15 second denaturing step at 95°C and a 1 minute annealing/ extension step at 60°C. If the target gene was present, the DNA polymerase cleaved the reporter from the TaqMan probe, which emitted a fluorescent signal. Baseline and threshold for amplification curves were calculated automatically by the sequence detection system in the StepOne software and checked manually. Samples were quantified using the comparative $\Delta\Delta C_t$ quantitation method, which was implemented using Microsoft Excel. Quantitations were first normalised to an endogenous control, namely *SNRPD3* or *ACTB*, to generate the ΔC_t. The $\Delta\Delta C_t$ of each sample was calculated by normalisation to a calibrator sample, usually the sample from the earliest day of differentiation in the experiment. Relative fold change was determined through calculation of $2^{(-\Delta\Delta Ct)}$. For housekeeping gene test, samples were not normalised to an endogenous control but to day 0 of the iPSC differentiation.

Gene	Reporter/Quencher	Gene Purpose	Amplicon length	Assay ID
SNRPD3	FAM/MGB-NFQ	Reference gene	68	Hs00188207_m1
LDHA	FAM/MGB-NFQ	Reference gene	66	Hs01378790_g1
TBP	FAM/MGB-NFQ	Reference gene	91	Hs00427620_m1
PGK1	FAM/MGB-NFQ	Reference gene	73	Hs00943178_g1
GAPDH	FAM/MGB-NFQ	Reference gene	93	Hs02758991_g1
ACTB	FAM/MGB-NFQ	Reference gene	63	Hs01060665_g1
HPRT1	FAM/MGB-NFQ	Reference gene	82	Hs02800695_m1
NANOG	FAM/MGB-NFQ	Pluripotency marker	101	Hs04399610_g1
PAX6	FAM/MGB-NFQ	Eye field marker	76	Hs00240871_m1
RAX	FAM/MGB-NFQ	Eye field marker	106	Hs00429459_m1
MITF	FAM/MGB-NFQ	RPE marker	81	Hs01117294_m1
VSX2	FAM/MGB-NFQ	Neural retina marker	125	Hs01584047_m1
CRX	FAM/MGB-NFQ	Photoreceptor specific	73	Hs00230899_m1
RCVRN	FAM/MGB-NFQ	Photoreceptor specific	70	Hs00610056_m1

Table 2.8 – List of TaqMan probes and assay IDs used for qPCR.

2.10.7 Sanger sequencing of p.M390R

Primers were designed at least 50bp either side of the codon for amino acid position 390 in *BBS1* exon 12. Genomic DNA extracted from 5 iPSC samples was inputted into a touchdown PCR with *BBS1* primers (described previously). The primers were assessed for specificity by running the PCR product on a 2% agarose gel (described previously). Excess primers and nucleotides were removed by addition of 2µl ExoSap-IT PCR product cleanup reagent (Affymatrix) to 5µl of PCR product. Samples were incubated at 37°C for 15 minutes followed by 80°C for 15 minutes. Samples prepared to a concentration of 3ng/µl were sent with *BBS1* primers at 3.2pmol/µl to Source Bioscience, Nottingham, UK for Sanger sequencing. Sanger sequencing traces were visualised using Sequencher v.5.2.3 software (Gene Codes Corporation).

2.11 Immunofluorescence and imaging

2.11.1 Collection and cryosectioning optic cups at day 98, 147 and 198

Where possible, 5-10 optic cups were collected when differentiations reached day 98, 147 and 198. These were washed in PBS and fixed in 4% Paraformaldhyde (PFA) (VWR) for 20 minutes at room temperature. After 3x 10 minute PBS washes, the optic cups were cryoprotected by incubation in 30% sucrose solution at room temperature for 1 hour. The cups were then embedded in optimal cutting temperature (OCT) compound (VWR) and fast frozen in a methylbutane/dry ice slurry. They were stored at -80°C until cryosectioning and immuno-

histochemical staining. Embedded optic cups were equilibrated to -20°C before cryosectioning to a thickness of 5 µm with the Leica Biosystems cryostat.

2.11.2 Fixing of cultured cells

Fibroblasts and iPSC colonies for immunocytochemical staining were grown in 8-well Nunc Lab-Tek II chamber slide systems (ThermoFisher). Cells were washed in PBS and fixed in 4% PFA for 15 minutes at room temperature. Cells were washed 3 times in PBS for 5 minutes each before immunofluorescent staining.

2.11.3 Immunofluorescence staining

Slides of cryosectioned optic cups were dried at room temperature overnight and washed 3 times in DPBS for 5 minutes prior to staining. Samples were permeabilised by incubation in 0.1% Triton X-100 (Sigma-Aldrich), 0.1% Tween-20 (ThermoFisher) in PBS for 10 minutes at room temperature. After washing with 3 times with PBS for 5 minutes each, samples were blocked in 3% BSA (Sigma-Aldrich), 10% donkey serum (Sigma-Aldrich) in PBST (PBS, 0.1% Tween) for 1 hour at room temperature. Antibody diluent was prepared with 0.3% BSA, 1% donkey serum in PBST. Primary antibodies were diluted as indicated and samples were incubated with primary antibodies overnight at 4°C. Samples were washed in 3 times in PBS for 5 minutes each prior to application of species-specific secondary antibodies. Either anti-IgG Alexa Fluor 594 or 488 (ThermoFisher) were used as appropriate, which were diluted 1:1000 in PBST and incubated for 1 hour at room temperature in the dark. Samples were washed in PBS and coverslips were mounted with ProLong Gold antifade mountant with DAPI (ThermoFisher).

2.11.4 Antibodies

The following antibodies were used to stain the primary cilium, basal body and anterograde intraflagellar transport machinery: anti-mouse acetylated tubulin at 1:500 (Sigma-Aldrich), anti-rabbit BBS5 at 1:100 (Proteintech) and anti-rabbit IFT88 at 1:300 (Proteintech). The following antibodies were used to establish pluripotency of iPSCs: anti-goat NANOG at a dilution of 1:100 (R&D Systems), anti-goat SOX2 at 1:200 (R&D Systems) and anti-mouse OCT-4 at 1:100 (Santa Cruz Biotechnology). The following antibodies were used to follow development in iPSC derived optic cups: anti-mouse CRX at a dilution of 1:800 (Sigma-Aldrich), anti-rabbit Recoverin at 1:800 (Merck Millipore), anti-mouse BRN3 at 1:200 (Merck Millipore), anti-goat VSX2 at 1:200 (Santa Cruz Biotechnology), anti-rabbit rhodopsin 4D2 at 1:500 (Abcam) and anti-mouse cone-arrestin 7G6 at 1:10,000 was a kind gift from Professors Mike Cheetham and Alison Hardcastle (UCL

Institute of Ophthalmology, London, UK). Donkey anti-mouse, donkey anti-rabbit and donkey anti-goat Alexa Fluor 488 or Alexa Fluor 594 secondary antibodies from ThermoFisher were utilised at a dilution of 1:1000.

2.11.5 Imaging

All fluorescence images were obtained using the Carl Zeiss Z1 imager at 20x, 40x or 63x magnification and captured with the AxioCam MRm camera. For images at 63x magnification, immersion oil (Sigma-Aldrich) was used. Images were taken using AxioVision 4.8 software and exported to Fiji image processing package where scale bars were calculated for each image. Images at x20, 40x or 63x magnification were assigned 50µm, 20µm or 20µm bars, respectively.

2.11.6 Calculating the percentage of ciliated fibroblasts

Serum starved and serum fed fibroblasts from 5 lines were immunocytochemically stained with acetylated tubulin to identify primary cilia. 6 images representative of each line were taken, resulting in 30 images for each condition. Nuclei were identifiable by DAPI counterstain, which were counted for each image using an ImageJ macro. Similarly, primary cilia number was identified using a second ImageJ macro. The mean number of ciliated cells for each condition was calculated by dividing the number of cilia by the number of nuclei per image and averaging over 30 images. ImageJ macros were written by Miss Grace Freke.

2.11.7 Transmission Electron Microscopy (TEM)

Optic cups from patient line BBS001 (aetc_1) at day 198 were fixed in 3% glutaraldehyde, 1% PFA and 0.08 M sodium cacodylate-HCl buffered to pH 7.4. Following this, the optic cups were processed for transmission electron microscopy (TEM) and imaged by Dr Peter Munro (UCL Institute of Ophthalmology, London, UK). They were washed twice in 0.1M sodium cacodylate-HCl, pH 7.4 for 5 minutes each and incubated for 2 hours in 1% aqueous osmium tetroxide. They were dehydrated in alcohol, passing through once at 50-90% alcohol, and 3 times at 100%, for 10 minutes each, before incubation overnight in equal parts araldite and propylene oxide. Optic cups were embedded in araldite resin and cured at 60 °C overnight. Sections were prepared at 0.75µM and stained with 1% toluidine blue-borax and Reynolds lead citrate. The apical layer of the optic cups was imaged by JEOL 1010 transmission electron microscope at 80 kV and captured using the Gatan Orius CCD camera.

2.12 Software and Databases

With millions of genetic variants, tens of thousands of transcripts, and thousands of proteomics data points, the greatest challenge of large omic datasets is how to reduce the data size to allow accurate and biologically meaningful inferences. This can be aided through use of different software, analytical techniques, and open source databases. The software and databases utilised within this thesis are presented in table 2.9.

2.13 Raw Data

The raw data for transcriptomic and proteomic experiments, as well as outputs for GSEA and other pathway analyses, can be found on the USB stick attached to this thesis. A table demonstrating the file names and locations can be found in the supplemental material section, table S.1.

Software, database, open source software	Supplier/ provider	Version	Website	Use in thesis	Reference
Ingenuity Variant Analysis (IVA)	QIAGEN	-	https://www.qiagenbioinformatics.com/products/ingenuity-variant-analysis/	Chapter 3	-
Progenesis QI for Proteomics (QIP)	Nonlinear Dynamics by Waters Corporation	v.2.0	http://www.nonlinear.com/progenesis/qi-for-proteomics/	Chapter 4	-
Exome aggregate consortium (ExAC)	Broad Institute	v.0.3.1	http://exac.broadinstitute.org	Chapter 3	(Lek et al. 2016)
Genome aggregation database (GnomAD)	Broad Institute	v.2	http://gnomad.broadinstitute.org	Chapter 3	(Lek et al. 2016)
Sorting Intolerant From Tolerant (SIFT)	J. Craig Venter Institute	v.1.03	http://sift.jcvi.org	Chapter 3	(Ng and Henikoff 2003)
PolyPhen-2	Harvard University	v.2.2.2	http://genetics.bwh.harvard.edu/pph2/	Chapter 3	(Adzhubei et al. 2013)
Protein Variation Effect Analyser (PROVEAN)	J. Craig Venter Institute	v.1.1.3	http://provean.jcvi.org/index.php	Chapter 3	(Choi et al. 2012; Choi and Chan 2015)
Combined Annotation-Dependent Depletion (CADD)	University of Washington	v.1.3	http://cadd.gs.washington.edu	Chapter 3	(Kircher et al. 2014)
PANTHER Classification system	Gene Ontology Phylogenic Annotation Project	v.12.0	http://pantherdb.org/	Chapter 4 & 5	(Mi et al. 2013; Mi et al. 2017)
Gene Ontology	Gene Ontology Consortium	-	http://geneontology.org/	Chapter 4	(Ashburner et al. 2000; Gene Ontology Consortium 2015)
Reactome	Reactome consortium	v.54.0	http://reactome.org	Chapter 5	(Matthews et al. 2009; Croft et al. 2011)
Gene set enrichment analysis (GSEA)	Broad Institute	v.3.0	http://software.broadinstitute.org/gsea/index.jsp	Chapter 5	(Subramanian et al. 2005; Mootha et al. 2003)
Molecular Signatures Database (MSigDB) Hallmark	Broad Institute	v.6.0	http://software.broadinstitute.org/gsea/index.jsp	Chapter 5	(Liberzon et al. 2015)
STRING	STRING consortium	v.10.5	https://string-db.org	Chapter 3, 4 & 5	(Szklarczyk et al. 2015)

Table 2.9 – Software, databases, and open source software utilised within this thesis.

2.14 Patients and controls

2.14.1 Recruitment criteria

Patients were recruited to the study from April to December 2014. All clinical data described below was relevant at the time of recruitment. The conditions for recruitment were as follows:

- patients possess a definitive molecular diagnosis comprised of the common p.M390R mutation in *BBS1*. Patients may be homozygous for p.M390R or compound heterozygous with p.M390R on one allele and an additional segregating pathogenic allele elsewhere in *BBS1*
- patients were white European from non-consanguineous families
- patients must have attended the multidisciplinary BBS clinic at Guy's Hospital, London for genetic and phenotypic assessment
- patients consented to use of a skin biopsy, whole blood, and urine for research purposes

2.14.2 Clinical details

The clinical details for each patient is outlined here and detailed in **table 2.10**.

2.14.2.1 BBS001

Gender: Female Genotype: *BBS1* p.M390R/ p.M390R Age: 30-39

Retinal pigmentation was observed at the age of 7 and rod-cone dystrophy diagnosed at 13 years. The patient had a clinically obese BMI of 35.6 when samples for this study were taken and has moderate learning difficulties. She suffered from recurrent renal and urinary tract infections (UTIs), including vesicoureteral reflux and pyelonephritis with renal abscess formation. Despite this, she has no structural renal dysfunction to classify as a primary BBS feature. Physical secondary features of BBS include mild cataract, brachydactyly, epicanthic folds and a high palate.

2.14.2.2 BBS002

Gender: Male Genotype: *BBS1* p.M390R/ p.M390R Age: 60+

Patient BBS002 has been registered blind since his 20s due to advanced rod-cone dystrophy. The patient was born with foot polydactyly and suffers from mild learning difficulties. At the time of recruitment to the study, he had a clinically obese BMI of 38.5. He had type II diabetes mellitus and hypertension, likely as a secondary feature to obesity.

2.14.2.3 BBS003

Gender: Female Genotype: *BBS1* p.M390R/ p.M390R Age: 40-49

Patient BBS003 suffers from severe rod-cone dystrophy and was born with polydactyly of the left hand. She presented with persistent fetal lobulation, a structural defect of the kidney. At the time of study recruitment, the patient had a severely obese BMI of 41 and also displayed severe learning difficulties. Secondary development defects include brachydactyly, syndactyly and a high arched palate. She also had hypothyroidism and hypertension.

2.14.2.4 BBS004

Gender: Male Genotype: *BBS1* p.M390R/ c.723+1G>A Age: 20-29

Patient BBS004 has severe rod-cone dystrophy and moderate learning difficulties but displays no other symptoms indicative of BBS. He was identified as a patient due to a sibling who presents a more typical phenotype. At the time of study recruitment, BBS004 had an overweight BMI of 25.6, but was not clinically obese.

2.14.2.5 BBS005

Gender: Female Genotype: *BBS1* p.M390R/ p.M390R Age: 60+

Night blindness was first reported in this patient at the age of 8 years. Deterioration of eyesight has rendered this patient completely blind. She was borderline obese at the time of recruitment with a BMI of 30.2 and was born with polydactyly of the feet and the right hand. BBS005 also experienced renal cysts and cardiac issues including angina pectoris and hypertension. She had some urorenal abnormalities, including pyelonephritis and nephrolithiasis (kidney stones).

2.14.2.6 BBS006

Gender: Male Genotype: *BBS1* p.M390R/ p.G73* Age: 40-49

BBS-related symptoms were first diagnosed in this patient at the age of 11 and he was registered blind from rod-cone dystrophy at 21 years. 4-limb polydactyly was present at birth and has orodental abnormalities including high arched palate, hypodontia and dental crowding. Weight gain was noticeable at an early age and had a clinically obese BMI of 36.2 at recruitment. Other secondary features of BBS include cataracts, hypercholesterolemia, polyuria and polydipsia.

2.14.2.7 BBS007

Gender: Female Genotype: *BBS1* p.M390R/ p.M390R Age: 40-49

Patient BBS007 was diagnosed with BBS in her 40s due to a late onset of rod-cone dystrophy. Despite this, her vision has diminished quickly. This patient had a severely obese BMI of 51.3 at recruitment and presents stage III chronic kidney disease. Developmental features include moderate learning difficulties, postaxial foot polydactyly, midface retrusion, a high palate and dental crowding.

2.14.2.8 BBS008

Gender: Male Genotype: *BBS1* p.M390R/ p.M390R Age: 40-49

Patient BBS008 has profound rod-cone dystrophy and was registered blind in his 20s. At the time of recruitment, he was clinically obese with a BMI of 35.1. He was born with polydactyly of the feet and has mild learning difficulties. BBS008 has type I diabetes mellitus, cataracts, dental crowding and suffers from recurrent UTIs.

2.14.2.9 BBS009

Gender: Male Genotype: *BBS1* p.M390R/ p.M390R Age: 30-39

Patient BBS009 presents mild rod-cone dystrophy but shows a progression decline of colour vision. Polydactyly of the feet was present at birth. He was clinically obese with a BMI of 31.3 at the time of recruitment.

2.14.2.10 BBS010

Gender: Male Genotype: *BBS1* p.M390R/ p.Y284Sfs*5 Age: 50-59

Patient BBS010 is a monozygotic (MZ) twin with patient BBS011 and brother of BBS012. He has severe rod-cone dystrophy and is registered to have severely impaired vision. He was born with 4-limb polydactyly and presents moderate learning difficulties. Secondary BBS features include cataracts, high palate and dental crowding. At the time of recruitment, he was clinically obese with a BMI of 31.2.

2.14.2.11 BBS011

Gender: Male Genotype: *BBS1* p.M390R/ p.Y284Sfs*5 Age: 50-59

Patient BBS011 has severe rod-cone dystrophy and early cataract formation. In contrast to his MZ twin, he was not born with polydactyly. He has moderate learning difficulties and was clinically obese with a BMI of 33 at recruitment.

2.14.2.12 BBS012

Gender: Female Genotype: *BBS1* p.M390R/ p.Y284Sfs*5 Age: 50-59

Patient BBS012 is the affected sibling of BBS010 and BBS011. In comparison to her siblings, she had later onset of rod-cone dystrophy and milder learning difficulties. She was clinically obese at recruitment with a BMI of 33.4. Secondary features include cataracts, high palate, dental crowding and nephrolithiasis.

2.14.2.13 BBS013

Gender: Female Genotype: *BBS1* p.M390R/ p.M390R Age: 20-29

Patient BBS013 was diagnosed at the age of 3 years due to onset of visual decline, but unusually did not present with night blindness. BBS013 has mild learning difficulties and was born with polydactyly. Secondary BBS features include cardiac, endocrine and urorenal abnormalities.

2.14.2.14 BBS014

Gender: Female Genotype: *BBS1* p.M390R/ p.G73* Age: 30-39

Patient BBS014 was diagnosed with rod-cone dystrophy at 7 years. She was born with 4-limb polydactyly and has mild learning difficulties. BBS014 had stage V chronic kidney disease, which required transplantation. At the time of recruitment, she was reported to be obese, but her BMI was not recorded. Secondary endocrine features include uterine leiomyoma and polycystic ovaries.

2.14.2.15 BBS015

Gender: Male Genotype: *BBS1* p.M390R/ p.M390R Age: 40-49

Night blindness was first reported in BBS015 at the age of 9 years, developing into rod-cone dystrophy in later life. He presents with hypogonadism and some urorenal abnormalities, including polyuria and polydipsia. At the time of recruitment, he was reported to be overweight, with a BMI of 27.

Patient ID	Gender	Allele 1	Allele 2	Primary features							Secondary features						
				Retinal dystrophy	Obesity (BMI)	Hypo-gonadism	Postaxial Polydactyly	Renal dysfunction	Learning difficulties	Count	Eye abnormalities	Developmental features	Endocrine	Orodental abnormalities	Cardiac issues	Urorenal abnormalities	Count
BBS001	Female	*BBS1* p.M390R	*BBS1* p.M390R	Severe	Yes (35.6)	-	-	-	Moderate	3	Cataract	Brachydactyly, epicanthic folds	-	High palate	-	Vesicoureteral reflux, recurrent UTI, pyelonephritis	4
BBS002	Male	*BBS1* p.M390R	*BBS1* p.M390R	Severe	Yes (38.5)	-	Feet	-	Mild	4	Cataract	-	Type II diabetes mellitus	-	Hypertension	-	3
BBS003	Female	*BBS1* p.M390R	*BBS1* p.M390R	Severe	Yes (41.0)	-	Hand	Persistent fetal lobulation	Severe	5	-	Brachydactyly, Syndactyly	Hypothyroidism	High palate	Hypertension	-	4
BBS004	Male	*BBS1* p.M390R	*BBS1* c.723+1G>A	Severe	OW (25.6)	-	-	-	Moderate	2	-	-	-	-	-	-	0
BBS005	Female	*BBS1* p.M390R	*BBS1* p.M390R	Severe	Yes (30.2)	-	3-limb	Renal cysts	-	4	Cataract	-	-	Dental crowding	Hypertension, angina pectoris	Pyelonephritis, polyuria, nephrolithiasis	4
BBS006	Male	*BBS1* p.M390R	*BBS1* p.G73*	Severe	Yes (36.2)	Yes	4-limb	-	Mild	5	Cataract	-	Hyper-cholesterolemia	Hypodontia, high palate, dental crowding	-	Polyuria, polydipsia	4
BBS007	Female	*BBS1* p.M390R	*BBS1* p.M390R	Mild	Yes (51.3)	-	Feet	Stage III CKD	Moderate	5	Strabismus	Midface retrusion	Hypothyroidism	High palate, dental crowding	-	-	4
BBS008	Male	*BBS1* p.M390R	*BBS1* p.M390R	Severe	Yes (35.1)	-	Feet	-	Mild	4	Cataract, Strabismus	-	Type I diabetes mellitus	Dental crowding	-	Recurrent UTI	4

BBS009	Male	*BBS1* p.M390R	*BBS1* p.M390R	Mild	Yes (31.3)	-	Feet	-	-	3	-	-	-	-	-	Recurrent UTI	1
BBS010	Male	*BBS1* p.M390R	*BBS1* p.Y284Sfs*5	Severe	Yes (31.2)	-	4-limb	-	Moderate	4	Cataract	-	-	High palate, dental crowding	-	-	2
BBS011	Male	*BBS1* p.M390R	*BBS1* p.Y284Sfs*5	Severe	Yes (33)	-	-	-	Moderate	3	Cataract	-	-	-	-	-	1
BBS012	Female	*BBS1* p.M390R	*BBS1* p.Y284Sfs*5	Mild	Yes (33.4)	-	-	-	Mild	3	Cataract	-	-	High palate, dental crowding	-	Nephrolithiasis	3
BBS013	Female	*BBS1* p.M390R	*BBS1* p.M390R	Mild	Yes (37.3)	-	4-limb	-	Mild	4	-	-	Polycystic ovaries, diabetes insipidus	-	Aortic valve stenosis, Aortic regurgitation	Recurrent UTI, polysplenia	3
BBS014	Female	*BBS1* p.M390R	*BBS1* p.G73*	Severe	Yes (BMI N/k)	-	4-limb	Stage V CKD	Mild	5	-	-	Uterine leiomyoma, polycystic ovaries	-	-	Recurrent UTI	2
BBS015	Male	*BBS1* p.M390R	*BBS1* p.M390R	Severe	OW (27)	Yes	-	-	-	2	-	-	-	-	-	Polyuria, polydipsia	1

Table 2.10 – Patient clinical information. Grey highlighted headings indicate cardinal features of BBS, whereas secondary features are on the right-hand side of the table. A count for primary and secondary features is displayed for each patient. N/k indicates a feature that was not recorded; OW = overweight.

2.14.3 Control individuals

Throughout this project, biological replicates from BBS patients were compared against a set of clinically healthy non-obese control individuals (detailed in table 2.11). Urine and plasma from control samples CTRL019 and CTRL021-26 were utilised as biological replicates for biomarker discovery in chapter 4. Fibroblasts obtained from CTRL019-020 and CTRL030-033 were profiled by RNA-seq and iTRAQ mass spectrometry in chapter 5. Finally, iPSC cell lines from CTRL037 and CTRL038 were reprogrammed from healthy individuals, which were provided for use in chapter 6 by the HipSci project (Wellcome Trust Sanger Institute, Cambridge, UK).

Sample ID	Gender	Age	Ethnicity	Material/cells obtained
CTRL019	Female	40-49	European	Urine/ plasma/ fibroblasts
CTRL020	Female	30-39	European	Fibroblasts
CTRL021	Female	20-29	European	Urine/ plasma
CTRL022	Male	30-39	European	Urine/ plasma
CTRL023	Female	30-39	European	Urine/ plasma
CTRL024	Male	20-29	European	Urine/ plasma
CTRL025	Male	30-39	European	Urine/ plasma
CTRL026	Male	50-59	European	Urine/ plasma
CTRL030	Female	20-29	European	Fibroblasts
CTRL031	Female	30-39	European	Fibroblasts
CTRL032	Female	20-29	European	Fibroblasts
CTRL033	Male	30-39	European	Fibroblasts
CTRL037	Male	65-69	European	iPSCs
CTRL038	Female	40-44	European	iPSCs

Table 2.11 – Sample information about control individuals used within this study.

Chapter 3: Use of next generation sequencing data to investigate the presence of genetic modifiers in BBS

3.1. Background

A genetic modifier is defined as a gene that has the potential to alter the expression of another gene at the primary locus (Cutting 2010). In the case of disease modifiers, these can transform the expressivity or manifestation of a primary driving mutation, with little or no effect in a healthy individual (Génin et al. 2008). Modifiers can either act in close proximity to the primary gene as a *cis*-modifier, or can be modulated by a genetic variant far from the primary locus, as a *trans*-modifier (Harper et al. 2015).

3.1.1. Genetic modifiers in ciliopathies

Ciliopathies are an intriguing group of diseases. The proteins involved are closely organised in a single organelle, and yet pathogenic defects in ciliopathy related genes can cause a vast spectrum of disease. These can present as a variety of seemingly unrelated phenotypes, at varying levels of severity. The diversity observed at the phenotypic level is accredited, in part, to the genetic background of each patient, where it is proposed that disease expressivity is controlled by genetic modifiers (Beales et al. 2003; Katsanis, Lupski, et al. 2001; Zaghloul and Katsanis 2009; Badano et al. 2006; Zaki et al. 2011). Since the early genotyping of pathogenic variants in BBS proteins there has been identification of additional alleles in a second BBS locus in a small subset of families (Beales et al. 2003; Zaki et al. 2011; Katsanis, Ansley, et al. 2001). Following this phenomenon, modifiers have been described in several different ciliopathies, such as additional heterozygous *AHI1* or *NPHP6* alleles found in patients with *NPHP1* nephronophthisis, modifying the phenotype from isolated neuropathy towards a JBTS-like syndrome (Tory et al. 2007). In addition, BBS proteins have been shown to alter phenotypic expression in other ciliopathies, particularly in *CEP290* related syndromes (Zhang et al. 2014; Rachel et al. 2012).

A particularly relevant example is of three siblings of northern European descent, all of whom were homozygous for *BBS1* p.M390R alleles. However, two of the siblings were heterozygous for an additional missense allele in *BBS2*, which exhibited a more severe retinal phenotype in these two, compared to the third sibling (Beales et al. 2003). Moreover, there have been remarkable cases where siblings have displayed characteristics of distinct syndromes, such as one child

exhibiting BBS inclusive of renal dysfunction, whereas their two siblings displayed the molar tooth sign indicative of JBTS, without obesity or nephropathy (Zaki et al. 2011). These findings demonstrate the close interconnectivity of the ciliary proteome and highlight the delicate nature of these syndromes, where an additional allele could easily alter the degree of phenotypic expressivity.

Upon the discovery of these triple allelic combinations, it was hypothesised that an effect would only be seen between alleles from interacting proteins (Katsanis 2004). This hypothesis supported the relatively high frequency of variants detected in both *BBS2* and *BBS6* (Katsanis, Lupski, et al. 2001; Katsanis 2004; Beales et al. 2003), as these proteins directly interact via their respective complexes, the BBSome and chaperone complex. More recently, modifying variants have also been predicted to affect BBS phenotypic expressivity, identified in *CCDC28B* (Badano et al. 2006; Cardenas-Rodriguez et al. 2013), *TTC21B* (Davis et al. 2011) and *RPGRIP1L* (Khanna et al. 2009). Although it has been shown that *CCDC28B* directly interacts with the BBSome, *TTC21B* and *RPGRIP1L* have yet to demonstrate an experimentally determined interaction with BBSome proteins. It is not yet understood whether modifiers in any ciliary gene can exhibit a *trans*-modifying effect to any non-interacting ciliary protein (Boldt et al. 2016).

3.1.2. Identification of modifiers

Although modifiers are predicted to explain the substantial variability observed in Mendelian disorders, relatively few numbers of potential modifiers have been identified so far (Génin et al. 2008). The reason for this can be attributed to the techniques for modifier identification, which can be difficult to replicate and validate. Detection of *cis*-modifiers can be relatively straight forward compared to *trans*-modifiers, as they can be detected through linkage analysis (Wright et al. 2011). *Cis*-modifiers act within close range of the primary driving mutation, and thus segregate with the disease variant. The presence or absence of a *cis*-modifier can occasionally explain variability in disease penetrance (Azaiez et al. 2004; Shawky 2014). By contrast, *trans*-modifiers exhibit their effects in other regions of the genome and can often appear as multifactorial-like diseases (Génin et al. 2008). Association studies, and more recently, next generation sequencing data are being utilised in an attempt to identify modifier alleles (Gonzaga-Jauregui et al. 2015; Li and Leal 2008).

In contrast to identification of the primary disease locus, modifier alleles are more challenging (Boyle 2007). Firstly, the majority of modifiers in ciliopathies are reported as heterozygous

variants, and thus cannot be determined through genetic segregation. Sibling or twin investigations can provide additional insight into the function of a potential modifier, if discordant phenotypes are observed with the presence or absence of a third allele. Additionally, these studies can further assist in discriminating between variability caused by genetic or environmental contributing factors (Czyz et al. 2012).

Secondly, quantifying the severity of a modifier can prove challenging (Khanna et al. 2009), especially when it is hypothesised that they implement temporal changes, such as age of onset or developmental defects. These changes are particularly difficult to validate through functional assays, as some developmental features may express differently in model organisms, as is the case with the lack of polydactyly phenotype in BBS mouse models (Davis et al. 2007). Furthermore, prediction of phenotype is made harder by variable transcript expressivity, as discrepancies may exist for the same mutant transcript in different tissue types (Shaheen et al. 2016). This phenomenon is thought to account for some of the disparity observed for splice variants, such as the common deep intronic *CEP290* variant that causes isolated retinopathy (Parfitt et al. 2016).

A further challenge is to obtain a sufficient sample size and a comparable control group (Boyle 2007), an issue that is exacerbated in a rare disease cohort. Furthermore, potential modifying variants should not be excluded if common in the general population, as these alleles can still augment an aberrant pathway when expressed in *trans* with a disease locus (Zaghloul and Katsanis 2009). Moreover, predicting the outcome of a potential modifier presents another obstacle, as they can act in either a protective or deleterious manner (Harper et al. 2015; Kousi and Katsanis 2015).

3.1.3. Characterisation of ciliopathy modifiers

Current published BBS modifiers were identified in large patient cohorts of heterogeneous ciliopathy patients (Badano et al. 2006). Typically, patient DNA was screened for an enrichment of predicted loss of function variants in known ciliopathy genes (Khanna et al. 2009; Davis et al. 2011). After identification, potential modifiers were tested to establish functionality through large scale functional assays using either a secondary cell line known to ciliate, such as kidney derived IMCD3 cells, or zebrafish morphants (Davis et al. 2011; Wheway et al. 2015). Zebrafish models are commonly used to archetype ciliopathies as they display relevant physiological

features and introduction of human wild type mRNA can robustly rescue mutant phenotypes (Khanna et al. 2009; Ross et al. 2005; Cardenas-Rodriguez et al. 2013).

Zaghloul et al. (2010) undertook a challenging project to morphometrically screen 125 common alleles, previously considered benign or reported as modifiers in BBS (Zaghloul et al. 2010). Out of the 125 alleles screened, 14 were reported as benign, 45 as unrescuable null alleles, 15 hypomorphs, and 35 alleles predicted to exhibit a dominant negative function. This demonstrates that common alleles have the potential to be detrimental, and the importance of testing variants in a model system. Further studies have corroborated that the characteristics of the variant, not just the gene, is critical to expected phenotypic outcome. Davis et al. tested 48 previously unreported polymorphisms in *TTC21B*, a member of the IFT-B complex that can cause NPHP and JS. The effect of mutant alleles on ciliogenesis was assayed in IMCD3 cells and varied from benign, hypomorphic or null effects, likely depending on their position in the gene and proximity to conserved tetratricopeptide domains (Davis et al. 2011).

Although the majority of ciliopathy modifiers have been linked to a negative effect on phenotype presentation, an intriguing study deemed to increase the complexity of ciliopathies further by revealing a protective effect of a triple allelic mouse. Separately, mice exhibiting *BBS6*$^{-/-}$ or homozygous *Cep290* truncating mutations mimicked a severe ciliopathy phenotype. Astoundingly, researchers found that generation of a triple allelic mouse for these loci reduced retinal defects and olfactory abnormalities compared to homozygous mutants alone (Rachel et al. 2012). This study further demonstrates the intricate and unpredictable relationships present in the primary cilium.

3.1.4. Mutational burden

In-depth sequencing, either at the exome or genome level, has enabled the development of new approaches for identification of disease associated genes. One of these approaches is known as mutational burden testing, which attempts to quantify the mutational load of a disease based on the total presence of pathogenic variants in relevant genes (Guo et al. 2016). Mutational load is defined by the accumulation of deleterious mutations in the general population. A landmark study determined that mutational load would tend to be attributed to more common, less damaging variants, as very harmful mutations would quickly be eliminated from the population (Kimura et al. 1963). Recently, rare mutations have been recognised to accumulate in disease genes of certain disease populations, such as patients with Charcot-Marie-Tooth disease

(Gonzaga-Jauregui et al. 2015), schizophrenia (Girard et al. 2015) and atrioventricular septal defect (AVSD) (Alessandro et al. 2016). It is hypothesised that this increased mutational burden cumulatively acts to further destabilise the network of affected biological pathways and influence disease penetrance or expressivity (Gonzaga-Jauregui et al. 2015). Due to the interconnected relationship of proteins within the primary cilium, it is possible that ciliopathies, such as BBS, will adhere to the mutational burden model and help explain phenotypic heterogeneity observed through patient cohorts (Shaheen et al. 2016).

3.1.5. Aims of chapter

In this study, a cohort of 15 BBS patients with mutations in *BBS1* has been investigated. These patients were selected due to their similar genotypes, namely either homozygous *BBS1* p.M390R or compound heterozygous for p.M390R with a second deleterious allele elsewhere in the *BBS1* locus. Despite all presenting with pathogenic variants in the same gene, there is considerable interfamilial heterogeneity shown throughout the cohort. This heterogeneity presents as differences in onset of disease, severity, or the presence/absence of a clinical feature of BBS.

I have hypothesised that the difference in phenotypic presentation between patients is due to modifier alleles that can be uncovered from whole genome sequencing data. In particular, potentially modifying variants would be present in genes that encode proteins that interact with BBS1 in the primary cilium.

Detection of modifying alleles can be more complex than identification of disease causing variants. Association studies have yielded some success for Mendelian disorders (Cutting 2010; Génin et al. 2008), though large cohort sizes are required, which was not possible for this study. Therefore, in this chapter I describe the development of a 4-tier categorisation system based on predicted functional outcomes of mutations and their population frequencies to assess the role of the modifier genes in disease pathology. A combination of three functional prediction programmes has been used, PROVEAN (Choi and Chan 2015), PolyPhen-2 (Adzhubei et al. 2013), and Combined Annotation Dependent Depletion (CADD) scoring (Kircher et al. 2014) to predict the outcome of the sequence variants. As a result, 13 out of 15 patients were found to exhibit at least one potential modifying variant that functions in interacting networks with BBS1. Further selection criteria involved identification of known ciliopathy modifiers from the literature, such as mutations in *TTC21B* (Davis et al. 2011) and *RPGRIP1L* (Khanna et al. 2009), as well as localisation of encoding modifier proteins within the primary cilium.

The second stage of this chapter addresses whether the presence of modifying variants contribute to the overall mutational burden of BBS. To undertake mutational burden testing, the mean aggregate of qualifying variants was compared between patient and control cases for a relevant gene panel. The hypothesis that there is an enrichment of pathogenic variants in the patient group compared to the control subjects was investigated. Qualifying variants were defined as rare (MAF<1%) and predicted pathogenic by SIFT, PolyPhen-2, or CADD>20. Three gene panels were examined; the ciliopathy disease panel (n=73; appendix, table A.1), an extended set of primary cilia genes curated by the SYSCILIA consortium (SYSCILIA gold standard v1 (SCGSv1) panel n=235; appendix, table A.2) (van Dam et al. 2013), and a control gene panel associated with an unrelated neuropathy, Charcot-Marie-Tooth disease (n=74; appendix, table A.3). Additional data from whole exome sequencing of 16 BBS patients and 25 controls was included in the mutational burden analysis. These were analysed separately and then integrated with WGS data to enhance statistical power. Fishers exact testing was performed to determine the statistical significance of the burden tests. Overall, this analysis uncovered that mutational burden exists only with the presence of the pathogenic driving mutation in BBS, and there is no evidence that modifying variants contribute to the total mutational load of the disease.

3.2. Results

3.2.1. Quality control for whole genome sequencing

Prior to whole genome sequencing, DNA was assessed for quality and concentration, using a Qubit 2.0 fluorometer (Thermo Fisher Scientific) and Nanodrop-1000 spectrophotometer (Thermo Scientific). Concentration and nanodrop output featuring A_{260}/A_{280} absorbency ratios are displayed in **table 3.1**. The ratio of absorbency can help determine the purity of DNA, where a ratio of 1.8 is considered pure. A_{260}/A_{280} ratios for BBS001-BBS015 range between 1.71 to 2.03. Nonetheless, samples that deterred from a ratio of 1.8 were accepted for sequencing by Beijing Genomics Institute (BGI). The integrity of total genomic DNA was assessed through electrophoresis on a 1% agarose gel (**figure 3.1**). Sample BBS007 was flagged by BGI as slightly degraded, but sequencing was completed with the understanding that sequencing quality may be affected.

Patient no.	Concentration (ng/µl)	A_{260}/A_{280}
BBS001	108.2	1.71
BBS002	39.5	1.83
BBS003	66.1	1.85
BBS004	50.1	1.91
BBS005	142	1.79
BBS006	81.8	1.97
BBS007	53.8	1.82
BBS008	58.9	1.85
BBS009	51.4	2.03
BBS010	73.5	1.83
BBS011	63.4	1.84
BBS012	65	1.89
BBS013	61.4	1.76
BBS014	78.6	1.91
BBS015	77.8	1.95

Table 3.1 – Concentrations of BBS patient genomic DNA samples with corresponding A_{260}/A_{280} absorbency ratios.

Whole genome sequencing was also completed on 30 samples that were used as a control set. 20 samples were obtained from parents of 10 trios, where the probands were Silver-Russell Syndrome (SRS) patients. The parents used in this study were asymptomatic, nonrelated and of European ethnicity. The other 10 control samples were obtained from patients with very early onset inflammatory bowel disease (VEO-IBD), which had no overlapping symptoms with BBS, and were also nonrelated and of European ethnicity. The quality of all 30 samples was assessed by the same method previously utilised for the 15 BBS patients (completed by Miss Stefanie Dowle

and Dr Jochen Kammermeier, UCL GOS Institute of Child Health, London) and were sequenced by the same sequencing technology at Beijing Genome Institute (BGI). Similarly, Dr. Louise Ocaka (UCL GOS Institute of Child Health, London) executed whole exome sequencing quality control analysis and sequencing was completed by PerkinElmer Inc (Connecticut, USA). Batch and sequencing information for all samples are presented in the appendix (table A.4).

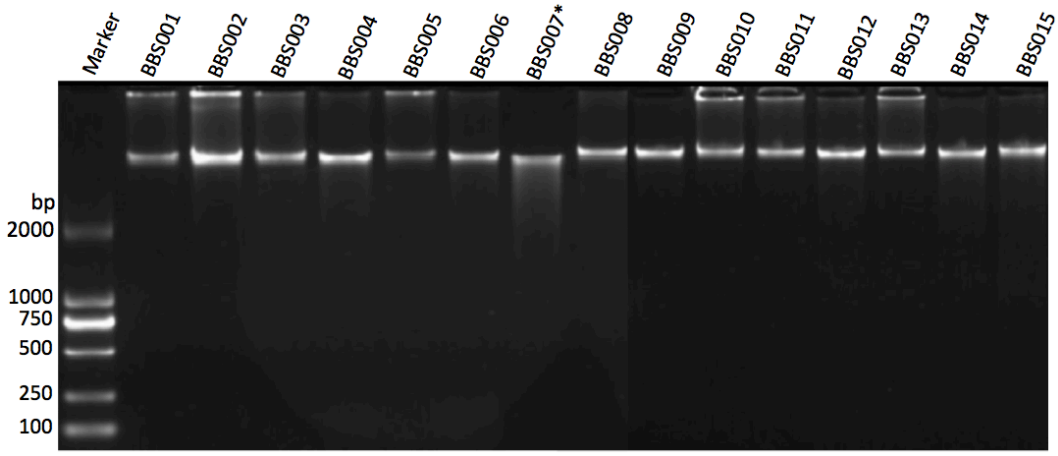

Figure 3.1 – 1% agarose gel of total genomic DNA for BBS001-015 to test DNA integrity. *BBS007 was reported by BGI as slightly degraded.

3.2.2. Identification of modifiers

The hypothesis that phenotypic heterogeneity in *BBS1* p.M390R patients is accounted for by genetic modifiers was initially investigated. Possible modifiers were identified by looking at whole genome data generated from BBS001-BBS015 in Ingenuity Variant Analysis (IVA). In total, there were 11,590,993 variants detected across all 15 genomes. Due to the unmanageable number of variants observed, several filtering methods were implemented to reduce the variant count (table 3.2).

Filtering criteria	Number of variants	Number of genes
Total variants	11,590,993	23,904
Exonic only	85,351	16,640
Variants not seen in controls	11,843	6,736
Confidence	11,672	6,710
Ciliopathy panel	89	41

Table 3.2 – Filtering criteria and resulting number of calls via IVA analysis. Confidence filter removes calls with low quality (call quality <20) and those within the most exonically variable regions in the genome.

A factor that markedly reduced the number of variants was the exclusion of variants in non-protein coding regions. Although these may impact disease pathogenicity, these were excluded from this study as their functionality is challenging to predict. Next, to reduce calls further, 30 control samples were included in the analysis, where variants were excluded if seen in any control sample. A confidence filter reduced the number of calls further by removing low quality variants. In order to assess genes that may affect ciliary function, a list containing 73 genes known to cause the spectrum of ciliopathies were screened (appendix, **table A.1**).

3.2.2.1. Development of the 4-tier categorisation system

Further classification of variants was required to aid the discovery of potential disease modifiers. To efficiently manage the variants obtained from IVA, I developed a 4-tier categorisation system, which featured relaxed and stringent tiers for extensive identification. This classification system aimed to encompass potential modifiers of varying characteristics without producing an unmanageable number of variants for further investigation. The first tier contained the most stringently called variants and at each stage the filters were relaxed, allowing more variants to be called (**figure 3.2**).

At each stage of analysis, either the population frequency based on ExAC European frequency data (Tier 1 and 2, MAF<1%; Tier 3 and 4, MAF<5%) or the prediction of deleterious calls by PROVEAN or PolyPhen-2 was manipulated. To help assign functionality when PolyPhen-2 and PROVEAN were discordantly called, a CADD score >20 filter was added, where variants <20 were classified as innocuous (Kircher et al. 2014). To qualify as deleterious, variants had to be called by 2 out of 3 functional prediction algorithms.

Tier 1 analysis identified 12 rare nonsynonymous variants (MAF<1%) in 8 genes that were predicted deleterious or were high impact variants such as frame shift, nonsense or splice site loss. Included in this group of variants were the underlying pathogenic variants in *BBS1* that were previously diagnosed (**highlighted in table 3.3i, tier 1**). Of note, *BBS1* p.M390R has a relatively high allele frequency for a rare disease-causing variant, with a MAF of 0.3% from a heterozygous allele count of 347/126644 from the Gnomad European dataset. *BBS1* p.M390R was predicted as benign by PolyPhen-2, which demonstrates the importance of using multiple functional prediction algorithms.

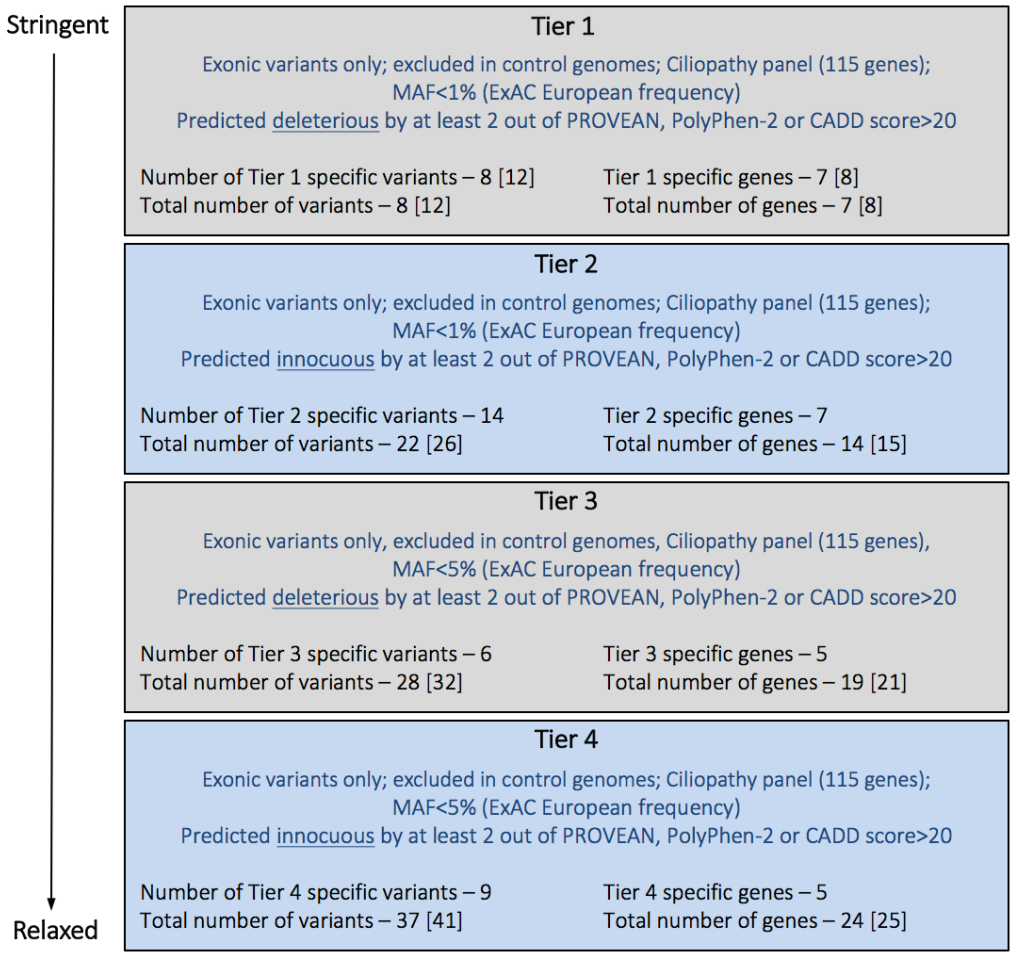

Figure 3.2 – Strategy for genetic modifiers 4-tier categorisation where deleterious prediction and allele frequency were manipulated. The number of detected variants and genes are displayed at each tier. The numbers in square brackets are inclusive of primary driving mutations in *BBS1*.

Next, Tier 2 analysis generated rare nonsynonymous variants that were predicted to be innocuous by at least 2 out of PROVEAN, PolyPhen-2 or CADD (score<20) (**table 3.3ii**). These were included in the analysis because modifiers may not be disease causing by themselves, but in combination with a pathogenic background could alter the disease expressivity (Cutting 2010). Here, 14 heterozygous missense variants in 9 different genes were identified.

For tier 3 and 4 (**table 3.3iii; 3.3iv**), the objective was to widen the net to detect variants that were less rare (MAF<5%), as some known modifiers of ciliopathies are described as common (Khanna et al. 2009; Zaghloul et al. 2010).

Table 3.3(i) Tier 1: <1% ExAC European frequency (MAF<0.01), predicted damaging by 2 or more functional prediction algorithms, where † denotes primary driving mutation

Gene Symbol	Variant	Affected patients	Translation Impact	PolyPhen-2	PROVEAN	CADD score	ExAC European Frequency (%)	Gnomad European Frequency (%)	Association of gene with disease
NPHP4	p.R735W	BBS005	Missense	Possibly Damaging	Deleterious	32	0.026	0.026	NPHP (OMIM: 606966), SLS (OMIM: 606996)
SDCCAG8	p.R82C	BBS013	Missense	Possibly Damaging	Deleterious	23.8	0.008	0.011	BBS (OMIM: 615993); SLS (OMIM: 613615)
C2ORF71	p.S1085L	BBS007	Missense	Possibly Damaging	Deleterious	17.67	0.004	0.006	RP (OMIM: 613428)
TTC21B	p.L1002V[1]	BBS006; BBS008	Missense	Possibly Damaging	Deleterious	25.2	0.800	0.874	NPHP (OMIM: 613820); JATD (OMIM: 613819)
TTC21B	p.R616C	BBS006	Missense	Benign	Deleterious	20.7	0.304	0.289	
ARL6	p.T32M	BBS005	Missense	Probably Damaging	Deleterious	33	0.000	0.000	BBS (OMIM: 600161)
BBS1†	p.G73*	BBS006; BBS014	Stop Gain			36	0.000	0.000	BBS (OMIM: 209900)
BBS1†	c.723+1G>A	BBS004	Splice Site Loss			27	0.000	0.000	
BBS1†	p.Y284fs*5	BBS010; BBS011; BBS012	Frameshift			34	0.001	0.002	
BBS1†	p.M390R	All (BBS001, 002, 003, 005, 007, 008, 009, 013, 015 as homozygous)	Missense	Benign	Deleterious	28.2	0.227	0.274	
CEP164	p.Q1410*	BBS014	Stop Gain			40	0.109	0.135	NPHP (OMIM: 614845)
XPNPEP3	p.P493A	BBS012	Missense	Probably Damaging	Deleterious	24	0.169	0.174	NPHP-like (OMIM: 613159)

Table 3.3(ii) Tier 2: <1% ExAC European frequency (MAF<0.01), predicted innocuous by 2 or more functional prediction algorithms

Gene Symbol	Variant	Affected patients	Translation Impact	PolyPhen-2	PROVEAN	CADD score	ExAC European Frequency (%)	Gnomad European Frequency (%)	Association of gene with disease
USH2A	p.S5188G	BBS014	Missense	Benign	Neutral	22.7	0.314	0.367	RP (OMIM: 613809); USH (OMIM: 276901)
USH2A	p.K5026E	BBS014	Missense	Benign	Neutral	<10	0.315	0.367	
USH2A	p.R4848Q	BBS014	Missense	Benign	Neutral	10.5	0.320	0.368	
USH2A	p.G4838E	BBS014	Missense	Benign	Neutral	<10	0.330	0.378	
USH2A	p.Q4541P	BBS002	Missense	Benign	Neutral	16.94	0.001	0.001	
WDR35	p.I285F	BBS002	Missense	Benign	Deleterious	<10	0.055	0.074	JATD (OMIM: 614091)
WDPCP	p.R290H	BBS010; BBS011; BBS012	Missense	Benign	Neutral	<10	0.000	0.001	BBS (OMIM: 615992)
BBS12	p.S429T[2]	BBS009	Missense	Benign	Neutral	<10	0.094	0.062	BBS (OMIM: 615989)
PKHD1	p.G698D	BBS010; BBS011	Missense	Benign	Neutral	<10	0.015	0.015	Polycystic kidney and hepatic disease (OMIM: 263200)
CEP164	p.P127L	BBS013	Missense	Benign	Neutral	19.55	0.223	0.168	NPHP (OMIM: 614845)
PKHD1	p.I3309V	BBS010; BBS011	Missense	Possibly Damaging	Neutral	15.02	0.016	0.039	Polycystic kidney and hepatic disease (OMIM: 263200)
PKHD1	p.V3263A	BBS014	Missense	Benign	Neutral	23.5	0.267	0.315	
CDH23	p.D1806E	BBS009	Missense	Probably Damaging	Neutral	18.96	0.585	0.646	USH (OMIM: 601067)
BBS4	p.V266G	BBS013	Missense	Benign	Deleterious	14.75	0.001	0.001	BBS (OMIM: 615982)

Table 3.3(iii) Tier 3: <5% ExAC European frequency (MAF<0.05), predicted damaging by 2 or more functional prediction algorithms

Gene Symbol	Variant	Affected patients	Translation Impact	PolyPhen-2	PROVEAN	CADD score	ExAC European Frequency (%)	Gnomad European Frequency (%)	Association of gene with disease
NPHP4	p.E618K	BBS013	Missense	Probably Damaging	Neutral	23.3	1.507	1.440	NPHP (OMIM: 606966), SLS (OMIM: 606996)
CC2D2A	p.K507E	BBS009	Missense	Probably Damaging	Neutral	27.1	1.080	1.038	JBTS (OMIM: 612285); MKS (OMIM: 612284)
PCDH15	p.N637S	BBS014	Missense	Possibly Damaging	Neutral	21.5	3.155	2.727	USH (OMIM: 601067/602083)
BBS10	p.D142N	BBS008; BBS010; BBS011; BBS012	Missense	Possibly Damaging	Neutral	28.3	1.291	1.245	BBS (OMIM: 615987)
CEP290	p.I2134T	BBS005	Missense	Probably Damaging	Neutral	29	1.693	0.688	BBS (OMIM: 615991); JBTS (OMIM: 610188); LCA (OMIM: 611755); MKS (OMIM: 611134); SLS (OMIM: 610189)
AHI1	p.R548H	BBS015	Missense	Benign	Deleterious	24.1	1.545	1.716	JBTS (OMIM: 608629)

Table 3.3(iv) Tier 4: <5% ExAC European frequency (MAF<0.05), predicted innocuous by 2 or more functional prediction algorithms

Gene Symbol	Protein Variant	Affected patients	Translation Impact	PolyPhen-2	PROVEAN	CADD score	ExAC European Frequency (%)	Gnomad European Frequency (%)	Association of gene with disease
NPHP4	p.R740H	BBS003	Missense	Benign	Neutral	<10	4.079	3.304	NPHP (OMIM: 606966), SLS (OMIM: 606996)
USH2A	p.E478D	BBS005	Missense	Benign	Neutral	23.8	1.720	1.790	RP (OMIM: 613809); USH (OMIM: 276901)
TMEM67	p.D261N	BBS005; BBS009	Missense	Benign	Neutral	<10	1.700	1.548	JBTS (OMIM: 610688); MKS (OMIM: 607361); NPHP (OMIM: 613550)
PCDH15	p.P1718S	BBS005	Missense	Benign	Neutral	<10	2.752	1.603	USH (OMIM: 601067/602083)
TCTN1	p.I412V	BBS002 (Hom)	Missense	Benign	Neutral	<10	1.005	0.926	JBTS (OMIM: 614173)
TCTN2	p.R200Q	BBS003; BBS008	Missense	Benign	Neutral	<10	1.242	1.240	JBTS (OMIM: 616654)
RPGRIP1L	p.A229T[3]	BBS003	Missense	Benign	Neutral	15.69	2.783	2.716	JBTS (OMIM: 611560); MKS (OMIM: 611561)
TCTN1	p.M1I	BBS002 (Hom)	Start Loss		Neutral	<10	2.201	0.996	JBTS (OMIM: 614173)
RPGRIP1L	p.A1183G	BBS013	Missense	Possibly Damaging	Neutral	10.95	1.539	1.558	JBTS (OMIM: 611560); MKS (OMIM: 611561)

Table 3.3 – Putative modifying variants identified through the 4-tier categorisation strategy. Modifiers that are known to act in BBS were uncovered: *TTC21B* p.L1002V[1] (Davis et al. 2011); *BBS12* p.S429T[2] (Zaghloul et al. 2010); *RPGRIP1L* p.A229T[3] (Khanna et al. 2009). Disease definitions: Bardet-Biedl Syndrome, BBS; Nephronophthisis, NPHP; Senior-Loken Syndrome, SLS; Retinitis Pigmentosa, RP; Jeune Asphyxiating Thoracic Dysplasia, JATD; Joubert Syndrome, JBTS; Leber Congenital Amaurosis, LCA; Meckel Syndrome, MKS; Usher syndrome, USH.

Tier 3 analysis revealed 6 heterozygous missense variants predicted to be deleterious. Finally, tier 4 analysis, where functional predictions were innocuous, identified an additional 9 variants in 7 genes. 2 of these were homozygous mutations in *TNTN1*, both found in patient BBS002.

The number of variants per patient at each tier is summarised in **table 3.4**. Notably, patients BBS001 and BBS004 have no additional qualifying variants other than the underlying pathogenic alleles. On the other hand, patient BBS005 and BBS014 have the greatest number of potential modifying alleles, with 6 alleles in 6 genes and 7 alleles in 4 genes, respectively.

Patient ID	Allele 1	Allele 2	Number of potentially modifying alleles					Affected genes	
			Tier 1	Tier 2	Tier 3	Tier 4	Total	Number	Gene names
BBS001	*BBS1* p.M390R	*BBS1* p.M390R	0	0	0	0	0	0	
BBS002	*BBS1* p.M390R	*BBS1* p.M390R	0	2	0	4	6	3	*USH2A; WDR35; TCTN1*
BBS003	*BBS1* p.M390R	*BBS1* p.M390R	0	0	0	3	3	3	*NPHP4; TCTN2; RPGRIP1L*
BBS004	*BBS1* p.M390R	*BBS1* c.723+1G>A	0	0	0	0	0	0	
BBS005	*BBS1* p.M390R	*BBS1* p.M390R	2	0	0	3	6	6	*ARL6; NPHP4; CEP290; USH2A; TMEM67*
BBS006	*BBS1* p.M390R	*BBS1* p.G73*	2	0	0	0	2	1	*TTC21B*
BBS007	*BBS1* p.M390R	*BBS1* p.M390R	1	0	0	0	1	1	*C2ORF71*
BBS008	*BBS1* p.M390R	*BBS1* p.M390R	1	0	1	1	3	3	*TTC21B; BBS10; TCTN2*
BBS009	*BBS1* p.M390R	*BBS1* p.M390R	0	2	1	1	4	4	*BBS12; CDH23; CC2D2A; TMEM67*
BBS010†	*BBS1* p.M390R	*BBS1* p.Y284fs*5	0	3	1	0	4	3	*WDPCP; PKHD1; BBS10*
BBS011†	*BBS1* p.M390R	*BBS1* p.Y284fs*5	0	3	1	0	4	3	*WDPCP; PKHD1; BBS10*
BBS012†	*BBS1* p.M390R	*BBS1* p.Y284fs*5	1	1	1	0	3	3	*WDPCP; XPNPEP3; BBS10*
BBS013	*BBS1* p.M390R	*BBS1* p.M390R	1	2	1	1	5	5	*SDCCAG8; CEP164; BBS4; NPHP4; RPGRIP1L*
BBS014	*BBS1* p.M390R	*BBS1* p.G73*	1	5	1	0	7	4	*CEP164; USH2A; PKHD1; PCDH15*
BBS015	*BBS1* p.M390R	*BBS1* p.M390R	0	0	1	0	1	1	*AHI1*

Table 3.4 – A summary of the number of potentially modifying variants per patient that were detected at each tier. The number and names of affected genes are also presented. † denotes related individuals where BBS010 and BBS011 are monozygotic twins and BBS012 is their sibling.

3.2.2.2. Location of modifiers within the cilium

Exclusive of the primary driving mutations in *BBS1*, 37 variants were identified in 23 ciliopathy genes. The image represents the localisation of each protein that is encoded by qualifying genes within the cilia (figure 3.3) as established by the SYSCILIA project (van Dam et al. 2013) (appendix, table A.2).

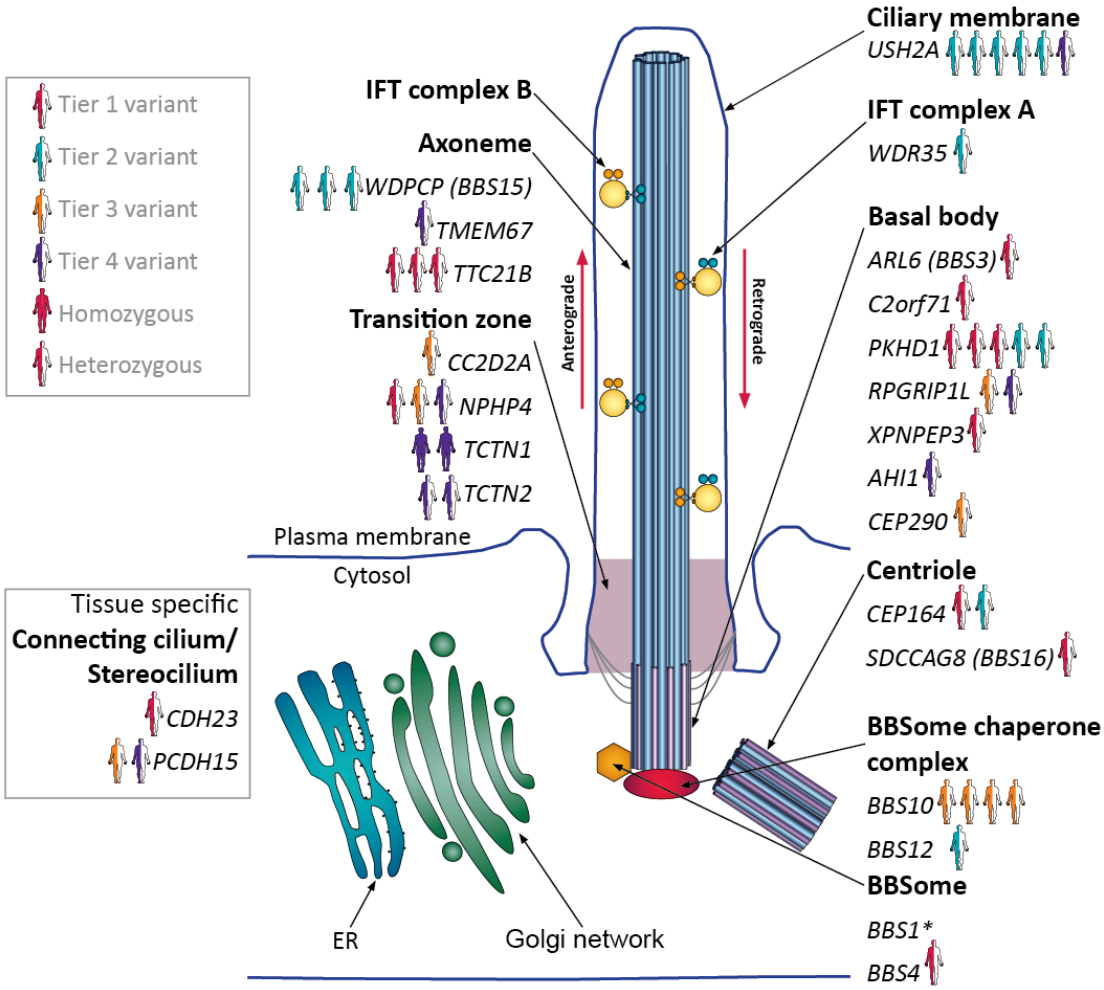

Figure 3.3 – The location of proteins encoded by qualifying genes within the primary cilium, where the presence of a variant is illustrated by a silhouette. Each silhouette is coded by tier (colour) and zygosity (50% = heterozygous; 100% = homozygous). Tiered strategy is as previously described, where tier 1 = MAF<1%, predicted deleterious; tier 2 = MAF<1%, predicted innocuous; tier 3 = MAF<5%, predicted deleterious; tier 4 = MAF<5%, predicted innocuous.

This demonstrates that the majority of affected genes (7/23) localise to the basal body, where the BBS proteins within the BBSome also localise (Wei et al. 2012). However, the image also illustrates a global effect of variants on the cilium, as almost all structural elements of the cilium are affected. Furthermore, three variants lie in tissue specific genes, such as *CDH23* and *PCDH15*, which localise to the connecting cilium of photoreceptors and stereocilium of cochlear cells (van

Dam et al. 2013). These variant findings may contribute to an overall mutational burden in ciliopathy genes for these patients.

Moreover, to determine the effect that these modifiers may have at a functional level, the STRING database (Szklarczyk et al. 2015) was utilised to predict direct and indirect protein-protein interactions with BBS1, which were assembled into a network (**figure 3.4**). This interaction map demonstrates the close-knit relationship between these genes. BBS1 is known through experimental means or curated databases to interact directly with BBS10, ARL6, BBS4, BBS12, and CEP290. A further 7 interactions are predicted to interact by knowledge obtained from text mining (namely, WDPCP, TTC21B, NPHP4, RPGRIP1L, CC2D2A, AHI1, and TMEM67). Through investigation of indirect interactions, BBS1 is predicted to network with all remaining gene products through intermediate proteins, predominantly via CEP290. The only exception is C2ORF71, which has no known interactors.

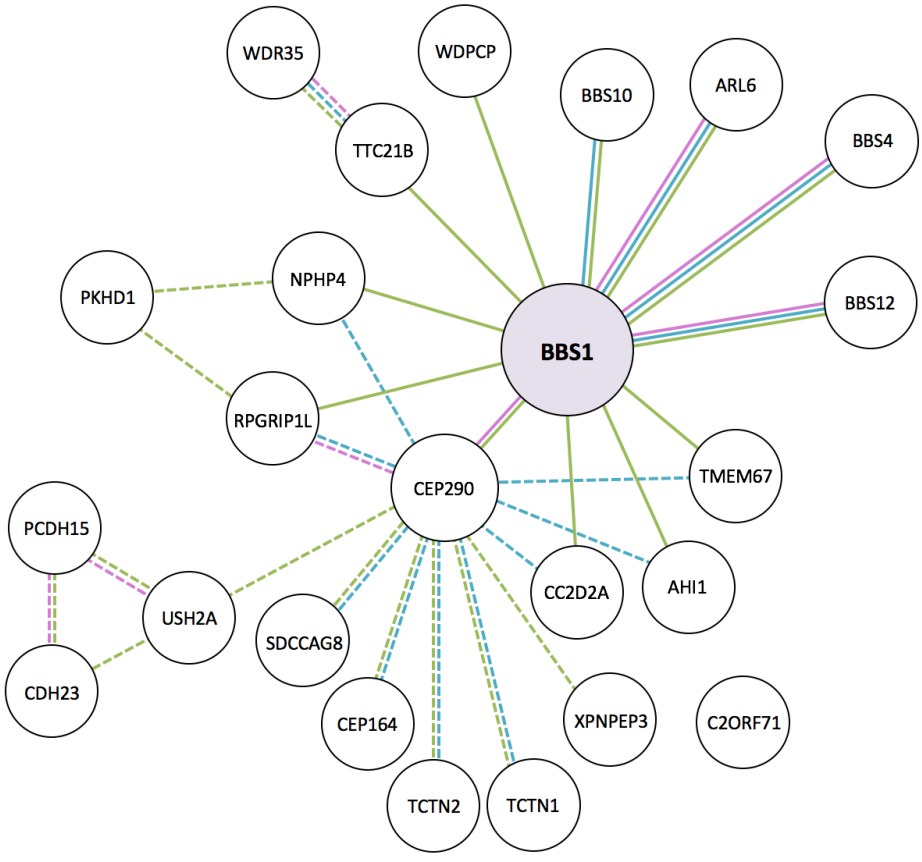

Figure 3.4 – The interacting network of BBS1 with the protein products of genes identified through 4-tier categorisation. Interactions are experimentally determined (pink connectors), from literature mining (green connectors) or from curated databases (blue connectors). Direct interactions with BBS1 are indicated by solid lines, whereas indirect interactions through intermediates are denoted as detached lines. Image adapted from information provided by the STRING database.

3.2.2.3. Known modifiers

Following identification of these variants, a literature search was executed to determine whether patients were harbouring any established modifier mutations. In total, three known hypomorphic variants were found. Namely, *BBS12* p.S429T in patient BBS009 (Zaghloul et al. 2010), *TTC21B* p.L1002V in patients BBS006 and BBS008 (Davis et al. 2011), and *RPGRIP1L* p.A229T in patient BBS003 (Khanna et al. 2009). This demonstrates that the 4-tier strategy was successful in uncovering variants known to influence BBS expressivity.

Furthermore, a previously reported variant in *AHI1* p.R548H was detected in BBS015, but was disregarded as pathogenic in a study of JBTS patients as it did not segregate exclusively with the disease (Parisi et al. 2006). However, lack of segregation does not exclude this variant from having a role as a modifier. Similarly, the heterozygous variant in *NPHP4* p.R735W found in patient BBS005 was previously identified in a nephronophthisis patient, though it was not determined whether this variant was a main contributor to the disease (Hoefele et al. 2005).

In addition to finding known modifiers in the literature, two variants identified in this study were also previously predicted to be benign through morphometric analysis. Specifically, *BBS10* p.D142N found in BBS008 and the siblings BBS010, BBS011 and BBS012 (Zaghloul et al. 2010), as well as *TTC21B* p.R616C in patient BBS006 (Davis et al. 2011).

A variant that may be of interest is *ARL6 (BBS3)* p.T32M that was uncovered in patient BBS005. This is a rare variant that has never been reported in an individual of European descent. Gnomad database reports a frequency of 2/245868 in the total population, MAF=$8.1x10^{-6}$, due to two heterozygous variants identified in the South Asian population. Furthermore, this variant is adjacent to an amino acid change known to cause BBS (p.T31M). Indeed, *ARL6* p.T31 was screened by Zaghloul and colleagues, which exhibited a dominant negative effect when substituted to either methionine or arginine (Zaghloul et al. 2010). Further investigation of the variant p.T31M discovered that this residue is in a highly conserved GTP binding pocket of ARL6, where the nucleotide binding sequence is known as the P-loop [G-N(x4)-G-K-T/S] (Fan et al. 2004) (figure 3.5). Although not directly part of the P-loop, this variant could still play a crucial role in maintaining the structure of the GTP binding pocket.

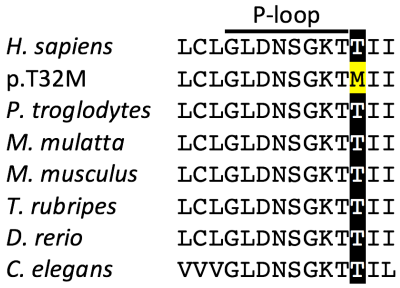

Figure 3.5 – Conservation of ARL6 p.T32M. The amino acid p.T32 is conserved through vertebrates and adjacent to the functional phosphate binding loop (P-loop). The variant p.T32M is highlighted in yellow.

3.2.2.4. Candidate modifiers

From the tiered modifier analysis, potential variants have been assessed and table of the more likely candidate mutations has been compiled **(table 3.5)**. These have been chosen after assessment of the commonality, functional prediction, the location of affected proteins within the primary cilium, and the network of interactions in close proximity with BBS1. Where applicable, known modifiers have also been included. A confidence grading of low or high was assigned to each candidate depending on the strength of evidence as a modifier of BBS. The justification for each confidence rating is further examined in the discussion section of this chapter.

Patient ID	Gene	Variant	Tier	Confidence	Confidence justification
BBS002	*TCTN1*	p.M1I	4	Low	Uncertain significance
BBS003	*RPGRIP1L*	p.A229T	4	High	Known modifier
BBS005	*ARL6*	p.T32M	1	High	Highly conserved/next to known BBS variant/direct interaction with BBS1
BBS006	*TTC21B*	p.L1002V	1	High	Known modifier
BBS007	*C2ORF71*	p.S1085L	1	Low	Uncertain significance
BBS008	*TTC21B*	p.L1002V	1	High	Known modifier
BBS009	*BBS12*	p.S429T	2	High	Known modifier
BBS010/011†	*PKHD1*	p.I3309V	2	Low	Uncertain significance
		p.G698D	2	Low	Uncertain significance
BBS010/011/012†	*WDPCP*	p.R290H	2	Low	Uncertain significance
BBS012†	*XPNPEP3*	p.P493A	1	Low	Uncertain significance
BBS013	*BBS4*	p.V266G	2	High	Direct interaction with BBS1
BBS014	*CEP164*	p.Q1410*	1	High	Deleterious nonsense variant
BBS015	*AHI1*	p.R548H	3	High	Known modifying gene

Table 3.5 – The candidate modifying variant for each patient with a confidence grading of high or low.

† denotes related individuals where BBS010 and BBS011 are monozygotic twins and BBS012 is their sibling.

3.2.3. Mutational burden analysis

The second hypothesis of this chapter, that enhanced background variation attributes to mutational burden and contributes to disease severity in patients, was subsequently investigated.

It was explored whether there was an accumulation of mutational load in BBS patients compared to controls, using both whole genome and exome datasets. These datasets were first considered separately and then combined to observe the effect of a larger sample size on the analysis. Only rare variants (MAF<1%) predicted pathogenic by SIFT or PolyPhen-2 or CADD score>20 were considered. First, variant counting was restricted to a panel of genes known to cause ciliopathies of the primary cilium (n=73). The panel was then expanded to encompass a broader range of genes from the primary cilium network, referred to as SYSCILIA's SCGSv1 panel (n=235) (van Dam et al. 2013). As the ethnicity of the WES BBS patients was not exclusively European, the total ExAC population frequency data was utilised.

3.2.3.1. Mutational burden of patients from WES data

The WES patient data were taken from a set of 16 patients exhibiting BBS who were sequenced to identify the disease-causing alleles (**appendix, table A.4**). Mutational analysis for a genetic diagnosis was implemented by GOSgene, UCL GOS Institute of Child Health, who confirmed disease causing variants in 10 patients (**table 3.6**). The remaining 6 were unresolved without a genetic diagnosis.

Sample ID	Gene Symbol	Allele 1	Allele 2
WES_508	*BBS12*	p.G540V	p.V409M
WES_511	*BBS12*	p.L530P	p.L530P
WES_512	*MKKS*	p.T259fs*21	p.T259fs*21
WES_513	*MKKS*	p.T259fs*21	p.T259fs*21
WES_514	*BBS12*	p.L530P	p.L530P
WES_519	*BBS12*	p.C426Y	p.E365fs*18
WES_521	*BBS1*	p.R570*	p.I200_T201del
WES_522	*MKKS*	p.I339fs*3	p.I339fs*3
WES_524	*TTC8*	c.116+2T>C	c.116+2T>C
WES_525	*BBS12*	p.S21*	p.S21*
WES_506	No diagnosis confirmed		
WES_507	No diagnosis confirmed		
WES_502	No diagnosis confirmed		
WES_518	No diagnosis confirmed		
WES_520	No diagnosis confirmed		
WES_523	No diagnosis confirmed		

Table 3.6 – Known genetic diagnoses for 10 of 16 patients sequenced by WES.

The first set of burden analysis conducted was with all 16 WES patients compared against a set of 25 non-ciliopathy control individuals (**appendix, table A.4**). There was a significant enrichment

of variants in patients compared to controls (mean variant count 4.8 vs. 3.1 respectively; p=0.006) (table 3.7i).

As an autosomal recessive disease, there are two pathogenic segregating alleles per patient that are diagnosed to be the main contributor to disease. In this study, the combination of these alleles has been defined as the pathogenic driving mutation (PDM). By definition, patients with a definitive molecular diagnosis are expected to have additional burden on disease genes that is augmented from the typical background level of variation seen in the general population. Next, it was tested whether there was still an enrichment of background pathogenic variants when the PDMs were excluded from analysis. When PDMs were removed, the mean number of pathogenic variants reduced to a level that was similar to controls (table 3.7ii). This suggested that the significant difference seen in table 3.7i was being driven by the PDMs and that there was no additional mutational burden without PDMs within the 73 ciliopathy genes tested.

The hypothesis that the PDMs were the significant factor in the analysis was further supported when the group was separated into cases with and without a definitive molecular diagnosis. The count of variants in 10 diagnosed cases (including PDMs) was significantly greater with a mean of 5.5 alleles per patient compared to 3.1 for controls (p=0.001). However, when the PDMs were removed the mean variant count diminished to a comparable count as the 6 undiagnosed patients alone (3.5 and 3.7, respectively) (table 3.7iii). Both of these were not significantly different to controls (p=0.524 and 0.439, respectively), demonstrating that the driving mutations for the undetermined cases may be in genes not found in the ciliopathy panel.

	Filtering criteria	Data comparison	Mean number of variant alleles per patient	Mean number of variant alleles per control	Fisher Exact p-value
i)	+PDM	16 BBS vs. 25 controls	4.8	3.1	0.006*
ii)	-PDM	16 BBS vs. 25 controls	3.6	3.1	0.419
iii)	+PDM	10 diagnosed vs. 25 controls	5.5	3.1	0.001*
	-PDM	10 diagnosed vs. 25 controls	3.5	3.1	0.524
		6 undiagnosed vs. 25 controls	3.7	3.1	0.439

Table 3.7 – Burden analysis comparison against WES patient and control datasets with ciliopathy gene panel (n=73), with and without primary driving mutation included in the analysis (PDM).

* denotes statistically significant p-value where α=0.05.

3.2.3.2. Mutation burden of WGS data

An identical burden analysis strategy for the WES dataset was implemented with the WGS data generated from 14 BBS patients compared with 30 non-ciliopathy control individuals. As genetic material from monozygotic twins is highly similar, BBS011 was excluded from the burden analysis to prevent artificial bias created from pseudo duplicate samples. As the 14 WGS patients all had a molecular diagnosis in *BBS1*, it was expected that the burden analysis with PDMs would be similar to the WES analysis. However, the number of variants per patient was closely matched to controls, with mean counts of 3.7 and 3.6, respectively (p=0.863) (table 3.8i). This did not mimic the high significance observed between WES patient and control sets. To investigate the effect the PDMs in *BBS1* were having on the outcome of the burden test, these alleles were excluded from further analysis. As a result, there was a large decrease in mean variants per patient compared to control (1.7 vs. 3.6, respectively), which elicited a highly significant p-value of 0.0004 (table 3.8ii).

	Filtering criteria	Data comparison	Mean number of variant alleles per patient	Mean number of variant alleles per control	Fisher Exact p-value
i)	+PDM	14 patients vs. 30 controls	3.7	3.6	0.863
ii)	-PDM	14 patients vs. 30 controls	1.7	3.6	0.0004*

Table 3.8 – Burden analysis comparison against WGS patient and control datasets with ciliopathy gene panel (n=73), with and without primary driving mutation included in the analysis (PDM). * denotes statistically significant p-value where α=0.05.

3.2.3.3. Combination of WES and WGS

Next the two datasets were integrated to give a combined analysis of 30 patients vs. 55 controls. A collective average for BBS patients was 4.3 variants compared to 3.4 variants for controls (p=0.035) (table 3.9i). With the exclusion of the PDMs from the analysis, the mean variant count per patient was not significantly lower than the controls (unlike the result seen with WGS alone) (table 3.9ii).

	Filtering criteria	Data comparison	Mean number of variant alleles per patient	Mean number of variant alleles per control	Fisher Exact p-value
i)	+PDM	WES + WGS combined	4.3	3.4	0.035*
ii)	-PDM	WES + WGS combined	2.7	3.4	0.109

Table 3.9 – Burden analysis comparison against integrated datasets with ciliopathy gene panel (n=73), with and without primary driving mutation included in the analysis (PDM). * denotes statistically significant p-value where α=0.05.

3.2.3.4. Mutational burden within a larger gene panel

The results of the ciliopathy panel suggested that there was no mutational burden in the 73 genes when the PDMs were subtracted. The next line of investigation was to expand the gene list to encompass an additional 162 genes from the SYSCILIA project, which have central roles within the primary cilia but are not directly associated with ciliopathies (van Dam et al. 2013).

The results of mutational burden testing with a larger gene panel mimicked prior analyses. When PDMs were included in the calculations, there were significantly more variants in the WES patients compared to controls (mean of 12.8 vs. 10.6 variants, respectively; p=0.006) (table 3.10i). However, like the ciliopathy panel, there was no additional mutational burden compared to controls when the PDMs were subtracted from the analysis. Similarly, there was no enrichment of variants in WGS patients compared to controls when PDMs were present (mean of 9.3 vs. 9.9 variants, respectively; p=0.522). When the PDMs were removed from the analysis, there was a significant decrease in the mean number of variants compared to controls (7.2 vs. 9.9, respectively; p=0.006) (table 3.10ii). In contrast to the ciliopathy panel, there was no significant difference between patients and controls when data sets were combined, both with and without PDMs (table 3.10iii). This analysis demonstrated that even within a larger primary cilia specific gene panel, there was no mutational burden observed for these patients compared to controls.

	Filtering criteria	Data comparison	Mean number of variant alleles per patient	Mean number of variant alleles per control	Fisher Exact p-value
i)	+PDM	WES 16 BBS vs. 25 controls	12.8	10.6	0.006*
	-PDM		11.6	10.6	0.350
ii)	+PDM	WGS 14 patients vs. 30 controls	9.2	9.9	0.522
	-PDM		7.2	9.9	0.006*
iii)	+PDM	WES + WGS combined	11.2	10.2	0.206
	-PDM		9.6	10.2	0.378

Table 3.10 – Burden analysis comparison with SCGSv1 gene panel (n=235), with and without primary driving mutation included in the analysis (PDM). * denotes statistically significant p-value where α=0.05.

3.2.3.5. Burden analysis with control gene panel

To establish whether the significant differences observed (or not observed) were specific to the ciliopathy and SCGSv1 panels, a similarly sized unrelated disease panel was utilised. This featured 74 genes associated with the inherited motor neuropathy, Charcot-Marie-Tooth (CMT) disease (**table 3.11;** appendix**, A.3**). This gene list has previously been investigated by Gonzaga-Jauregui and colleagues who observed mutational burden in a group of CMT patients compared to controls (Gonzaga-Jauregui et al. 2015).

The average number of variants were concordant between patient and controls (**table 3.11**), though some variability was observed between WES and WGS sets (combined mean of 1.9 and 2.5, respectively). This inconsistency could be due to sequencing artefacts distinct to one technique over the other. As there was no significant difference in the number of variants between patients and controls, this analysis offers evidence that the differences observed in the burden testing are specific to ciliopathy genes.

Filtering criteria	Data comparison	Mean number of variant alleles per patient	Mean number of variant alleles per control	Fisher Exact p-value
CMT gene panel	WGS only	1.8	2.0	0.723
	WES only	2.6	2.4	0.678
	Combined data sets	2.2	2.15	0.931

Table 3.11 – Comparison between WGS, WES, and combined datasets with a control gene panel (n=74)

3.2.3.6. Technical investigations

The unexpected finding presented here was the observation that there were significantly fewer variants in BBS patients than controls when *BBS1* PDMs were removed from the WGS cohort. I set out to determine whether this could be a technical issue caused by more variants overall in the controls than patients. Although the sequencing company, technology, and bioinformatics analysis were identical for all, it was possible that batch effects may exist as the 15 BBS patients, 20 SRS, and 10 VEO-IBD controls were sequenced in different batches. To investigate potential technological inconsistencies, the number of total variants was counted separately for each individual. It was then recorded how many variants were considered after each filter was applied in IVA. The mean number of qualifying variants for patients and controls were calculated and a Student's t-test conducted to assess whether there were significantly more variants in the control than patient cohort. The number of qualifying variants with the respective filtering criteria is presented in **figure 3.6**.

On average for total variants, there were ~300 more variants in the control group than the BBS cohort (**figure 3.6A**). However, at α=0.05 there was not sufficient evidence to reject the null hypothesis that the group means differ significantly (p=0.078; **supplemental data table S.40**). Further testing was completed throughout the filtering process, where the conclusion that differences were not statistically significant was found at each level (**figure 3.6B and 3.6C**). As previously shown, when the PDMs were subsequently removed from the analysis, there was a significant decrease in the number of variants per patient compared to controls (**figure 3.6C**).

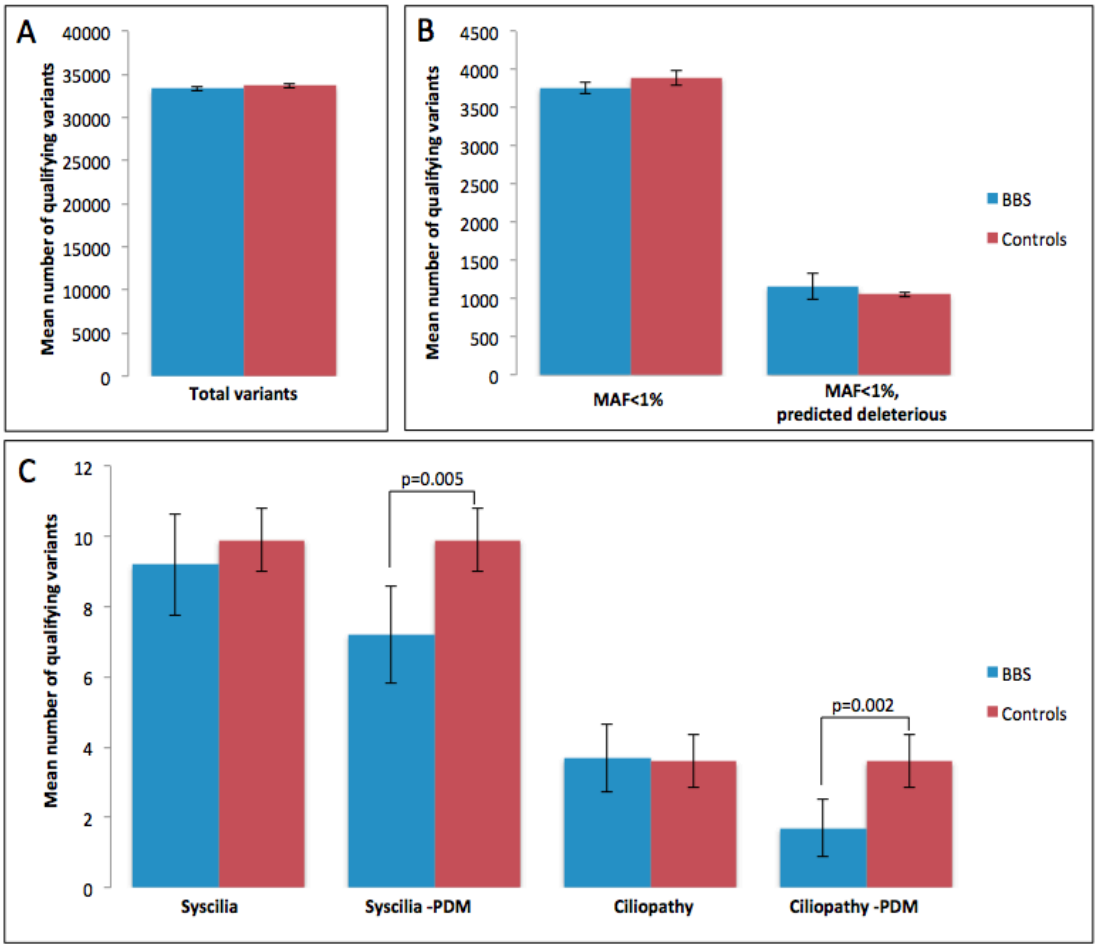

Figure 3.6 – The mean number of variants throughout IVA filtering for BBS patients (blue) and controls (red). A) mean number of total variants without filtering. B) mean number of variants with MAF<1% ExAC population frequency filter applied (left), and with the addition of predicted deleterious by SIFT, PolyPhen-2 or CADD>20 (right). C) SCGSc1 (n=235) and Ciliopathy (n=73) gene panels were added with and without the presence of PDMs in the analysis. Error bars correspond to 95% confidence intervals and statistically significant differences when α=0.05 are indicated.

All samples were whole genome sequenced to an average coverage of 30X. Although the sequencing is expected to be robust across different batches, there can be some variability due to sequencing errors caused by GC rich or repetitive regions of the genome. It was therefore inspected whether there were any irregularities in sequence coverage over 71 genes from the ciliopathy panel for patients and controls (*OFD1* and *RPGR* found on chromosome X were removed from this analysis, as they would add sex bias) (Figure 3.7).

All subjects had high sequence coverage over the ciliopathy panel, exceeding 30X in all cases (figure 3.7A). On average, BBS patients had a higher coverage compared to control subjects (36X and 33X, respectively), demonstrating that both were sequenced to the expected standard. The coverage was also assessed at an individual level by calculating the percentage of bases covered at different read depths (figure 3.7B). On average, 76% of bases were covered with depth greater than 30X for BBS patients, compared to 69% for controls. Both groups had an average of 96% of bases covered with coverage greater than 20X. Taken together, these data show no clear technical sequencing abnormalities in both datasets, and they were sequenced to a high quality. This supports the conclusion that the mutational burden testing was not driven by technical anomalies.

Figure 3.7 – Mean coverage data to assess the quality of WGS data at a gene (A) and individual (B) level. A) The mean coverage calculated across each gene in the ciliopathy panel for 15 BBS patients (blue) and 30 controls (red). B) The percentage of bases averaged over the ciliopathy gene panel covered to a depth greater than 30X (blue), between 20-30X (red), between 10-20X (green) and less than 10X (purple), for BBS patients (left) and control individuals (right).

3.3. Discussion

3.3.1. Identification of potential modifier alleles

In this study, whole genome sequencing was utilised to uncover potential modifying alleles that could explain the phenotypic heterogeneity commonly observed in BBS patients. A 4-tier categorisation system was devised to filter over 11 million variants down to 41 potential modifying mutations. In addition to the underlying pathogenic mutations in *BBS1*, 13 out of 15 patients possessed potential secondary mutations with MAF<5%. 7 of these patients had variants identified that were predicted to be deleterious by functional prediction algorithms, as well as defined as rare (MAF<1%). Furthermore, the tiered system identified previously established hypomorphic modifying mutations in *BBS12*, *TTC21B* and *RPGRIP1L* in 4 patients. From the 41 candidate modifiers, each affected gene was considered for their distribution throughout the primary cilium and assessed for their direct interaction with *BBS1*. As a result of this, as well as acknowledgement of predicted pathogenicity and rare minor allele frequencies, a shortlist of candidate modifier variants was constructed (table 3.5).

3.3.1.1. High confidence candidate modifier alleles

For comprehensive characterisation of these candidate modifying variants, functional analysis should be undertaken, as well as further assessment of these candidates in a larger population of BBS patients. However, as this is beyond the scope of this project, I have selected a number of variants where evidence exists in literature and databases to support their role as a modifier. These are classed as high confidence modifiers and discussed here in detail.

BBS003: RPGRIP1L

RPGRIP1L p.A229T was identified in BBS003, who presents with severe BBS including early onset RP and renal defects. *RPGRIP1L* is mutated in MKS and JBST, but has been shown to act as a modifier specifically for RP in other ciliopathies (Khanna et al. 2009). Khanna and colleagues (2009) identified p.A229T, a common polymorphic variant in *RPGRIP1L* that was enriched in 18 ciliopathy patients of European descent compared to 3016 control alleles. 4 of these 18 ciliopathy cases were BBS patients. Furthermore, through morphogenic zebrafish experiments, *RPGRIP1L* p.A229T was validated to act as a hypomorph. In particular, this variant was only present in ciliopathy cases where RP is a diagnostic feature, such as BBS or SLS. The manuscript continued to suggest that the absence of this variant may protect from RP in other ciliopathy patients. However, this claim is not supported by the fact that p.A229T is absent from the 14 remaining

BBS patients in this study who also suffer from RP. The proposed mechanism involved the collaborative relationship between *RPGRIP1L* and *RPGR*, which was abrogated by substitution to p.T229. The study by Khanna et al. suggested that this interaction may play a larger role in photoreceptors, due to the significant role of *RPGR* in RP (Ayyagari et al. 2002). It may be postulated that this phenomenon also affect kidney cells, as both *RPGRIP1L* and *RPGR* are expressed in renal tissue and aberrations can lead to ciliopathies featuring nephropathy (Wolf et al. 2007; Delous et al. 2007; Patil et al. 2010). Despite this variant being fairly common and predicted innocuous, this evidence indicates that it may play a pivotal role in the severity of BBS in patient BBS003.

BBS005: ARL6 (BBS3)

Out of all 15 patients, BBS005 has the greatest number of qualifying variants, with 6 mutations in 6 genes. The most convincing candidate modifier is the variant in the GTP-binding protein *ARL6*. A direct interaction between ARL6 and BBS1 has been shown by X-ray crystallography, which is required for recruitment of the BBSome to ciliary membrane (Mourão et al. 2014). Mourão et al. (2014) hypothesised that the p.M390R variant destabilises this interaction with ARL6, thus an additional hit in *ARL6* could exacerbate this disruption further. Moreover, the adjacent residue to T32 is another threonine, which contributes to a conserved nucleotide binding site called a P-loop. Mutation of p.T31M is known to cause BBS by disruption of GTP binding (Fan et al. 2004). Structural analysis by Fan et al. (2004) determined that p.T32 is also within the nucleotide binding pocket of ARL6, and thus a substitution to a methionine could sufficiently alter the structure of this pocket and disrupt GTP binding. The rarity of the mutation (MAF<0.001% of total Gnomad database) and position within a functional domain of the protein could render p.T32M as a deleterious variant, perhaps more so in collaboration with *BBS1* p.M390R.

BBS006, BBS008: TTC21B

TTC21B is mutated in NPHP and the skeletal ciliopathy JATD, and encodes an intraflagellar transport protein in the IFT-A particle (Davis et al. 2011). Tiered modifier analysis identified 3 hits in *TTC21B* that affect 2 patients, where 2 of these hits are the same variant (p.L1002V) in BBS006 and BBS008. This mutation was identified as a modifier by Davis et al. (2011) after it was found as a second site hit in 5 ciliopathy patients. An mRNA construct of p.L1002 was subsequently expressed in the photoreceptors of a *Ttc21b* knock-down rat retina. It was observed that p.L1002V could not fully rescue TTC21B mislocalisation within the cilia caused by the gene knock down, but did partially restore ciliary length. Due to the partial rehabilitation of the phenotype,

this variant was considered to have a hypomorphic effect. In a similar experiment, Davis and colleagues (2011) determined that the third hit in TTC21B (p.R616C), identified in BBS006, was benign.

BBS009: BBS12

A final known modifier, *BBS12* p.S429T, was identified in patient BBS009. BBS12 is a member of a chaperone complex required for assembly of the BBSome (Seo et al. 2010). This point mutation was interpreted by PROVEAN and PolyPhen-2 to be neutral/benign and had a CADD score<10. The reason for this prediction is due to the substitution from a serine to a threonine, which both contain polar side chains. Regardless, *BBS12* p.S429T was screened in a zebrafish morphometric experiment and was found to have a hypomorphic affect (Zaghloul et al. 2010). This demonstrates how modifying variants may not be deleterious when expressed alone but induce a variable phenotype when in combination with the primary driving mutation.

BBS013: BBS4

In patient BBS013, another suitable candidate modifier was identified in *BBS4* (p.V226G), which is a fellow member of the BBSome with *BBS1*, and has previously been highlighted as a candidate for oligogenicity in BBS (Katsanis et al. 2002). Additional roles for BBS4 are known, including the interaction with centriolar satellite proteins. One such interactor is pericentriolar material 1 (PCM1), which is required for correct localisation of proteins to the centrosome for cilia assembly (Kim et al. 2004; Chamling et al. 2014). The BBS4 protein has a well conserved structure, consisting of 13 tandem tetratricopeptide repeat (TPR) motifs (Kim et al. 2004). These motifs are found in many proteins involved in complex formation, and critically, are highly enriched in ciliary proteins (Xu et al. 2015; Davis et al. 2011). The amino acid change p.V226G found in patient BBS013 lies within the 6th TPR of BBS4 and thus may abrogate the binding properties of BBS4 for assembly of centriolar satellite or BBSome complexes.

BBS014: CEP164

Except for two variants, the majority of candidate modifying variants discovered were heterozygous missense mutations. One exception is a heterozygous nonsense mutation in *CEP164* (p.Q1410*), identified in BBS014. *CEP164* is linked to NPHP with extra-renal features, such as RP and obesity (Chaki et al. 2012). Hence, this variant is a good candidate as a modifier for BBS014 as this patient had stage-5 chronic kidney disease. *CEP164* encodes for a centrosomal protein that is required for primary cilium formation, as well as having important roles in the DNA

damage response pathway and cell cycle progression (Graser et al. 2007; Slaats et al. 2014). Furthermore, CEP164 interacts with SDCCAG8 (Chaki et al. 2012), a centrosomal protein that has been shown to cause BBS (Schaefer et al. 2011). As a nonsense mutation, this shortened transcript could disrupt the function of the protein, especially since this truncation prevents the translation of a known phosphorylation site (Zhou et al. 2013).

BBS015: AHI1

Patient BBS015 has a fairly mild presentation of BBS but suffered from early onset retinal dystrophy. The only variant that was identified in this patient was in *AHI1* (p.R548H), which is mutated in JBTS with high penetrance for retinal dystrophy (Valente et al. 2006; Parisi et al. 2006). *AHI1* has been suspected of demonstrating modifying effects in other ciliopathies, including NPHP (Tory et al. 2007; Louie et al. 2010) and LCA (Coppieters et al. 2010). It also exhibited a vital role in the formation of the outer segments of photoreceptors and suppression of *Ahi1* in mice resulted in early photoreceptor degeneration (Louie et al. 2010; Westfall et al. 2010). Therefore, it is possible that p.R548H acted as a modifying variant specifically towards rod-cone dystrophy and enhanced retinal degeneration in BBS015.

3.3.1.2. Low confidence candidate modifier alleles

Some of the variants chosen as candidate alleles have little evidence for their potential role as modifiers. This may be due to tissue specific gene expression that contradicts the patient phenotype, or too little is known about the gene. These were classified as low confidence modifier alleles and are outlined here.

BBS002: TCTN1

In patient BBS002, a homozygous start loss variant was discovered in a transcript of *TCTN1* (p.M1I). *TCTN1* is a regulator of the Hedgehog pathway (Reiter and Skarnes 2006). In addition to this role, TCTN1 regulates ciliogenesis in certain tissues and controls the localisation of some transition zone proteins, such as MKS1 and TMEM67 (Garcia-Gonzalo et al. 2011). Mutations in *TCTN1* can cause the developmental ciliopathy, JBTS (Huppke et al. 2015). There are at least 5 protein coding transcripts for *TCTN1*, where the p.M1I variant only interrupts one transcript (RefSeq transcript ID: NM_001173975) and falls in an intronic region between exon 1 and 2, or in the 3' untranslated region (UTR) of the remaining transcripts. Although it is known that *TCTN1* expression differs in different tissue types (Garcia-Gonzalo et al. 2011), there is no evidence of

whether NM_001173975 has a distinct role that is not achieved by the other 4 transcripts. Therefore, it is uncertain whether the loss of a start site will actually present a phenotype.

BBS007: C2ORF71

Patient BBS007 is 1 of 3 patients to suffer from kidney failure in this cohort. However, the only qualifying variant identified in this patient was a heterozygous substitution in *C2ORF71*, which has only been found to cause nonsyndromic autosomal-recessive RP (Collin et al. 2010). It has been suggested that *C2ORF71* is consistent with a gene that is expressed specifically in the outer segment or connecting cilium of photoreceptors (Nishimura et al. 2010). However, the Human Protein Atlas (Uhlén et al. 2015) shows some expression in other tissue types, including the kidney, where there is weak expression in the tubules but high expression in the glomeruli (Human Protein Atlas 2017a).

BBS010, BBS011, BBS012: PKHD1, WDPCP, XPNPEP3

The discovery of genetic modifiers can be facilitated by examining discordant siblings, where segregation analysis can be utilised for discovery of unique variants that may explain a phenotype in one sibling and not the other (Génin et al. 2008). In this study, patients BBS010 and BBS011 are monozygotic (MZ) twins and BBS012 is an older sibling. Interestingly, despite the same underlying pathogenic variants in *BBS1*, BBS010 was the only patient to display polydactyly at birth and BBS012 had later onset retinal degeneration with no learning difficulties.

WDPCP (BBS15) was shown to cause BBS after a homozygous mutation was found to segregate in a patient by Kim et al. (2010). Although this has been the only case with published data linking *WDPCP* with BBS, it has been shown that WDPCP is required for ciliogenesis, as well as other key regulatory processes such as maintaining cell polarity (Cui et al. 2013). *WDPCP* p.R290H was the only candidate variant consistent across all 3 siblings, but predicted to be benign as position 290 is not well conserved in orthologues. Although this does not exclude this variant as a potential modifier, there is currently no evidence that it would have a significant effect on expressivity. It is of note that this variant is extremely rare, as only 1 out of 244598 alleles are found in the Gnomad database.

Although there was no distinction between BBS010 and BBS011 to explain the manifestation of polydactyly, there were two variants in *PKHD1* that were not present in BBS012 (p.I3309V; p.G698D). However, this variant was assigned low confidence due to expression primarily in renal

and hepatic cells (GTEx Portal 2017; Human Protein Atlas 2017b; Aguet et al. 2016), where mutations in *PKHD1* cause polycystic kidney and hepatic disease (Zhang et al. 2004). The MZ twins do not present a renal phenotype, ergo this variant does not seem a likely candidate.

In contrast to the MZ twins, BBS012 has a variant in *XPNPEP3* (p.P493A), which interestingly localises to the mitochondria of renal cells and causes a NPHP-like ciliopathy. Suppression of *XPNPEP3* exhibits a deleterious effect on the cilium and thus is predicted to have a dual role at the basal body (O'Toole et al. 2010). BBS012 does not present renal defects except for nephrolithiasis, which is not considered a symptom of BBS. Due to the presence of this mutation in the less severe sibling, it could be possible that the variant acts protectively and slows the progression of symptoms like retinal degeneration. *XPNPEP3* is undoubtedly an interesting gene that requires more investigation to determine the link between the primary cilium and mitochondria.

3.3.1.3. Undiscovered modifiers

As described above, there are some fitting candidate modifiers that may explain the phenotypic heterogeneity observed throughout the cohort. However, there were 2 patients who did not present any modifiers, and others exhibited less suitable candidates that could not explain the differences in expressivity. For example, I could not find a genetic explanation for why the MZ twins were discordant for polydactyly. This may be an obvious statement, since MZ twins are supposedly genetically identical. However, this assumption needs to be re-evaluated, as somatic mutations can develop in MZ twins postzygotically (Silva et al. 2011; Youssoufian and Pyeritz 2002). Indeed, somatic mutations can occur during any stage of development and may be tissue specific (Lupski 2010). Therefore, a somatic mutation that caused polydactyly in BBS010 may not be detectable in genomic DNA extracted from peripheral blood.

One of the benefits of whole genome sequencing is that the data are contiguous and areas that are difficult to capture in WES can be sequenced. However, complications with WGS technology still exist. For example, some repetitive regions are challenging to sequence at sufficient depth to pass quality control tests (Dewey et al. 2014). Moreover, WGS also allows investigation of the non-coding DNA accounting for ~98% of the genome. Emerging projects such as ENCODE has revealed that regulatory variation can play a role in explaining disease variability due to mutations within expression quantitative trait loci (eQTLs); regions of the genome shown to influence gene expression (Gerstein et al. 2012; Ward and Kellis 2012). For the purpose of this study, analysis

was restricted to functional coding regions. At present resources for identification of disease-related regulatory variation is limited, making it challenging to determine the functional outcome of the non-coding genome. Identification of eQTLs requires comparison by association in a large sample size. This could not be accomplished in this study due to insufficient numbers.

The gene panel that was utilised contained 73 genes known to exhibit the spectrum of ciliopathies specific to the primary cilium. However, the primary cilium is a vast network of interacting proteins, estimated to feature over 800 genes (van Dam et al. 2017), where novel ciliary gene discovery is a common occurrence (Boldt et al. 2016; Wheway et al. 2015; van Dam et al. 2013). Therefore, it is possible that a modifier may be present in any of the cilia genes not screened in this study. The reason that a larger panel was not screened was because a vast number of variants would have been unrealistic to individually evaluate.

The purpose of this study was to assess whether the phenotypic heterogeneity of BBS is altered by genetic factors. However, expressivity could also be impacted by lifestyle or environmental influences, such as gender, diet, physical activity etc. (Hunter 2005). Undoubtedly, environment can affect the degree of obesity and metabolic syndrome. On the other hand, there are other genetic factors not investigated here that may explain low genotype-phenotype correlations. These include larger variation types that may modify gene expression, such as copy number and structural variation (Lindstrand et al. 2016). Indeed, all 15 patients were assayed for copy number changes using a SNParray. In addition, *in silico* methods of structural variant detection were also applied to the WGS data, using the programmes LUMPY (Layer et al. 2014) and CNVnator (Abyzov et al. 2011). A screen of the resulting data with the ciliopathy gene panel identified no candidate modifying copy number or structural variations.

3.3.2. Future experiments for validation of modifiers

Further probing will be required to determine what affect the candidate modifier mutations have on downstream gene expression and protein function. The effect of secondary modifying variation will need to be resolved through further genetic or functional analysis.

3.3.2.1. Screening in larger cohort

One of the limitations of this experiment was a small sample size, an inevitable feature of studying a rare disease. As a result, it is more challenging to speculate the effect of a potential modifier on the disease phenotype when only found in one patient. However, with a larger cohort, it may be

possible to stratify patients according to identified modifiers and their phenotypes. The utility of this technique was displayed when 15 out of 117 JBTS patients were discovered to manifest retinal dystrophy and renal dysfunction as a result of modifying mutations in *AHI1* (Parisi et al. 2006). A larger experiment would also provide greater power that would be required for statistical testing.

3.3.2.2. Functional validation

In vivo validation method

Zebrafish assays are one model that has shown successful utility for validation of modifiers, such as those in *TTC21B*, *RPGRIP1L,* and *CCDC28B* (Khanna et al. 2009; Zaghloul et al. 2010; Lindstrand et al. 2014; Leitch et al. 2008; Cardenas-Rodriguez et al. 2013). Practicality comes from their number, short gastrulation, and transparent embryos that makes it possible to monitor their development in real time. In addition, gene expression can easily be manipulated through injection of an antisense morpholino oligonucleotide (MO) to suppress a specific gene of interest (Zaghloul and Katsanis 2011). Typically, zebrafish are easier to breed than mice, especially when engineering specific knock-in mutations, which would not be time or cost effective for a large panel of modifying variants. Primary cilia pathways are conserved in zebrafish and the main organ systems affected in BBS are equivalent to mammalian models (Song et al. 2016). The ciliopathy phenotype has also been well characterised, consisting of a down-curved body axis, kidney cysts, gastrulation defects, as well as a shortened and widened body axis (Tobin and Beales 2008; Oishi et al. 2006).

For future validation of candidate modifying variants uncovered in this study, the zebrafish would be a valuable model. Expression of target genes can be suppressed through injection of a specific MO and subsequently rescued with wild type mRNA. To investigate the effect of a modifying mutation, the chosen variant can be incorporated into the rescue mRNA. The consequence of the modifier can be determined by comparing the impact of the mutant and wild type rescue mRNAs. Mutant alleles that rescue the phenotype to the same degree as the wild type mRNA are regarded as benign variants. However, mutant mRNAs that fail or only partially rescue the phenotype are classed as null or hypomorphic variants, respectively. Furthermore, mutants that exacerbate the phenotype further than knock down alone are defined as dominant negative variants. Through this method of knock down and rescue, many gene/modifier combinations can be assayed effectively in a reliable *in vivo* model (Zaghloul et al. 2010; Zaghloul and Katsanis 2011).

In vitro validation methods

Modifier validation in *in vitro* models would also be highly useful. The benefit of *in vitro* over zebrafish experiments is that it is possible to directly visualise cilia via immunohistochemical staining. This can provide greater understanding of the functional role of the modifier, for example, whether it induces mislocalisation of proteins or impaired ciliogenesis. Here knock down is achieved through addition of specific short-hairpin RNAs and rescued by transiently transfecting with wild type or mutant constructs (Davis et al. 2011). The categorisation of the modifier can be assessed under the same criteria as above (i.e. hypomorphic, null, etc).

For a patient-specific understanding of how a modifier may function, induced pluripotent stem cell (iPSC) models could be used. The benefit of patient-derived cells is that the genetic background is the same. Furthermore, tissue specific effects can be explored through differentiation of iPSCs into relevant cell types, such as renal cells (Xia et al. 2013), neurons (Wang et al. 2015), or photoreceptors (Zhong et al. 2014). The effect of the modifier can be determined through mutation correction techniques, such as CRIPSR-Cas9 (Kim et al. 2014).

3.3.3. Mutational Burden Analysis

Mutational burden analysis is the assessment of the number of variants within disease relevant genes for a patient cohort compared to a control group. If the mutational burden hypothesis is true, there should be an enrichment of functional variants in the disease group compared to controls. It is thought that this cumulative burden within a pathway could destabilise the biological system and provoke an enhanced disease phenotype (Gonzaga-Jauregui et al. 2015). As the primary cilium contains an intricate network of tightly regulated protein complexes, this organelle provided a fitting system where this model would be expected to work well.

Two sets of ciliopathy patient cohorts were assessed for mutational burden against controls, featuring patients sequenced by either whole exome (16 patients vs. 25 controls) or whole genome sequencing (14 patients vs. 30 controls). The experimental justification of using two data sets was that these could be combined for a greater statistical power (30 patients vs. 55 controls). These were also analysed separately under identical criteria to detect whether findings were replicable in different patient cohorts. Only nonsynonymous variants predicted pathogenic by SIFT or PolyPhen-2 or CADD>20 were considered at a population frequency less than 1% of ExAC database.

3.3.3.1. Mutational burden associated with PDMs in WES data

The first set of analyses used WES data, and determined that there was a significant enrichment in the mean number of qualifying variants, specifically in ciliopathy cases and not controls. This was demonstrated for 73 ciliopathy genes (p=0.006) and an extended panel of 235 primary cilia genes from the SYSCILIA consortium (p=0.006) (van Dam et al. 2013). A control set of a similarly sized disease gene panel that did not overlap with primary cilia genes was also tested. These 74 genes were associated with the neuropathy CMT, and have previously been affiliated with increased mutational burden in a cohort of 40 CMT patients vs. 5748 control individuals (Gonzaga-Jauregui et al. 2015). For my analysis, the mean number of qualifying variants was consistent between control and patient groups (p=0.678), which suggests that the statistical enrichment of variants was specific to primary cilia genes.

Under an autosomal recessive disease model, it is expected that two segregating deleterious alleles are necessary and sufficient alone to drive disease pathogenicity. 10 out 16 patients had a definitive molecular diagnosis. When these 20 known pathogenic alleles were removed from the analysis, there was no longer a statistical increase in the number of qualifying variants compared to controls (p=0.419). This demonstrated that the significant difference observed was an artificial artefact, contributed to only by the PDMs. This observation was further supported when patients were segregated into those with and without a definitive diagnosis. The undiagnosed patients exhibited a similar average (3.5 variants) to the diagnosed patients when the PDMs were subtracted (3.7 variants). Thus, when compared to the control group, there was no evidence of mutational burden without the effect of the PDMs. This discovery supports previous findings in 31 ciliopathy patients vs. 64 controls (Shaheen et al. 2016). Here, rare (MAF<0.01) and pathogenic (CADD>20) alleles were counted in 89 ciliopathy genes and no mutational load beyond the causal variants was detected.

3.3.3.2. Differences found between WES and WGS burden analysis

The WES patient data was not corroborated by the WGS patient burden analysis. Indeed, the mean number of variants were matched for patients (with PDMs) and controls for the ciliopathy panel (3.7 vs. 3.6, respectively; p=0.863) and less than controls for the SCGSv1 panel (9.2 vs. 9.9, respectively; p=0.522). For both panels, there were significantly fewer variants in patients when PDMs were removed (ciliopathy, p=0.0004; SCGSv1, p=0.006). This is opposite to what would be expected if the burden analysis hypothesis were true. Initially, it was considered that this abnormality could be a technical error caused by discordances across different batches of WGS.

This thought was strengthened by the observation that there were slightly fewer variants in the CMT panel in patients compared to controls. However, small amounts of variation between individuals to a certain degree would be expected, as determined by Fisher's exact test, which suggested that the difference observed was due to chance.

Further investigation into technical error was carried out by counting the mean number of variants at each stage of the filtering analysis. This showed that there were overall more variants in the control data than the patients. However, the large statistical variance illustrated by 95% confidence intervals, suggested that the difference is due to biological variability, rather than technical variability. This finding was investigated by a Student's t-test, which further supported that the variability was due to chance (p=0.078).

One of the advantages of WGS technology is that reads are contiguous, which allows greater accuracy for mapping and uniformity of sequence coverage. Therefore, it is more likely that WGS would outperform WES for reliability across batches, as WES reads can differ in hybridisation efficiency resulting in technical artefacts in low coverage areas. As no abnormalities were observed across the WES data sets, this led to the conclusion that the difference between the WGS patients and controls was biological and not caused by technical artefacts.

3.3.3.3. Combination of datasets

The datasets from WES and WGS were integrated to increase the sample size for a better representation of the BBS population. Where previously there were a similar number of variants in WGS patients (with PDMs) and controls, the combined set increased the number of variants in favour of the patients (ciliopathy panel, p=0.035). Although it could be considered that the integration enhanced the experimental power to statistically distinguish between the patient and control sets, this data should be interpreted with caution. The significance observed may be artificial, which has skewed the analysis due to a strong contrast between WES patients and controls. This demonstrates the importance of analysing the two datasets separately first, for the differences between the WGS and WES datasets would have been masked.

3.3.4. Conclusions of burden analysis

The key biological differences between the WGS and WES patient cohorts were the underlying causal variants. The variants in 10 diagnosed WES patients consisted largely of frameshift, nonsense, or splice site mutations, where 9 out of 11 alleles were not found in Gnomad database

of 138,632 individuals. On the other hand, the WGS cohort were selected for sequencing due to the presence of the common p.M390R allele in *BBS1*, for which 436 out of 277,150 control alleles were reported in Gnomad. With this knowledge, it was surmised that the discordance between the analyses for these 2 datasets is due to the differences in patient genotype. This deduction has led to two conclusions to be established relating to burden testing in BBS. Firstly, mutational burden analysis has shown that there was no enrichment of variants in patients when PDMs were subtracted. Therefore, this result suggests that the number of variants and disease expressivity do not correlate. Secondly, a significant increase in patients (with PDMs) compared to controls was only seen for private or very rare PDMs. Thus, a significant increase was not observed for patients with p.M390R allele as this is a relatively common variant (Beales et al. 2003; Mykytyn et al. 2002). These two conclusions are discussed below.

3.3.4.1. Number of variants does not correlate with disease expressivity

The results indicate that PDMs are the driving factors that lead to a significant enrichment of pathogenic variants in WES patients vs. controls. When the PDMs were removed from the analysis, the mean number of variants resembled the mean of the controls. This suggests that the PDM alone is driving disease expressivity. On the other hand, it has been shown that a single locus mutation does not sufficiently explain the clinical heterogeneity observed in BBS. Studies including this one, have provided evidence of modifying mutations that, to a degree, determine phenotypic expressivity (Cardenas-Rodriguez et al. 2013; Zaghloul et al. 2010; Badano et al. 2006). The burden analysis suggests that modifying mutations do not occur in addition to the background level of variation, but are inherited by chance as part of the background. Contrary to burden analysis of other diseases (Gonzaga-Jauregui et al. 2015; Ji et al. 2016; Girard et al. 2015), evidence suggests that BBS expressivity is driven by a small number of rare deleterious variants, which induce a large detrimental effect. In conclusion, it is not the number of variants that control disease expressivity, but the functional effect that the mutations have at the protein level.

3.3.4.2. Private mutations drive statistical significance in burden testing

Large scale sequencing projects, such as the 1000 genomes or ExAC projects, have determined that there is an abundance of rare and private mutations at a background level in the general population (Xue et al. 2012; Lek et al. 2016). The burden analysis in this study has determined that a level of background variation is present within all investigated individuals, which resides at around 3 background variants per person for the ciliopathy gene panel.

The hypothesis of mutational burden is that there is an accumulation of variants in the patients compared to controls, which drives disease expressivity. Testing in this study determined that there was only a statistical increase of variants in patients when the PDMs were included in the WES analysis and not for WGS patients. The nature of the WES patient's PDMs is that they are private/novel or very rare variants, and predicted to be pathogenic, either by a missense mutation in a highly conserved region or a protein truncating variant. The definition of a private mutation is a variant that is discovered for the first time in a population, but is not necessarily *de novo* (Gao and Keinan 2014). Through negative selection, very deleterious private mutations are typically removed from the population (Gazave et al. 2013). This study has demonstrated that these private mutations have an increased burden on BBS compared to the p.M390R allele, which by comparison is relatively common.

Unlike the WES, the WGS patients exhibited a similar mean number of variants (with PDMs) as the control group. This led to the conclusion that the p.M390R allele was mimicking a background variant, as possession of this mutation did not differentiate the patients from controls. This explains why there was a significant difference when PDMs were removed from the analysis. As the number of variants were similar between the patients and controls, the removal of p.M390R created an artificial bias that led to a statistical difference.

BBS1 p.M390R manifests in up to 30% of all BBS patients (Mykytyn et al. 2003). Individuals with p.M390R are arguably some of the most complex BBS cases. Family members homozygous for p.M390R can express widely different onsets of disease, severity or even be asymptomatic (Beales et al. 2003; Shaheen et al. 2016; Azari et al. 2006; Estrada-Cuzcano et al. 2012). The basis of p.M390R as a background variant relies on its fairly high population frequency in Europeans (MAF=0.3%). The commonality of the allele is supported by the fact that all homozygous patients in this study are from non-related parents; typically, in rare disease cases homozygous variants arise from consanguineous families (Shawky et al. 2013). Studies have suggested that p.M390R resides on an ancient haplotype that has propagated through the European population by genetic drift or a selective advantage in the heterozygous form (Beales et al. 2003; Mykytyn et al. 2003). Perhaps it could act as a beneficial allele in times of famine, as it has been suggested that heterozygous carriers are prone to obesity (Croft et al. 1995). Furthermore, patients with *BBS1* p.M390R mutations exhibit a spectrum of severity, where with modern treatments some individuals are able to raise families. Therefore, there is limited opportunity for the allele to be negatively selected against. Moreover, as p.M390R is an ancient mutation, it may have less

influence on disease susceptibility (Lupski et al. 2011). This strengthens the case that p.M390R patients have modifying mutations that control the penetrance or phenotypic variability, as demonstrated in the first section of this chapter.

3.3.5. Further investigations and experimental limitations

Firstly, it must be considered that a small sample size may influence the statistical testing due to an underpowered study design. Therefore, further investigations must be undertaken with a larger group of patient and control subjects in order to provide greater evidence of the conclusions made in this study.

In order to determine whether the hypothesis that p.M390R mimics a background variant is true, I would need to validate my findings with a larger cohort of patients carrying the p.M390R allele. Another line of investigation would be to determine whether asymptomatic patients exhibit protective variants. For example, asymptomatic carriers may have an enrichment of variants compared to syndromic patients, which safeguard against the onset of BBS. This investigation could be difficult to conduct, as it is unknown how many healthy individuals are harbouring homozygous p.M390R alleles and there are no homozygous variants in control databases, such as Gnomad. Thus, a large-scale sequencing study would be required.

The availability of a larger cohort would strengthen the study design for burden testing as more sophisticated analyses could be completed with greater statistical power (Guo et al. 2016; Ionita-Laza et al. 2011). Studies with large numbers have shown that it is possible to pinpoint the variants that are contributing the most to the statistical significance of the burden test (Lin 2016). This can help identify causative variants in undiagnosed patients or discover potential modifying alleles (Zhao et al. 2016; Lee et al. 2014). Especially for a rare disease, there are practical limitations for undertaking an experiment of this magnitude. In order to complete these analyses, potentially hundreds of case subjects would be required for a heterogeneous disease, such as BBS (Guo et al. 2016).

For this analysis, a relatively simplistic testing method was utilised. Two gene lists were investigated in order to encompass the complex interacting network of the primary cilium. However, it is possible that a large gene set was suppressing some mutational load, which gave the impression that patients have no additional burden compared to controls. Some genes may not be amenable to mutation testing as they have a high genic tolerance i.e. the gene can endure

multiple predicted deleterious mutations but not exhibit an effect (Petrovski et al. 2013). For genes in a mutational burden panel, the challenge is to decide between the biological importance of a gene and its genic tolerance. Furthermore, the true deleterious nature of a variant cannot be ascertained through computational means. In this study, the variants were filtered by functional prediction software, retaining calls predicted deleterious by SIFT or PolyPhen-2 or CADD>20. However, the use of these algorithms would have overestimated the deleterious calls, as variants can be retained if calling methods are discordant. Either manual or computational analysis would be required to accurately isolate deleterious calls by more than one algorithm.

3.3.6. Final conclusion

Overall, the modifier and burden testing experiments suggest that looking only at the number of variants is not sufficient to determine differences in disease expressivity. Within the modifier section of this study, the frequency of identified candidate modifiers did not correlate with disease severity. For example, patient BBS007 is one of the more severe patients in the cohort who presents renal abnormalities. This patient only had 1 qualifying variant. Furthermore, there was no evidence of mutational burden without the presence of the primary driving mutations in patients when compared to controls. Importantly, in order to validate the observations in this study, further testing in larger cohorts, as well as functional assays would be required.

Chapter 4: Utilisation of proteomic profiling techniques for identification of biomarkers in BBS

4.1 Background

4.1.1 Biomarkers

A biomarker is a measurable biological characteristic that can indicate normal physiology, disease state or pharmacological response to treatment (Biomarkers Definitions Working Group. 2001). A clinically informative biomarker can help clinicians make decisions for treatments or facilitate monitoring of disease progression, either to aid diagnosis or in response to therapy. Biomarkers can take many forms, for example, they may be a molecular marker identified from a biosample, a recording such as blood pressure, or an imaging scan. Within this chapter, the focus shall be molecular biomarkers comprised of protein or peptides. Established proteomic biomarkers are commonly identified in accessible biofluids, including plasma (Hathout et al. 2015) and urine (Park et al. 2016).

A common goal is to identify protein biomarkers that can be routinely used in clinics (Rifai et al. 2006). Cancer-specific markers have shown the greatest clinical utility, such as prostate specific antigen (PSA), which is indicative of prostate cancer. PSA is specifically expressed in prostate tissue, but is elevated in serum upon tumorigenesis, where abundance correlates with disease progression (Stamey et al. 1987). Furthermore, elevated levels of plasma CD340 were found to be indicative of an aggressive form of HER2/neu specific breast cancer. This marker can also be used to warrant treatment with trastuzumab, which targets the HER2 receptor (Lüftner et al. 2003). Both of these protein biomarkers are detectable due to increased secretion of proteins from diseased tissue, so that they are abnormally elevated in biofluids. Research studies for a range of non-cancer disease biomarkers have also been described, including Duchenne muscular dystrophy (Hathout et al. 2015), Parkinson disease (Shi et al. 2015), rheumatoid arthritis (Park et al. 2016), diabetes (Riaz 2015; Soggiu et al. 2012) and renal diseases (Rao et al. 2007; Kamińska et al. 2016; Zhao et al. 2015). Identified markers required comprehensive biomarker validation before they could be translated into the clinic.

4.1.1.1 Biomarkers for Bardet-Biedl Syndrome

For BBS patients, DNA sequencing allows identification of the genetic marker of disease. However, patients habitually express phenotypic variability and despite studies, a genotype-phenotype correlation is largely unestablished (Daniels et al. 2012; Deveault et al. 2011; Carmi et al. 1995; Hjortshøj et al. 2010). The mutual finding of these studies is that patients with mutations in *BBS1* are generally less severe than other genotypes. However, heterogeneity exists within the *BBS1* genotype, where patients with the same mutation can manifest different clinical features (Beales et al. 2003).

In clinic, patients are monitored through measurement of markers unspecific to BBS. Currently, circulating C reactive protein (CRP) is the standard marker to monitor disease progression. CRP is an acute phase protein, where plasma abundance correlates with inflammation (Gabay and Kushner 1999). This is used as an indicator of secondary factors to obesity, as low-grade chronic inflammation underpins cardiovascular disease (CVD), insulin resistance, and subsequent type II diabetes (Mirhafez et al. 2016). However, CRP is a generic marker of disease and does not provide information of the consequence of inflammation. Therefore, there is a need for markers that may help predict the onset of phenotypic features, such as renal failure or metabolic syndrome, to allow early treatment or management of disease. Within this study, 11 patients were clinically obese, patients BBS003 and BBS007 were morbidly obese, and BBS004 and BBS015 were overweight. Therefore, within this chapter obesity was the primary phenotypic feature under investigation.

Obesity is a highly prevalent feature of BBS, where an overweight to morbidly obese BMI (25-40+ kg/m^2) was reported to affect 92.5% of patients in a survey containing 105 cases by Deveault et al. (2010). Of these, 17.6% were overweight (25-29.9 kg/m^2), 64.9% were obese (BMI 30-39.9 kg/m^2) and 17.6% morbidly obese (BMI <40 kg/m^2). As a consequence of high adiposity, patients exhibited detrimental secondary features, such as elevated triglyceride levels, high cholesterol, hypertension, cardiovascular abnormalities, diabetes mellitus, and non-alcoholic fatty liver disease (phenotype frequency: 50%, 72.6%, 32.4%, 19.1%, 17.1% and 18% of cases, respectively) (Deveault et al. 2011). These features overlap closely with metabolic syndrome, which is predicted to affect 30-40% of people in the UK by age 65 (Han and Lean 2016). However, BBS subjects are thought to be molecularly distinct from obese subjects; BBS patients are known to exhibit leptin resistance, higher adiposity and greater abdominal visceral fat in comparison to

BMI-matched controls (Feuillan et al. 2011). Taken together, these features can attribute to high morbidity and mortality in BBS patients if improperly managed.

4.1.1.2 Strategy for biomarker identification

The goal of biomarker discovery is to uncover an easily measureable molecule related to a pathophysiological pathway that presents prognostic or diagnostic value. Regardless of many publications describing new biomarkers, only a handful are considered reliable enough for use in clinic (Poste 2011). The path from discovery to validation is a multi-step process (**figure 4.1**), where success lies in correct stratification of patients, independent preclinical validation, verification in large cohorts, and appropriate control groups. The initial discovery stage will typically utilise a relatively small number of samples and may identify thousands of potential markers. At each step of the process, the biomarker specificity and sensitivity is validated. Consequently, the sample size increases and false positive markers are eliminated, leaving a handful of high confidence markers that may be translated into the clinic (Rifai et al. 2006).

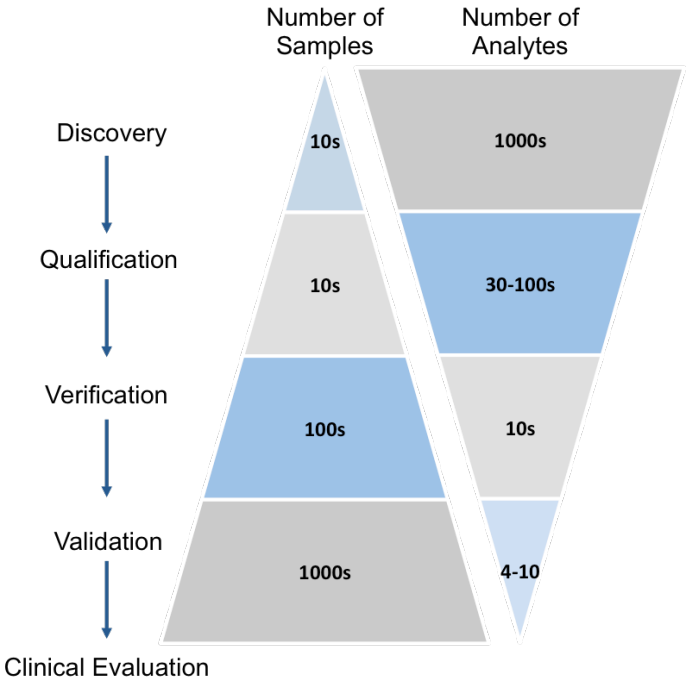

Figure 4.1 – The hierarchy for biomarker discovery (adapted from (Rifai et al. 2006)). The steps of biomarker identification are inversely proportional between the number of samples and detectable analytes.

4.1.2 Proteomic profiling of biofluids

4.1.2.1 Urine

Urine is the product of soluble components of blood filtered by the glomerulus of the kidney. Although the main constituents of urine are water, urea, ions, and creatinine, soluble molecules

such as proteins, hormones and metabolites are also detectable. Urine is an advantageous biofluid for study as it is readily and abundantly available, and can be obtained non-invasively. Furthermore, as it has been filtered by the glomerulus, it has relatively low complexity. Disease associated protein markers have previously been identified for diabetes, renal abnormalities including autosomal dominant polycystic kidney disease (ADPKD), and CVD (Soggiu et al. 2012; Rao et al. 2007; Gonzales et al. 2009; Dear et al. 2013; Pocsfalvi et al. 2015).

Studies have been undertaken to determine the protein composition of urine. Some have utilised the presence of extracellular vesicles (EVs) that are found in urine. These are vesicles ranging from 5-1000 nm in diameter and categorised into subtypes depending on their size; exosomes (50-100 nm), microvesicles (up to 1000 nm) and apoptotic bodies. EVs are used to traffic secreted proteins, lipids, and RNA between cells, and thus provide a valuable insight into intercellular communication (Raposo and Stoorvogel 2013). Furthermore, cells can release EVs under pathological conditions, hence soluble luminal proteins may provide additional insight into disease. A multistep fractionation study culminating in mass spectrometry uncovered that the urine proteome consists of at least 2362 proteins (Kentsis et al. 2009). In a separate investigation, 1132 proteins were discovered to exist as part of a fraction containing only urinal exosomes (Gonzales et al. 2009).

4.1.2.2 Plasma

Isolated plasma is the cell-free component of blood. Its physiological role is to maintain blood pressure and act as a transport system for blood cells and soluble macromolecules. Plasma contains proteins secreted by tissues, typically of sizes less than 45kDa, as larger molecules are filtered by the kidney (Anderson and Anderson 2002). The plasma proteome is comprised of immunoglobulin, hormones, cytokines, lysosomal proteins, and other non-hormone proteins that temporarily travel via the circulatory system. Moreover, plasma may also carry proteins that have leaked or were aberrantly secreted from diseased cells, as well as foreign proteins from infectious organisms (Anderson and Anderson 2002).

In theory, plasma is one of the most promising biofluids for biomarker discovery, due to the fact that proteins can be secreted into the blood by diseased cells (Parker and Borchers 2014). Clinically, it is suitable since it can be obtained through routine procedure. However, proteomic analysis of plasma is a formidable challenge due to the vast range of protein concentrations; the circulating proteome is estimated to cover 12 logs of concentration, where serum albumin

equates to approximately 55% of plasma mass (Anderson and Anderson 2002). Furthermore, proteomic analysis indicates that the remaining proportion of plasma may represent up to 3400 plasma proteins (Gold et al. 2010). Analysis of a highly complex sample complicates the sample preparation, where long depletion and fractionation methods are required for sensitive and reliable analysis. These steps can increase sample variability and decreases throughput (Alshammari et al. 2015).

4.1.3 Proteomic techniques for biomarker identification

Biomarkers may be present at a low abundance and thus a sensitive detection method is required (Hawkridge and Muddiman 2009). Two methods of detection have been utilised in this study; label-free mass spectrometry and targeted proteomic analysis.

4.1.3.1 Ultra-performance liquid chromatography mass spectrometry

For mass spectrometry experiments within this chapter, ultra-performance liquid chromatography (UPLC) coupled to a quadrupole time of flight (Q-ToF) mass spectrometer (Waters Corporation) was utilised for an untargeted, label-free strategy. Quadrupole and ToF are both types of mass analysers, which are coupled together to allow generation of MS/MS spectra by tandem mass spectrometry (MS/MS). This enhances resolution for accurate peptide identification. As mass spectrometry separates analytes by mass and charge, samples are ionised and desolvated upon injection into the instrument. Different variations of ionisation methods have been developed (Ashcroft 1997), but samples for this study were ionised by electrospray ionisation (ESI) for urine analysis, or by matrix assisted laser desorption/ionisation (MALDI) for plasma analysis. The ionisation chambers were coupled to the Synapt G2 Si QToF or QToF Premier mass spectrometers, respectively (both by Waters Corporation).

In UPLC Q-ToF mass spectrometry, the sample is loaded on to the chromatography column where it is subsequently eluted by pH and molecular weight over 45 minutes. The sample flows into the mass spectrometer, where it undergoes ionisation and desolvation by ESI or MALDI. Once charged through the loss of one or more electron, the analyte enters the vacuum system of the first (quadrupole) mass analyser, where the sample is separated by molecular weight and charge. The MS1 spectra is produced by the detector. The tandem MS/MS spectra is generated by a selection of ions from MS1 entering the collision chamber. Here, ions are fragmented by collision induced dissociation (CID), where accelerated neutrally charged atoms, such as helium, collide with analytes to induce bond breakage. These fragments progress to MS2, where once again ions

are separated by mass and charge, and detected on the second mass analyser. Analytes may have different charges, so ions are normalised through calculating the mass to charge ratio (m/z). Typically, MS/MS allows fragmentation of peptides to their amino acid constituents, and thus the spectra can be used to identify peptide sequence (**figure 4.2**).

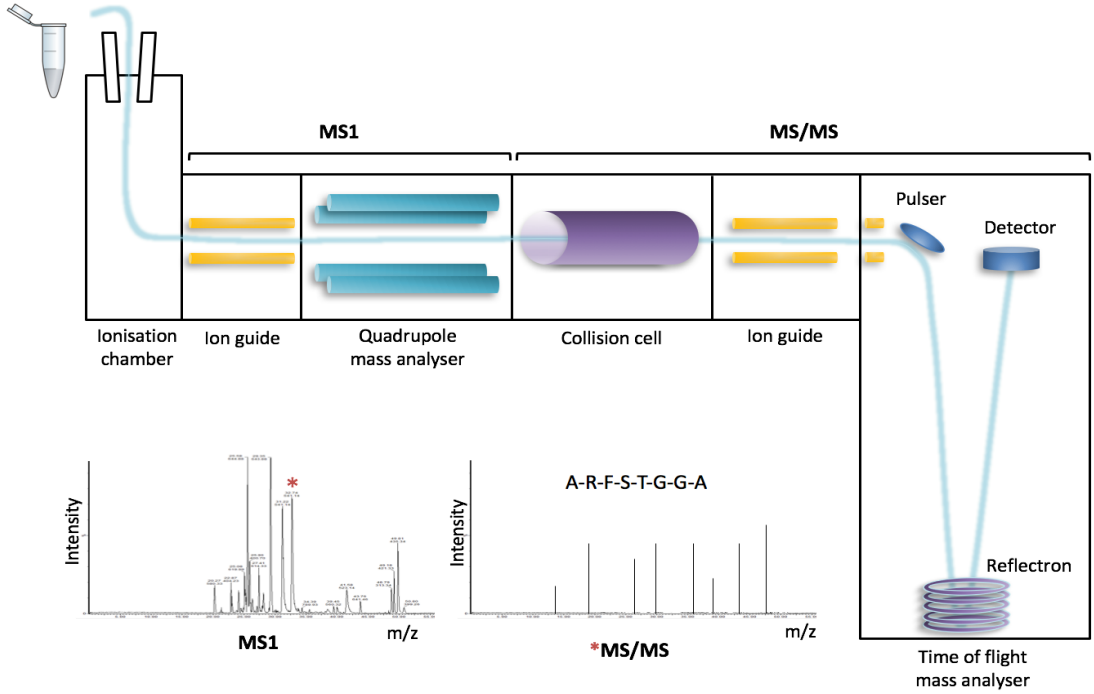

Figure 4.2 – Schematic representation of a quadrupole time of flight mass spectrometer. Sample enters the instrument and is ionised by ESI or MALDI. Peptide ions are separated by mass and charge through the quadrupole mass analyser in MS1 stage. Ions are selected for MS/MS and enter the collision cell for fragmentation. Fragment ions are separated by mass and charge by the ToF mass analyser (MS2), which are detected by the detector. An example MS1 spectrum of a sample (left) and a peptide fragmentation spectrum by MS/MS (indicated by *, right) is also shown, demonstrating how peptide identification can be determined from MS/MS data.

Although MS/MS aids peptide identification, not all peptides are selected to enter the collision cell for fragmentation (Glish and Vachet 2003). As only peptides that are fragmented can be reliably identified, this leads to a misrepresentation of the exact peptide composition of the sample. Recently, MS^E was developed by Waters Corporation, where the mass spectrometer rapidly cycles between high and low energy states during CID. At the low collision energy state the ions are distinctly separated, where upon induction of high collision energy the ions undergo maximal fragmentation. This enables all detectable ions to be fragmented without selection bias and thus increases reproducibility (Plumb et al. 2006). MS^E and high definition MS^E ($HDMS^E$) were used for plasma and urine analysis, respectively within this chapter.

4.1.3.2 Peptide identification and quantification from MS data

Progenesis QI for Proteomics (QIP) (Non-linear Dynamics, Waters Corporation) is computational software for label-free liquid chromatography (LC)-MS proteomics data analysis. It is designed to allow comprehensive processing of raw MS data by normalisation, peptide identification, protein quantitation, and statistical analysis of biological groups within a single programme (Qi et al. 2012). The functionality of Progenesis QIP is reviewed here.

4.1.3.3 Peptide identification

A peptide sequence can be obtained from efficient peptide fragmentation generated by tandem MS. For peptide identification, the ion fragment spectra produced by MS/MS is searched against the human proteome database. The quality of fragmentation is reflected by a peptide identification score. Peptides with high fragmentation efficiency, and thus a clear amino acid read-out, are assigned a high peptide score greater than 8. However, peptides that did not fragment well in the collision chamber receive a lower peptide score (**figure 4.3**).

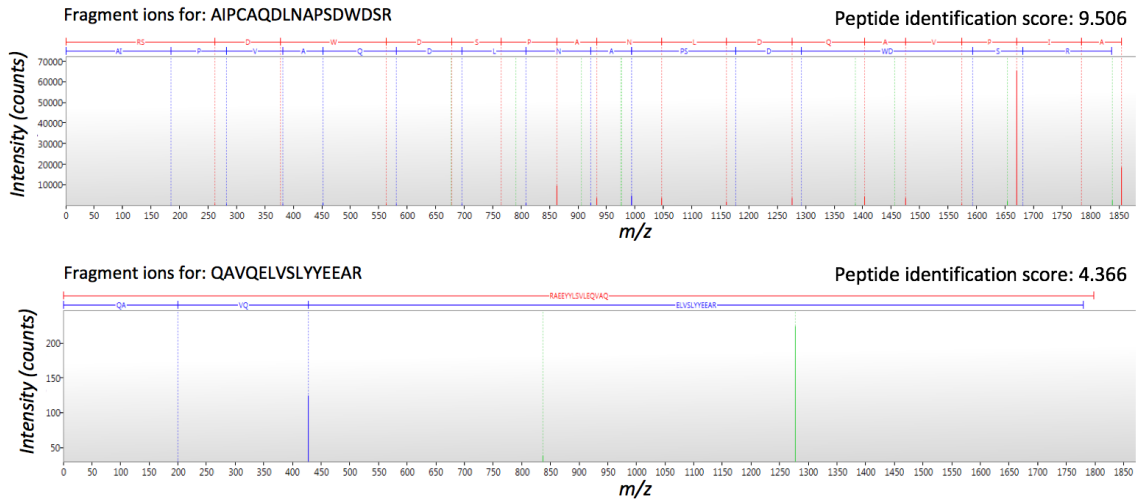

Figure 4.3 – The profile of a well fragmented peptide versus a low efficiency fragmentation profile. From the top peptide, an accurate amino acid sequence can be achieved and a high identification score is granted. The peptide profile at the bottom demonstrates poor fragmentation.

4.1.3.4 Label-free quantitation

The accurate quantification of proteins identified in a sample is crucial for statistical analysis between 2 biological groups. This is challenging since mass spectrometry is not intrinsically quantitative, due to ionisation efficiency differing between one analyte and another (Ho et al. 2003). As a consequence, the intensity peak of an ion is not comparable within the same run, but can be compared relative to the same ion in a different run (Zhu et al. 2010; Wang et al. 2012).

This method is called relative quantification and is used within this chapter for label-free quantitation.

Label-free methods are the least cumbersome approach to protein quantification, which makes them beneficial for large-scale discovery experiments (Griffin et al. 2010). Each sample is prepared and analysed separately without the need for expensive labelling compounds. However, to reduce the variability of this technique, there must be control over sample preparation and instrument stability. As this can be difficult to control, it is not uncommon for the same sample to result in differences in peak intensities (Nilsson et al. 2010). Therefore, spectra normalisation and peak alignment is crucial to ensure reliable results.

A sophisticated approach for peptide quantitation is employed by Progenesis QIP, called ion abundance quantification. This technique averages and normalises unique peptide ion peaks and calculates the abundance from the volume of each peak. The first step requires alignment of spectra to remove variability of retention time acquired for the same ion over different runs (figure 4.4). Consequently, high reproducibility of the instrument is essential for accurate quantification. The abundances are calculated from these aligned runs, only for ions that are unique and shared between runs. Therefore, if the ion is only found once, it cannot be quantified. This approach enhances the rate of protein quantification by up to 40% (Schulze and Usadel 2010). The main disadvantage is the requirement for expensive computational software, and that missing ions cannot be quantified.

4.1.3.5 Challenges of Mass spectrometry

Protein mixtures from biological samples can be highly complex; not only in number, which can range from 2000 proteomics from urine to over 10,000 proteins for cellular mixtures (Kentsis et al. 2009; Geiger et al. 2012), but also due to posttranslational modifications, complexing proteins, isoforms, or variants (Larance and Lamond 2015). Therefore, experimental conditions must be well controlled to ensure that experimental variability is not mistaken for biological variability. Complex mixtures are prone to undersampling, which can be prevented through extensive fractionation prior to profiling.

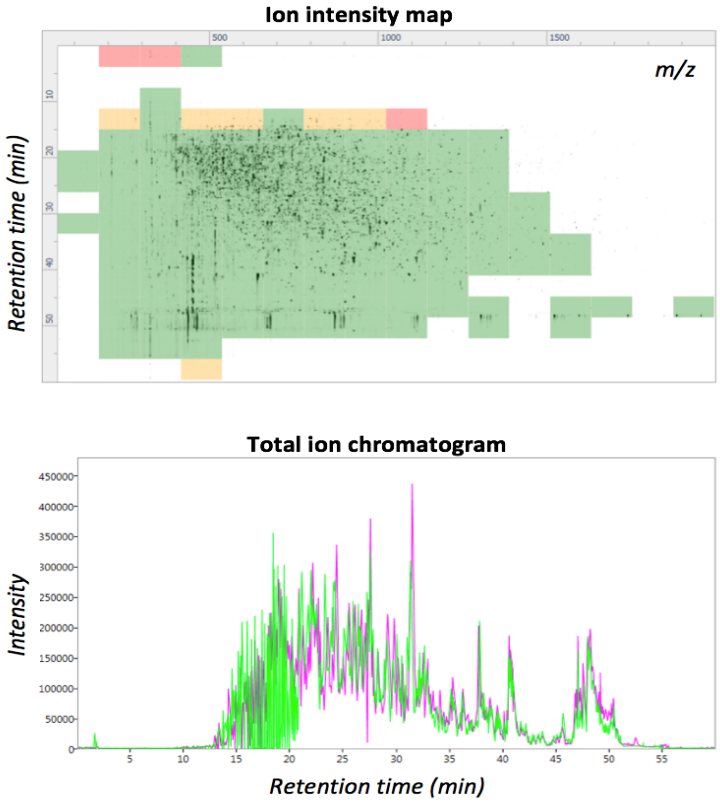

Figure 4.4 – Example of an ion intensity plot from Progenesis QIP, where each spot is a peptide ion plotted by m/z vs. retention time. Colours represent the level of alignment with a chosen reference run: well aligned (green), ok (orange), may need reviewing (red). This example has an alignment score of 94.7%. Ion intensity peaks (bottom) shows the intensity of peptide ions for the reference run (pink) compared to the target run (green). The relative abundance is calculated from the area of each peak.

Undersampling refers to when the analysis is not representative of the entire sample (Nilsson et al. 2010). For example, a mass spectrometer will profile an injected portion (~0.5 µl) of a peptide mixture, which will originate from a larger stock of resuspended peptides. The analysis will only be able to detect analytes present within the 0.5 µl of peptides, where low abundant peptides may not be sampled. This is why some mass spectrometry experiments have low reproducibility rates (Griffin et al. 2010). Undersampling can be combatted through multiple replicates of the same sample until new peptides are no longer discovered, or through reducing sample complexity by fractionation techniques (Jafari et al. 2012).

During mass spectrometry, a peptide mixture is separated by mass and charge. However, if the instrument is overwhelmed by analytes with very similar mass and charge, the resulting spectrum will contain a peak for the most abundant analyte, masking the low abundance ones (Hawkridge and Muddiman 2009). This is why a second form of separation is required prior to mass

spectrometry to reduce the sample complexity. As utilised in this project, ultra-performance liquid chromatography (UPLC) columns are commonly coupled to mass spectrometry for this purpose (Dunn et al. 2011). The column contains a stationary phase (such as C18 alkyl groups) and the sample is eluted through the column using different concentrations of a polar solvent (such as acetonitrile) under high pressure. The mixture is separated based on the chemical affinity with the stationary phase and the solvent mobile phase (Ardrey 2003). Different analytes are then profiled by the mass spectrometer with respect to their retention time (the time it takes for the analyte to be eluted from the column).

Fractionation takes advantage of the different chemical properties of proteins, and can separate by size, affinity, charge or pH. Samples can be fractionated by chromatography 'on-line' (coupled to the mass spectrometer), where different solvents are used to separate analytes by pH (Dunn et al. 2011). Off-line uncoupled fractionation may utilise other chromatography techniques (such as ion exchange or size exclusion chromatography) or gel electrophoresis (Sajic et al. 2015). In this project, proteins were fractionated by SDS-PAGE, which separates proteins by their molecular weight (Jafari et al. 2012). Fractionation can occur either at the protein level or as peptides, where the outcomes will differ depending on which method is chosen.

The quality and sensitivity of the mass spectrometer can be monitored through the course of the run to flag any potential anomalies. A quality control (QC) sample can be prepared from a pool of all samples within the experiment. This QC is then analysed multiple times and at regular intervals throughout the course of the run. Upon analysis, all samples and QCs can be plotted in a principle component space to observe the distribution of data points (**figure 4.5**). As the QC runs are technical replicates, these should cluster on the principle component analysis (PCA) plot. Furthermore, as a pool of all samples, the QC cluster should lie in the centre of the plot, with all other samples distributed around it (**figure 4.5B**). On the other hand, if the QC replicates are scattered on the PCA, this suggests that the mass spectrometer had reduced sensitivity and the data is not reliable for further analysis (**figure 4.5A**).

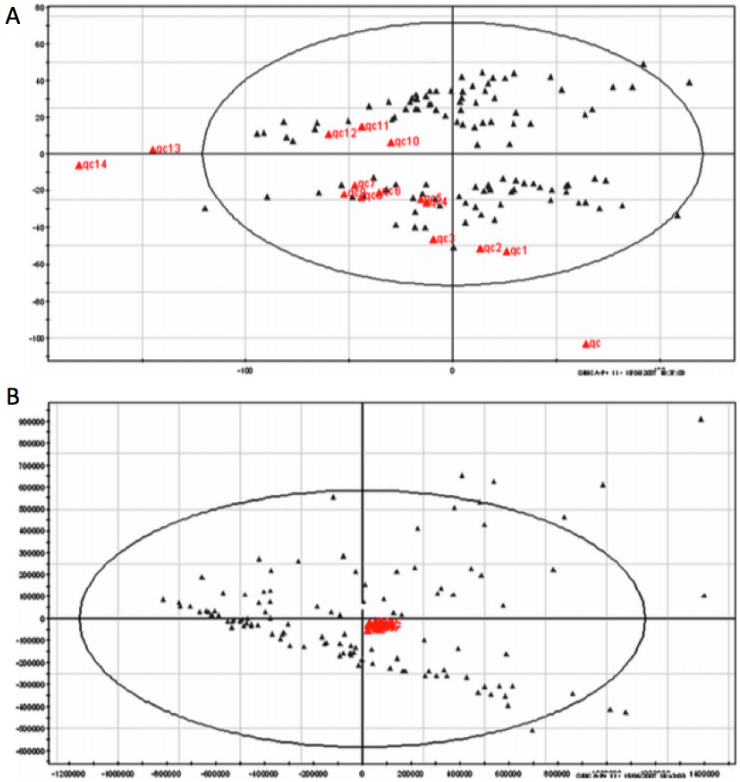

Figure 4.5 – Principle component analysis (PCA) of metabolomic analysis by mass spectrometry taken from Theodoridis et al. 2008 (with permission). Black triangles refer to samples, whereas red triangles are QC samples run at regular intervals throughout the course of the experiment. A) QCs are scattered indicating poor instrument performance and high variability. B) Tight clustering of QCs demonstrates sensitive system performance. The QC cluster is in the centre of the plot as it is a representative pool of all samples.

4.1.3.6 Alternative proteomic technique: SOMAlogic

Inspired by the specificity and sensitivity of ELISAs, SOMAlogic Inc. identified that for a complex proteomic study, large-scale multiplexing of assays is required. However, ELISAs have limited multiplexing ability due to inherent cross-reactivity of antibodies. The answer was found in oligonucleotide comprised three-dimensional molecules, first described over 20 years ago, called aptamers (Gold 1995). The nucleic acid equivalent of monoclonal antibodies, aptamers specifically recognise a protein epitope, which is possible due to nucleic acids innate ability to fold into 3D structures.

Aptamers are derived through an *in vitro* process called systemic evolution of ligands by exponential enrichment (SELEX) and have historically been used for therapeutics and catalysis (Brody and Gold 2000). Aptamers are comprised of RNA secondary structures and single stranded DNA, which can be individually engineered to specifically complex with respective protein targets. SOMAlogic have modified the aptamer selection process by generating slow off-rate modified

aptamers (referred to as SOMAmers), which have a greater affinity for the target than typical aptamers (Lollo et al. 2014). Specificity to the target can be altered through modifications of deoxyuridine bases, controlled by additions of benzyl, napthyl, tryptophan or isobutyl groups, which mimic amino acid side chains (Gold et al. 2010). As a result of years of optimisation, SOMAlogic have been able to develop aptamers that strongly associate with 1310 proteins and due to no cross-reactivity, these can be multiplexed to detect and quantify these proteins in one high-throughput assay, namely SOMAscan.

An advantage of SOMAscan is that it utilises robust nucleic acid hybridisation properties, and employs the same chemistry as a microarray (Gold et al. 2010). The oligonucleotide SOMAmer is able to distinctly fold into a unique structure, which specifically binds to the target protein. Through simple binding and washing steps, the target is cleaved and the remaining unique SOMAmer is hybridised to a complementary probe library. The output is a fluorescent signal that quantitatively reflects the protein constituents of the original sample (Kraemer et al. 2011).

The SOMAscan array is highly reproducible, sensitive, low cost and is high throughput with a fast turnaround. Unlike mass spectrometry, the SOMAscan is not limited by protein abundance; and can detect over 8 logs of concentration in one sample (Lollo et al. 2014). This results in a simpler sample preparation protocol, as fractionation steps are not required.

SOMAscan is limited only by its optimisation and library of SOMAmer probes. At the time of writing, the SOMAscan was optimised for detection of plasma proteins, but the scope is vast for optimisation of SOMAmers distinct to many different tissue-specific proteins (Gold et al. 2010). The SOMAmers are not designed to discriminate between posttranslational modifications (PTMs). PTMs could either block the epitope or induce a conformational change in the protein, which would compromise SOMAmer binding and prevent protein detection (Hathout et al. 2015). Similarly, proteins cannot be detected if the epitope is blocked by a complexing protein or a cofactor. Small linear polypeptides, typically with a molecular weight less than 5kDa, may not be detected by SOMAmers, since these oligonucleotides are designed to recognise globular proteins. Finally, difficulty may be found when distinguishing between close homologues, including those generated by alternative splicing, if the epitope is identical between proteins (Lollo et al. 2014). These weaknesses observed by SOMAscan are strengths for mass spectrometry.

4.1.4 Aims of investigation

Biomarkers can be beneficial for both patient and clinician to encourage changes in lifestyle or to aid administration of treatment. Currently, there are no specific molecular biomarkers to measure disease progression or detect secondary features of BBS. The current biomarker used in clinic is CRP, a generic indictor of inflammation. The aim of this chapter was to profile plasma and urine biofluids from BBS patients to uncover putative markers specific to BBS, or specific to phenotypic features of BBS. This was achieved through use of proteomic untargeted and targeted techniques. Furthermore, statistical analysis was utilised to determine if differentially expressed proteins were present between 15 BBS patients and 6 healthy non-obese controls. Although this was a small sample size, this is suitable for preliminary investigations to identify potential markers (Rifai et al. 2006). Importantly, significantly differentially expressed candidate markers must be validated in future investigations.

BBS patients are phenotypically heterogeneous, of which the patients in this cohort are no exception. For statistical comparison, ideally the biological groups need to be as homogeneous as possible. There are two phenotypes that are most prevalent in this study's cohort: retinal dystrophy (affecting 15/15 patients) and obesity (affecting 13/15 patients). In addition to obesity, some patients exhibited secondary features, such as type II diabetes in patient BBS002 and hypercholesterolemia in BBS006, as well as factors associated with CVD, including hypertension in BBS002, BBS003 and BBS005, angina pectoris in BBS005, and aortic valve abnormalities in BBS013 (**full patient details are available in table 2.9**). Therefore, comparisons between patient and control were conducted with a view to identify markers related to obesity and secondary phenotypes associated with obesity.

Additionally, the aim was to evaluate different proteomic techniques for use within future multiOmic studies. Protein from urine was extracted by ultracentrifugation before analysis by UPLC-QToF using the Synapt G2 Si with $HDMS^E$ capability. Method development was completed with plasma for analysis by SDS-PAGE and UPLC-MALDI-QToF with MS^E capability. Plasma was further profiled through a novel targeted proteomic technology, SOMAscan. 726, 222 and 1243 proteins were quantified from each technique, respectively. Statistical analysis identified differentially expressed proteins between patient and control groups, where PEDF, ORM1, leptin and ApoM may be of relevance to metabolic features of BBS. Finally, within this chapter it is discussed how putative markers may be validated using larger BBS and control cohorts.

4.2 Results

4.2.1 Label-free UPLC-MS/MS of urine

4.2.1.1 Quantitation of protein from urine

To ascertain the same amount of protein from each sample for mass spectrometry, protein quantitation by bicinchoninic acid (BCA) assay was completed. Protein absorbance was measured at 595nm in duplicate alongside bovine serum albumin (BSA) standards, which ranged from 125 to 2000ng/µl. Using the equation of the standard curve (**figure 4.6**), the average absorbance of each sample was used to calculate respective sample concentration and the volume required for 50µg of protein for further processing (**table 4.1**).

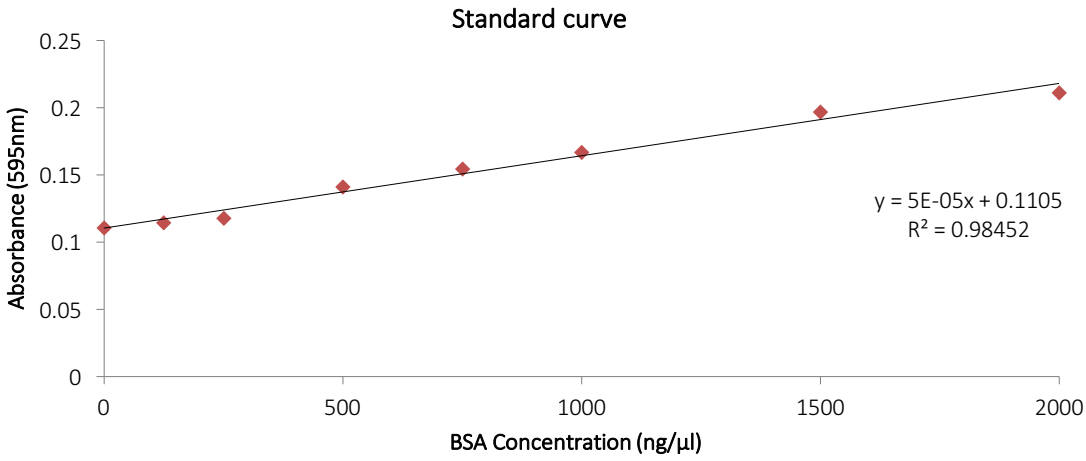

Figure 4.6 – Standard curve generated for quantitation of protein extracted from urine.

Progenesis QIP software version 2 (Nonlinear Dynamics, Waters Corporation) was utilised to analyse the HPLC-MS/MS experiment featuring 25 samples. These samples consisted of 15 BBS patients, 6 controls and 4 quality control (QC) technical replicates. The QC sample was comprised of a pool of all samples and was used to determine if instrument sensitivity was maintained throughout the course of the run. The first stage of the analysis pipeline required alignment of MS1 ion intensity maps to a reference run. The aim of this alignment was to minimise any variation between retention times for the same ion caused by instrument fluctuations. QC1 was chosen by the software as the best reference run, as this contained the majority of peptide ions represented across the experiment. Alignment scores were generated by Progenesis QIP, where scores greater than 80% are considered suitable for further analysis. High alignment scores ranging from 85.3% to 95.6% (mean score= 93.1%, n= 25) were produced, indicating good quality alignment.

Runs were next aggregated and normalised into a single map of peptide ion peaks. Normalisation aimed to reduce differences caused by sample injection and ionisation variability. Finally, before identification of peptides, filters were applied to remove unreliable features from the normalised runs. Ions with a retention time less than 10 minutes and greater than 45 minutes were removed from further analysis, as these ions were indicative of a salt front or other contaminants. In general, peptides are expected to have a charge state between 1 and 4, therefore ions with a charge greater than 10 were excluded. High charge states suggest the presence of large peptides, which are not suitable for mass spectrometry and thus would negatively impact the analysis.

Sample ID	Average absorbance	Concentration (µg/µl)	Volume for 50µg (µl)
BBS001	0.25	2.8	9.0
BBS002	0.26	3.0	8.5
BBS003	0.22	2.1	11.9
BBS004	0.18	1.3	19.2
BBS005	0.14	0.6	41.9
BBS006	0.28	3.3	7.5
BBS007	0.27	3.1	8.0
BBS008	0.23	2.3	10.8
BBS009	0.22	2.1	11.6
BBS010	0.18	1.4	18.0
BBS011	0.24	2.6	9.6
BBS012	0.21	1.9	13.0
BBS013	0.14	0.5	48.8
BBS014	0.15	0.7	35.2
BBS015	0.17	1.2	21.6
CTRL019	0.18	1.5	17.1
CTRL021	0.19	1.5	16.1
CTRL023	0.20	1.9	13.3
CTRL024	0.17	1.2	20.3
CTRL025	0.17	1.3	19.8
CTRL026	0.24	2.5	10.0

Table 4.1 – Values from sample quantitation after BCA assay of proteins extracted from urine and corresponding volume of protein required for downstream processing for mass spectrometry.

4.2.1.2 Identification of peptides and inferring proteins

For peptide identification, an MS^E search was conducted by Progenesis QIP software against a fasta file containing 20,186 reference human proteome sequences. In total, 8642 peptide ions were identified across all samples. Peptides were refined further by removing corresponding peptides with a sequence length less than 5, leaving 7901 peptide identifications remaining. These peptides were then used to infer protein identifications from MS/MS spectra and quantify proteins by relative quantitation. Overall, 815 proteins were identified. However, only proteins

with at least one unique peptide could be confidently called and quantified, thus 86 proteins were removed. In these cases, identified peptides aligned with other homologous proteins, so the exact specification of the protein could not be certain. The remaining 726 quantifiable proteins were identified through detection of 5090 peptides, of which 448 proteins contained 3 or more unique peptides, and 152 were identifiable by a single unique peptide (supplemental data, **table S.29**).

A confidence score was generated for each protein, corresponding to the cumulative sum of each peptide score that made up the protein. The peptide score refers to the fragmentation efficiency and confidence of peptide identification. Generally, proteins with a greater number of peptides had a larger confidence score, regardless of the number of unique peptides. In this experiment, protein scores ranged from 3.35 to 962.62. 212 proteins were flagged as having a confidence score less than 20. This score is recommended as a helpful guide, but should not be depended on. The confidence is negatively affected by spectra quality, where low coverage can lead to spurious calls. Additionally, the score is influenced by protein length and abundance.

4.2.1.3 Quantification metrics

Progenesis QIP software provided quality control metrics to monitor the presence of experimental anomalies. The graphs shown in **figure 4.7** were generated from each sample's MS/MS spectra and illustrate that individually there was a large range in peptide and protein abundances over the experiment. Samples BBS004, BBS005, BBS008, BBS015 and CTRL026 had a consistently low count for all metrics. This is indicative of poor identification data from the MS/MS spectra. The samples at the upper limit of the box plots, namely BBS006, BBS007, BBS010, BBS011, BBS012, BBS014 and CTRL021, had good quality identification data.

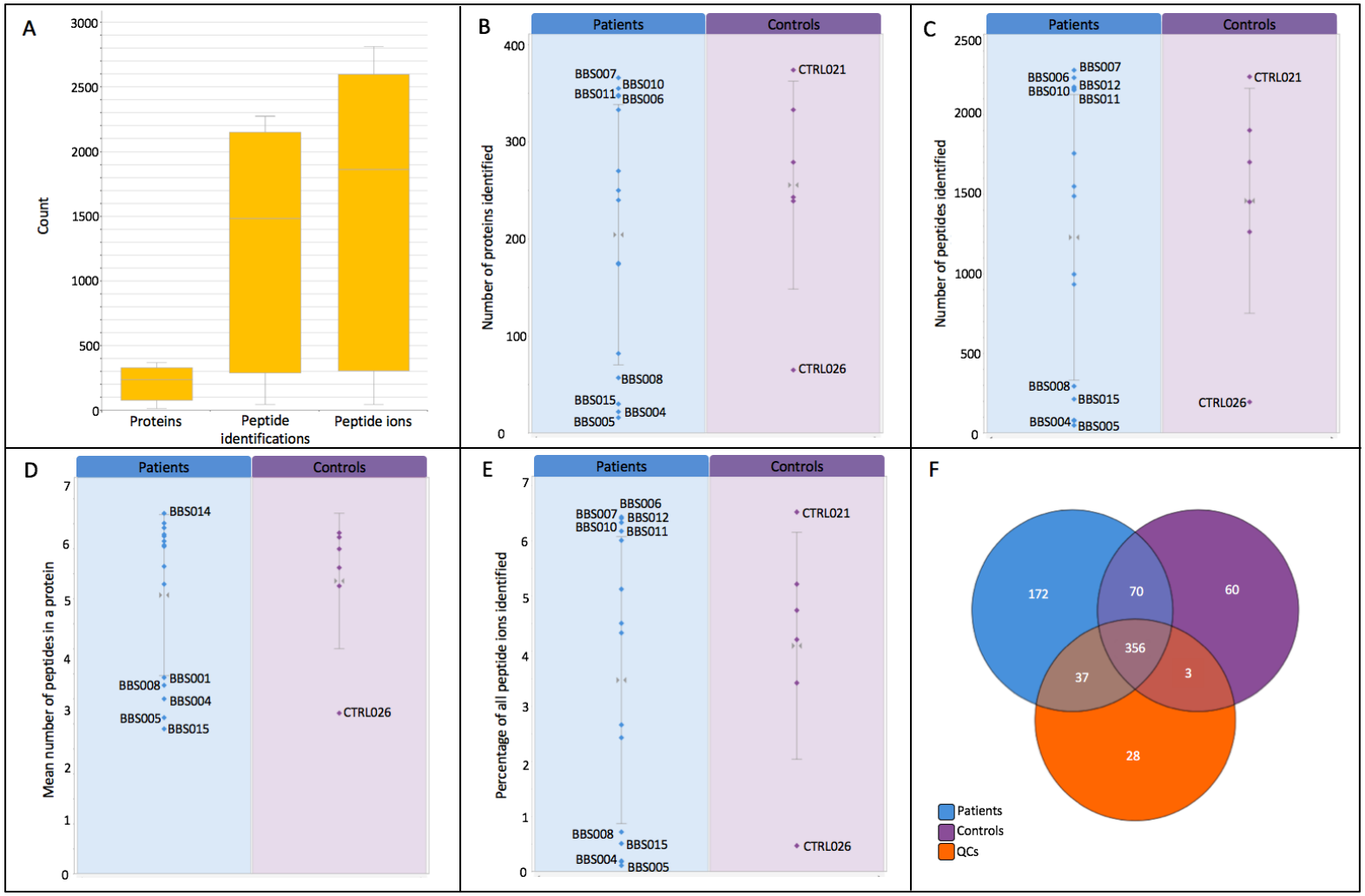

Figure 4.7 – Urine proteomics QC metrics read-outs (figure legend on next page).

Figure 4.7 (on previous page) – QC metric read-outs from Progenesis QIP for urine label-free HPLC-MS/MS analysis, inclusive of quantifiable proteins without duplications. A) Box plots for aggregated data sets (excluding QCs), displaying the spread of peptide ions, peptide identifications and subsequent quantifiable proteins. B) Number of proteins identified across patient and control samples (patient mean = 204 (70,228); control mean = 256 (148, 363). C) Number of peptides identified across patient and control samples (patient mean = 1230 (333, 2126); control mean = 1459 (752, 2165). D) Mean number of peptides identified in a protein for patient and control samples (patient mean = 5.1 (3.8, 6.7); control mean = 5.5 (4.1, 6.8). E) Percentage of peptide ions for each patient and control out of the total number of identified peptide ions in the experiment (patient mean = 3.5 (0.9, 6.1); control mean = 4.1 (2.0, 6.2). Outlying samples for high or low counts have been labelled. F) Venn diagrams illustrating the number of overlapping protein identifications between each biological group and QCs

These graphs help illustrate the purpose of MS1 spectra alignment and normalisation by Progenesis QIP for quantitation. If the MS/MS spectra is poor quality, or even was not generated at all, it does not prevent Progenesis QIP from accurately quantifying the ions from MS1 data. The identity of each ion can be found from the other samples run in the same experiment. Therefore, although it is important to identify anomalies in this experiment, these QC metric results would not have affected the final quantitation results. The Venn diagrams (**figure 4.7F**) display the total count of proteins acquired through unique peptide identifications from MS/MS spectra. This provides a visual overview of which biological groups the 726 quantifiable proteins originated from. The quantitation with MS1 spectra then aimed to identify and quantify all 726 proteins across all samples. 426 proteins were found in both patient and control groups, 70 of which were not also shared by the QCs. This may be due to undersampling or the increased complexity of the QCs compared to individual samples. On the other hand, 28 proteins were found in the QCs and not in either biological group

Proteins were quantified from the normalised MS1 ion spectra using relative label-free ion abundance quantitation. The abundance of 726 proteins was detected over 5 magnitudes of concentration (**figure 4.8**). To display the distribution of abundance, the mean abundance for all 726 quantifiable proteins was calculated and binned into groups. A normal distribution was observed when plotted on a log10 scale. Only 12 proteins were detected at a relative abundance less than 100. This may illustrate that low abundance proteins were not identifiable, either due to instrument sensitivity, sample complexity, or methodological error.

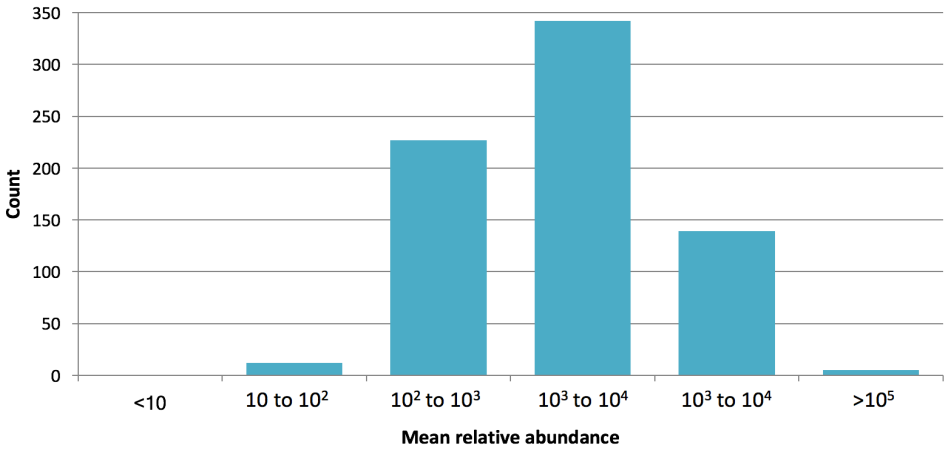

Figure 4.8 – Histogram of mean relative abundances for quantified proteins from urine HPLC-MS/MS.

4.2.1.4 Statistical testing and identification of differentially expressed proteins

Quantified data was plotted in a principle component space to identify sample clustering and outliers. As technical replicates consisting of a sample pool, the QCs helped to determine whether instrument sensitivity was maintained throughout the course of the experiment. Here, all 4 QCs clustered together (**figure 4.9A**). Because they grouped in the centre of the plot, this indicated that they were a good representation of all samples. The QCs were then removed to focus on the relationship between patient and control data points. Upon first observation, patient BBS004 was noted to outlie from the main plot. At this stage, it could not be established the reason for this outlier, as it may be due to a technical artefact or a biologically interesting finding. BBS004 was therefore removed from the PCA to observe the difference (**figure 4.9B**). As a result, there was an increase in spread of the patient data points, but no inherent structure observed in the PCA.

Using ANOVA statistical testing, p-values were generated to investigate differentially expressed (DE) proteins between patient and control groups. P-values were adjusted to account for multiple testing and a false discovery rate (FDR) for each protein was generated. Furthermore, relative quantities were transformed to log2 format (log2RQ) and 95% confidence intervals (CI) were calculated to observe the spread of the data between patient and controls.

When patient BBS004 was included, a single protein was significant after multiple testing, namely alpha-1 acid glycoprotein 1 (BBS mean log2RQ = 16.46, 95% CI=15.95, 16.98; control mean log2RQ = 14.15, 95% CI = 13.31, 15.00; log2 fold change (FC) = 2.31; FDR = 0.039). The top 10 differentially expressed proteins, ranked by FDR, are displayed in **table 4.2.**

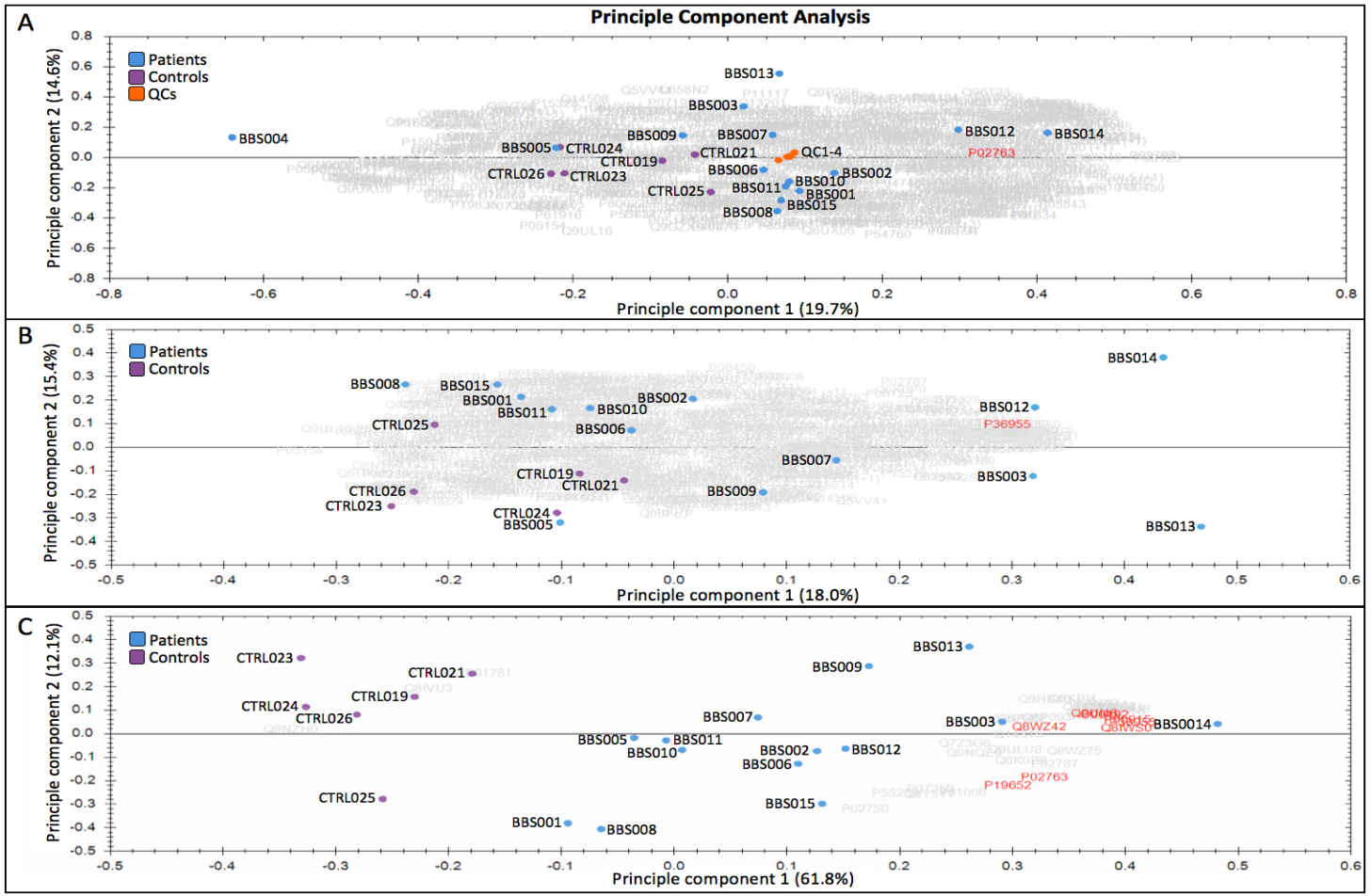

Figure 4.9 – Principle component analysis (PCA) plots generated from protein quantitation data obtained from urine HPLC-MS/MS analysis, where patient and control samples are blue and purple, respectively. The grey cloud represents where each protein (identifiable by UniProt ID) lies on the PCA. The red highlighted proteins are the most significant within each plot. A) PCA with all samples, including QCs (orange). B) PCA without BBS004 and QCs removed from analysis. C) PCA of significant proteins with FDR<0.1, where red highlighted proteins have FDR<0.05.

Gene name	Protein name	Peptide count	Unique peptides	Score	BBS mean	Control mean	FDR	Log2FC
ORM1	Alpha-1 acid glycoprotein 1	42	24	204.4	16.46 (15.95, 16.98)	14.15 (13.31, 15.00)	0.039	2.31
ORM2	Alpha-1 acid glycoprotein 2	33	19	211.1	15.13 (14.57, 15.69)	13.01 (12.34, 13.68)	0.061	2.12
ARHGAP35	Rho GTPase-activating protein 35	4	3	23.4	11.96 (11.47, 12.46)	10.53 (10.30, 10.77)	0.262	1.43
LYVE1	Lymphatic vessel endothelial hyaluronic acid receptor 1	19	17	121.7	12.70 (12.20, 13.19)	11.20 (10.48, 11.92)	0.262	1.50
SERPINC1	Antithrombin-III	21	16	137.9	11.44 (10.98, 11.91)	10.11 (9.74, 10.48)	0.262	1.33
DDX55	ATP-dependent RNA helicase DDX55	4	1	18.6	13.57 (12.98, 14.16)	12.03 (11.78, 12.28)	0.262	1.54
	Ig heavy chain V-III region GAL	6	4	56.0	12.85 (12.52, 13.18)	13.88 (13.44, 14.31)	0.262	-1.03
DNAI1	Dynein intermediate chain 2, axonemal	2	1	9.5	12.43 (11.85, 13.01)	10.89 (10.27, 11.52)	0.262	1.53
GP2	Pancreatic secretory granule membrane major glycoprotein GP2	1	1	3.4	6.26 (5.52, 6.99)	3.09 (0.58, 5.60)	0.262	2.97
GPRC5B	G-protein couple receptor family C group 5 member B	6	1	38.0	8.68 (7.98, 9.38)	10.49 (10.00, 10.97)	0.262	-1.80

Table 4.2 – Top 10 DE proteins ranked by FDR between patients and controls (including BBS004) from urine. Sample mean for BBS and control are presented, where data is expressed as log2(relative quantitation (RQ)) for sample mean (95% confidence interval (CI)). Peptide count refers to the total number of peptides identified for each protein, whereas unique peptides indicates the number of peptides that are distinct to each finding. 'Score' refers to the confidence score read out from Progenesis QIP.

Upon removal of BBS004 from the analysis, a greater difference in protein expression between patient and control became apparent. 8 proteins were significantly differentially expressed with an FDR<0.05. An additional 31 proteins were significant with a more relaxed FDR <0.1 (**figure 4.9C and table 4.3**). Proteins were mapped on a volcano plot by log2(FC) and -log10(FDR) to observed the distribution of differentially expressed proteins (**figure 4.10**).

Gene name	Protein name	Peptide count	Unique peptides	Score	BBS mean	Control mean	Log2FC	FDR	Relationship to disease
SERPINF1	Pigment epithelium-derived factor	7	2	32.3	9.36 (8.86, 9.85)	7.03 (6.60, 7.46)	2.56	0.015	*Metabolic syndrome (Yamagishi et al., 2006)*; Microvascular complications of diabetes (OMIM:603933)
ORM1	Alpha-1-acid glycoprotein 1	42	24	204.4	16.56 (16.05, 17.08)	14.15 (13.31, 15.00)	2.36	0.029	*Cardiovascular disesase and diabetes (Christiansen et al., 2002; Hou et al., 2014)*
ORM2	Alpha-1-acid glycoprotein 2	33	19	211.1	15.23 (14.66, 15.79)	13.01 (12.34, 13.68)	2.35	0.042	Rheumatoid arthritis (Park et al., 2016); Colorectal cancer (Zhang et al., 2012)
CA1	Carbonic anhydrase 1	5	3	29.6	12.41 (11.89, 12.94)	10.60 (10.36, 10.84)	2.19	0.042	*Microvascular complications of diabetes (OMIM:603933)*
ARVCF	Armadillo repeat protein deleted in velo-cardio-facial (VCF) syndrome	2	1	9.1	9.23 (8.69, 9.77)	7.13 (6.39, 7.88)	2.24	0.042	
TTN	Titin	91	36	367.3	13.96 (13.71, 14.21)	13.05 (12.80, 13.30)	0.96	0.042	*Cardiomyopathy (OMIM: 604145, 613765)*
DNAI1	Dynein intermediate chain 1, axonemal	2	1	9.4	12.61 (12.16, 13.09)	10.89 (10.27, 11.52)	1.83	0.047	Ciliary dyskinesia (OMIM: 244400)
PHF6	PHD finger protein 6	6	4	50.1	13.69 (13.29, 14.10)	12.30 (11.96, 12.65)	1.55	0.047	Borjeson-Forssman-Lehmann syndrome (OMIM: 301900)
GPRC5B	G-protein coupled receptor family C group 5 member B	6	1	38.0	8.49 (7.86, 9.13)	10.49 (10.00, 10.97)	-1.72	0.064	Type II diabetes (Soni et al. 2013)
PCDH10	Protocadherin-10	3	2	12.1	15.38 (14.90, 15.86)	13.88 (13.47, 14.29)	1.70	0.064	
RNF123	E3 ubiquitin-protein ligase	2	2	17.7	12.51 (11.77, 13.24)	9.94 (8.77, 11.11	2.72	0.064	
TF	Serotransferrin	86	67	686.7	14.60 (14.15, 15.06)	13.23 (12.89, 13.57)	1.59	0.064	Atransferrinemia (OMIM:209300); *type II diabetes (Riaz, 2015; Soggiu et al., 2012)*
HMX3	Homeobox protein	3	3	15.8	11.26 (10.54, 11.98)	9.08 (8.55, 9.62)	2.84	0.064	

GP1BA	Platelet glycoprotein Ib alpha chain	3	2	13.3	7.99 (7.40, 8.57)	5.80 (4.61, 6.99)	2.21	0.064	Bernard-Soulier syndrome (OMIM: 231200, 153670); von Willebrand disease (OMIM: 177820)
ASAH1	Acid ceramidase	16	10	101.37	16.39 (16.04, 16.75)	15.34 (15.14, 15.55)	1.19	0.064	Farber lipogranulomatosis (OMIM: 228000); Spinal muscular atrophy with progressive myoclonic epilepsy (OMIM: 159950)
POLD1	DNA polymerase delta catalytic subunit	2	1	8.36	15.50 (15.03, 15.97)	14.13 (13.90, 14.36)	1.60	0.064	Mandibular hypoplasia, deafness, progeroid features and lipodystrophy syndrome (OMIM: 615381)
HSPA12B	Heat shock 70 kDa protein 12B	11	6	49.70	13.81 (13.16, 14.45)	11.93 (11.64, 12.22)	2.37	0.064	*Type II diabetes (Soggiu et al., 2012)*
SERPINC1	Antithrombin-III	21	16	137.87	11.51 (11.04, 11.99)	10.11 (9.74, 10.48)	1.63	0.066	Thrombophilia due to antithrombin III deficiency (OMIM: 613118)
SCPEP1	Retinoid-inducible serine carboxypeptidase	5	5	26.31	10.71 (9.94, 11.49)	8.16 (6.93, 9.38)	2.52	0.068	
LYVE1	Lymphatic vessel endothelial hyaluronic acid receptor 1	19	17	121.72	12.79 (12.30, 13.28)	11.20 (10.48, 11.92)	1.61	0.068	
ROBO4	Roundabout homolog 4	28	18	182.39	15.07 (14.74, 15.39)	14.04 (13.59, 14.49)	1.06	0.068	
	Ig heavy chain V-III region	6	4	56.04	12.94 (12.64, 13.24)	13.88 (13.44, 14.31)	-0.92	0.076	
LRG1	Leucine-rich alpha-2-glycoprotein	17	13	133.77	12.36 (11.33, 13.38)	8.99 (7.20, 10.78)	2.50	0.077	*Type II diabetes (Soggiu et al., 2012)*
GP2	Pancreatic secretory granule membrane major glycoprotein	1	1	3.38	6.41 (5.69, 7.13)	3.09 (0.58, 5.60)	0.85	0.078	
DDX55	ATP-dependent RNA helicase	4	1	18.59	13.66 (13.05, 14.27)	12.03 (11.78, 12.28)	2.04	0.078	
ARHGAP35	Rho GTPase-activating protein 35	4	3	23.35	11.93 (11.41, 12.45)	10.53 (10.30, 10.77)	1.71	0.078	
AADACL2	Arylacetamide deacetylase-like 2	2	1	7.95	11.35 (10.42, 12.29)	8.62 (7.43, 9.80)	2.66	0.080	
CDH2	Cadherin-2	22	14	149.92	13.7 (13.09, 14.38)	11.97 (11.41, 12.52)	2.30	0.080	

BSPH1	Binder of sperm protein homolog 1	1	1	10.49	12.31 (11.52, 13.10)	9.92 (8.75, 11.09)	2.62	0.080	
TSPAN14	Tetraspanin-14	2	1	12.13	15.12 (14.71, 15.54)	14.02 (13.72, 14.32)	1.26	0.085	
CADPS	Calcium-dependent secretion activator 1	5	2	25.23	10.28 (9.78, 10.79)	9.00 (8.86, 9.13)	1.52	0.085	
MAN2B2	Epididymis-specific alpha-mannosidase	14	8	80.36	12.55 (12.09, 13.01)	11.33 (10.96, 11.70)	1.41	0.085	
FLNC	Filamin-C	12	6	63.05	12.94 (12.66, 13.22)	12.10 (11.66, 12.53)	0.84	0.085	*Cardiomyopathy/ myopathy (OMIM: 617047, 614065, 609524)*
PYHIN1	Pyrin and HIN domain-containing protein 1	2	2	11.63	10.00 (9.64, 10.36)	9.03 (8.68, 9.39)	1.10	0.087	
HINT3	Histidine triad nucleotide-binding protein 3	2	1	7.77	6.45 (5.22, 7.67)	1.57 (-2.11, 5.25)	1.77	0.091	
LRRC8E	Volume-regulated anion channel subunit	2	2	9.09	9.58 (8.74, 10.41)	7.51 (7.19, 7.82)	3.22	0.094	
FAHD1	Acylpyruvase, mitochondrial	1	1	6.33	14.93 (14.43, 15.42)	13.68 (13.37, 13.98)	1.51	0.094	
PRICKLE2	Prickle-like protein 2	5	1	19.10	9.08 (8.48, 9.68)	7.62 (6.10, 8.42)	1.65	0.097	
LPA	Apolipoprotein(A)	13	8	69.24	13.21 (12.76, 13.67)	14.36 (14.02, 14.69)	-0.95	0.097	LPA deficiency, Coronary artery disease (OMIM: 152200); *Type II diabetes* (Ruotolo et al. 1991)
HERC6	Probable E3 ubiquitin-protein ligase	6	3	25.35	7.79 (7.44, 8.13)	9.22 (8.03, 10.42)	-1.92	0.097	

Table 4.3 – DE proteins between patients and controls (excluding BBS004) with FDR<0.1 from urine HPLC-MS/MS experiment. Sample mean for BBS and control are presented, where data is expressed as log2(relative quantitation(RQ)) for sample mean (95% CI). Where applicable, an established link to disease has been noted, obtained from literature mining or from the Online Mendelian Inheritance in Man (OMIM) database. Diseases related to metabolic syndrome phenotypes are italicised.

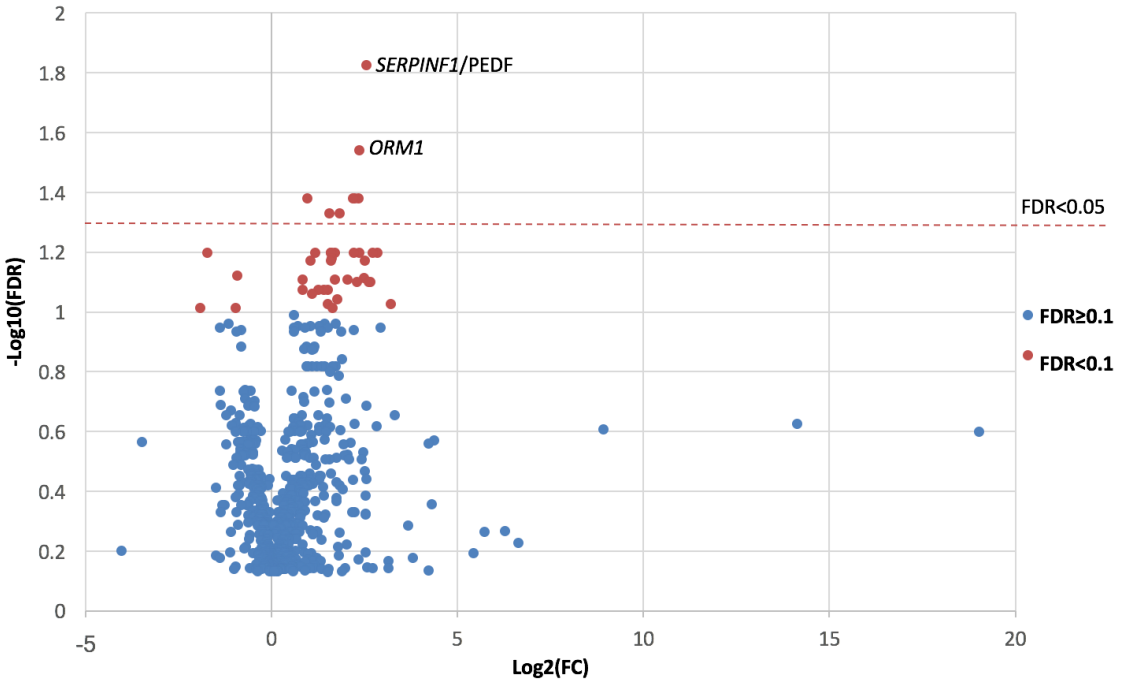

Figure 4.10 – Volcano plot displaying DE proteins between patient and controls (excluding BBS004). Red data points represent differentially expressed proteins with FDR<0.01. Data points above the dashed line have FDR<0.05, where the top 2 most significantly expressed proteins are labelled.

Pigment epithelium derived factor (PEDF, encoded by *SERPINF1*) was the most significant finding, which was upregulated in patients (BBS mean log2RQ = 9.36, 95% CI= 8.86, 9.85; control mean log2RQ = 7.03, 95% CI= 6.60, 7.46; log2(FC)= 2.56; FDR= 0.015). Other statistically significant DE proteins with FDR<0.05 included alpha-1 acid glycoprotein 1, alpha-1 acid glycoprotein 2, carbonic anhydrase 1, armadillo repeat protein, titin, dynein intermediate chain 1, and PHD finger protein 6 (encoded by *ORM1*, *ORM2*, *CA1*, *ARVCF*, *TTN*, *DNAI1* and *PHF6*, respectively). All proteins were significantly upregulated compared to controls.

The 39 proteins with FDR<0.1 were subjected to a literature and OMIM search to identify any known relationships to disease (**table 4.3**). 17 proteins were identified to be linked to disease, 10 of which were CVD or type II diabetes. 4 of the 8 proteins with FDR<0.05 were known to either have genetic links to CVD (*ORM1* and *TTN*) or were previously identified biomarkers for diabetes (PEDF, alpha-1-acid glycoprotein 1 and carbonic anhydrase 1). In addition, mutations in *PHF6* cause an X-linked disorder known as Borjeson-Forssman-Lehman syndrome, which has overlapping features with BBS, such as obesity, intellectual disability, and hypogonadism (Lower et al. 2002).

Subsequently, the difference in protein expression was observed at an individual level. The log2(RQ) was established for each individual for all proteins with FDR<0.01. The heatmap.2 package from gplots was used to generate a heatmap in *R*, with hierarchical clustering to determine related groups of proteins or biological groups (**figure 4.11**) (Warnes et al. 2005). BBS004 was also included in this analysis in an attempt to identify anomalies with this sample (highlighted by yellow bar). Within the heatmap, values were normalised for each protein and scaled by row, so that individuals can be compared relative to one another for each protein (red = high expression, blue = low expression).

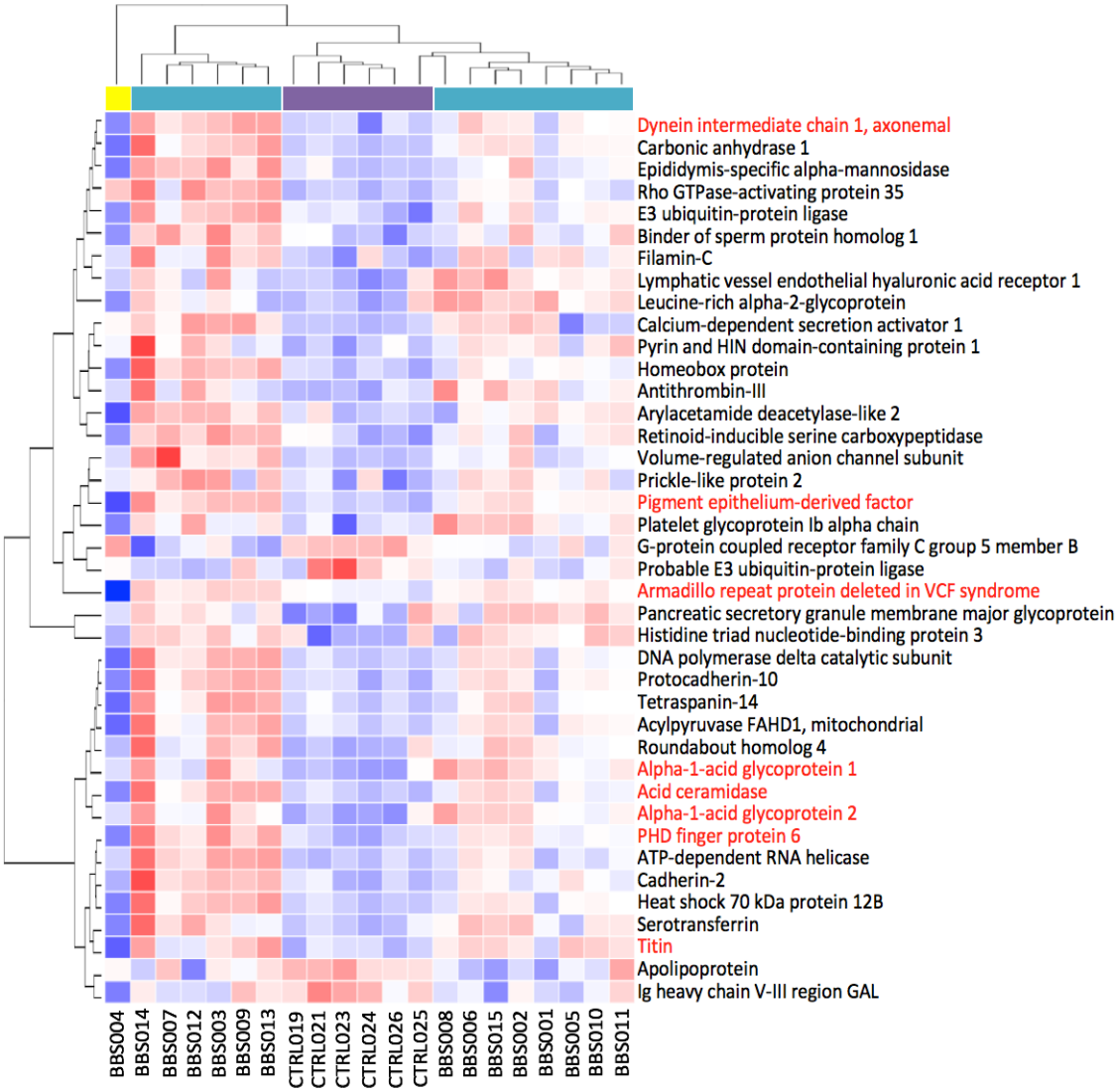

Figure 4.11 – Heatmap of DE proteins with FDR<0.1, generated by heatmap.2 in *R* from scaled and normalised log2(relative quantitation) data, where blue is low expression and red is high expression. Data is organised by hierarchical clustering, where the top bar highlights patients (light blue), controls (purple), and BBS004 (yellow). Proteins that were significantly differentially expressed to FDR<0.05 in patients, excluding BBS004, vs controls are highlighted in red.

The heatmap demonstrated that there were two clusters of patient groups (highlighted by the blue bar), where BBS014, BBS007, BBS012, BBS003, BBS009, and BBS013 formed one group, and BBS001, BBS002, BBS005, BBS006, BBS008, BBS010, BBS011, and BBS015 formed the second group. The control individuals clustered separately from the patients (highlighted by the purple bar), yet there was a close relationship between CTRL025 and BBS008. The heatmap validated that BBS004 was an outlier, as it did not cluster with any of the other samples. This sample had the lowest expression for proteins that were highly expressed in other patients, such as PEDF, suggesting that the reason for the outlier may be due to low protein abundance in this sample. For example, the mean for PEDF in patients excluding BBS004 was 9.36 log2RQ (95% CI 8.86, 8.85), control mean of 7.03 log2RQ (95% CI= 6.60, 7.46), and only 3.47 log2RQ for BBS004.

4.2.1.5 Pathway analysis

For pathway analysis, gene ontology (GO) term databases were utilised to determine whether identified proteins are known to act within the same pathways (Mi et al. 2017). 726 quantifiable proteins were entered into PANTHER database for statistical overrepresentation testing and searched against Reactome pathway, cell component, and biological process GO terms. 58 Reactome pathways, 123 cellular component terms, and 366 biological processes terms were identified to have an adjusted p-value <0.05 after Bonferroni correction (supplemental data, **table S.42**). The top 10 most significantly enriched GO terms are presented in **figure 4.12**.

Many of the biological processes GO terms were related to exocytosis and secretion, which would be expected for a biological fluid (Anderson and Anderson 2002). This was supported by cellular component GO terms, which consist of terms such as extracellular exosome, extracellular vesicle, and secretory granule. There were also terms related to immunological processes, including complement cascade from Reactome pathways, and myeloid cell activation, neutrophil activation, and leukocyte mediated immunity from biological processes.

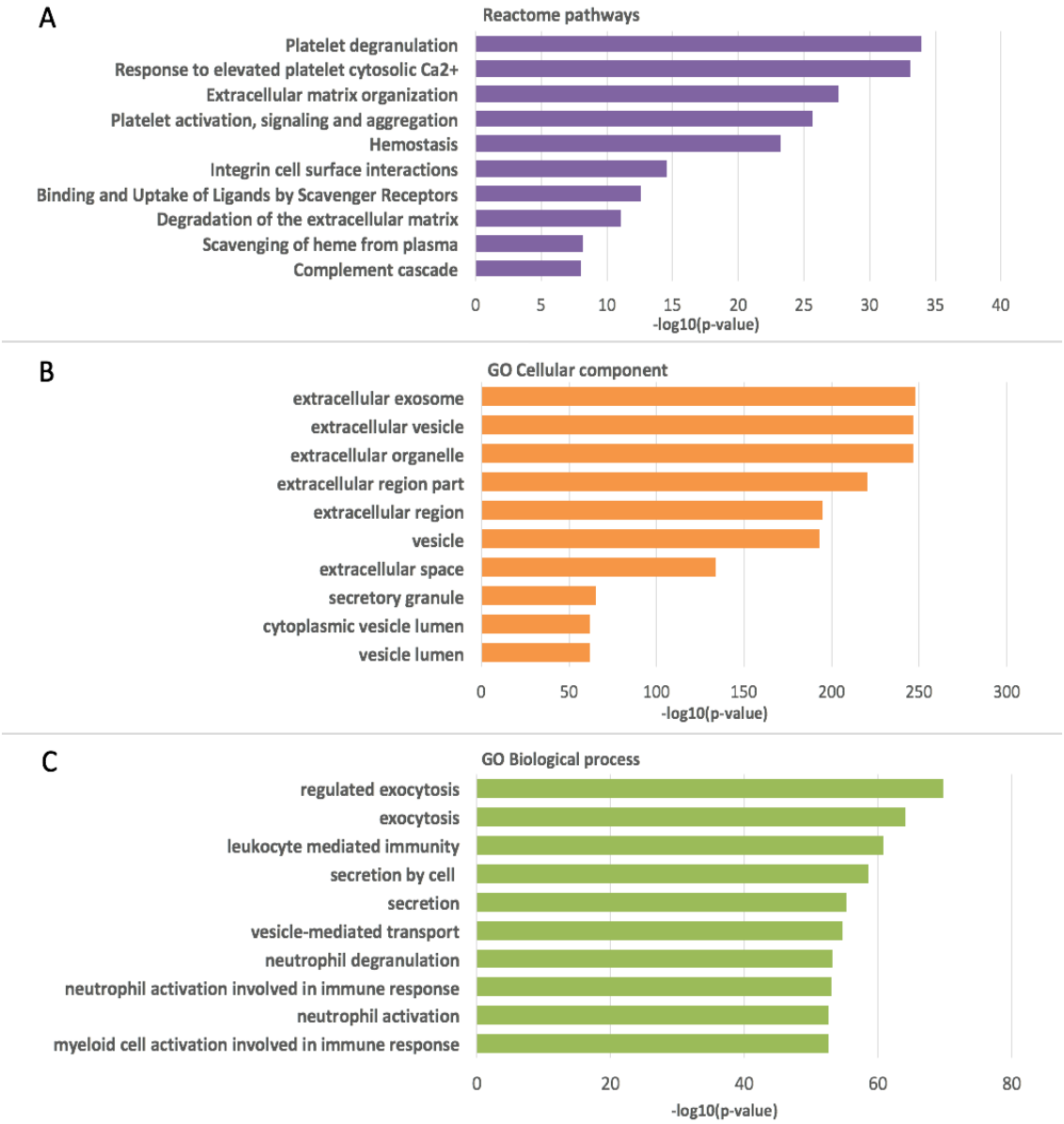

Figure 4.12 – PANTHER overrepresentation testing pathway analysis from urinary proteins. Top 10 most significant terms from Reactome pathways (A), GO cellular component (B), and GO biological processes (C) databases, from 726 quantifiable proteins identified from urine.

STRING software (Szklarczyk et al. 2015) was utilised to identify protein-protein interactions (PPI) between 39 identified proteins with FDR<0.1. 17 out of 39 proteins were predicted to interact with at least one other protein in the test list (**figure 4.13**). This revealed a PPI enrichment p-value of 0.023, suggesting that these proteins have a greater number of interactions among themselves than would be expected by chance. STRING investigated the connectivity of the proteins within the list and identified 18 proteins known to function within extracellular exosomes (p = 0.0002; highlighted in red). 13 edges/connections were present out of an expected 5 for the inputted

protein set size. The largest interaction with 7 edges was between ORM1, ORM2, SERPINC1, LPA, TF, GP1BA, and FLNC, where SERPINC1 (antithrombin-III) is at the centre of the network.

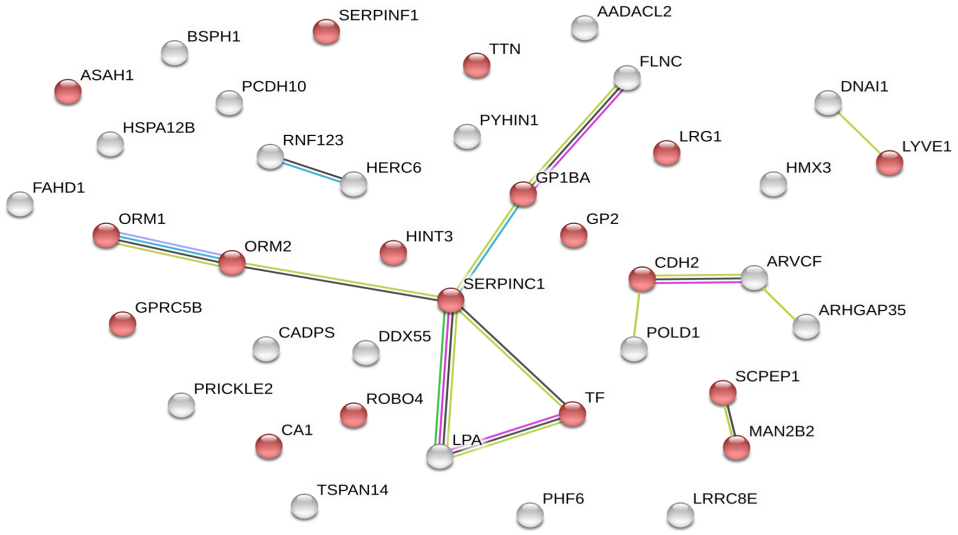

Figure 4.13 – Protein-protein interaction analysis of 39 proteins with FDR<0.1.Connecting edges correspond to proteins known to interact by experimental means (pink), from curated databases (blue), from literature (green), and co-expression analysis (black). Red nodes are known to localise to extracellular exosomes. Image generated by STRING v.10.5.

4.2.2 Proteomic profiling of plasma by mass spectrometry

Plasma requires a vast amount of preparation in order to increase the prospect of detecting the circulating proteome. The first necessity is to remove the highly abundant proteins that make up approximately 95% of plasma, such as albumin, fibrinogen, and globulins (Millioni et al. 2011; Anderson and Anderson 2002). These proteins can be removed through enrichment of low abundance proteins, technology demonstrated by ProteoMiner (Bio-Rad Laboratories), or depletion of highly abundant proteins by means of a ProteoPrep immunodepletion column (Sigma-Aldrich). Due to the high cost of immunodepletion, the ProteoMiner enrichment kit was utilised here.

Preliminary investigations were undertaken for method development purposes. In the first instance, plasma samples obtained from 2 individuals were enriched, fractionated by SDS-PAGE and analysed by UPLC-MS/MS. Samples were quantified by BCA assay, where the standard curve (**figure 4.14**) was utilised to obtain sample concentrations and protein volumes for further processing (**table 4.4**).

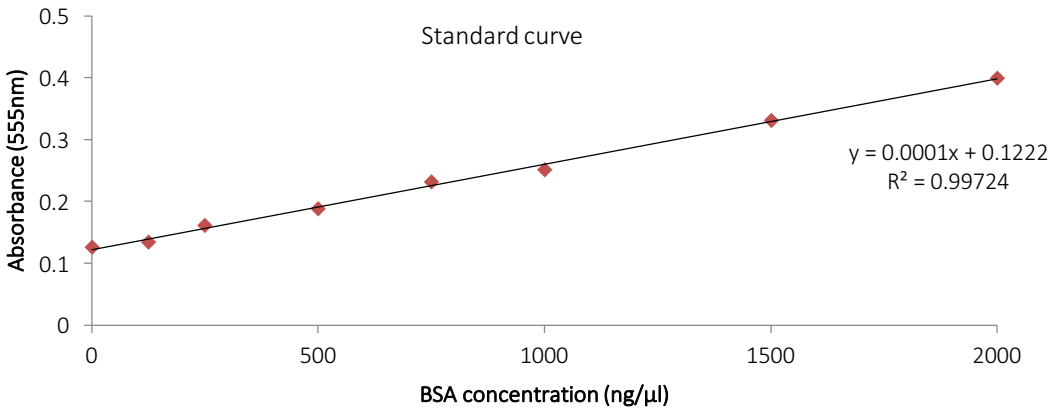

Figure 4.14 – Standard curve generated for quantitation of protein extracted from plasma.

Sample ID	Average absorbance	Concentration (µg/µl)	Yield (µg)	Volume for 50µg
BBS010	0.80	6.8	405.7	7.4
BBS011	0.82	7.0	418.7	7.2
Proteominer FT	0.43	3.1	187.1	16.0
Proteominer wash 1	0.36	2.3	139.8	21.5
Proteominer wash 2	0.42	3.0	177.5	16.9

Table 4.4 – Values from sample quantitation after BCA assay of protein obtained from plasma and corresponding volume of protein required for downstream processing for mass spectrometry

4.2.2.1 Reduction of plasma complexity

Each sample was enriched using a primed ProteoMiner column, where the plasma sample was loaded onto chromatographic beads containing a library of highly diverse hexapeptides. Each hexapeptide uniquely bound to a different protein. Highly abundant proteins oversaturated their ligand and were washed through the column in sequential washing steps, whereas simultaneously, low abundance proteins were concentrated. The final eluate contained a protein sample with reduced complexity. Each stage of the protocol is represented in **figure 4.15**, where each elution step was retained to be separated by SDS-PAGE. The resulting gel illustrates that highly abundant proteins were successfully washed through the column, and resulted in a comprehensible sample for further analysis. As an example, albumin (molecular weight 66.5kDa) saturated the gel in the first flow-through, but dramatically reduced throughout the subsequent washes. Although the final product contained albumin, it was notably reduced, revealing additional bands that were not originally visible.

Following ProteoMiner enrichment, samples were quantified by BCA assay and 50µg of protein was prepared for SDS-PAGE. This separation technique segregates proteins according to their molecular weight. Each sample was cut into 8 pieces of gel (**figure 4.15**). The size of the gel piece

varied depending on the approximate concentration within each band. An effort was made to keep the bands for one protein intact to prevent overlap of proteins into different fractions. The aim was to produce 8 fractions containing ~6.25µg of protein each. It was ensured that bands were selected at the same position for both samples, so that they could be compared.

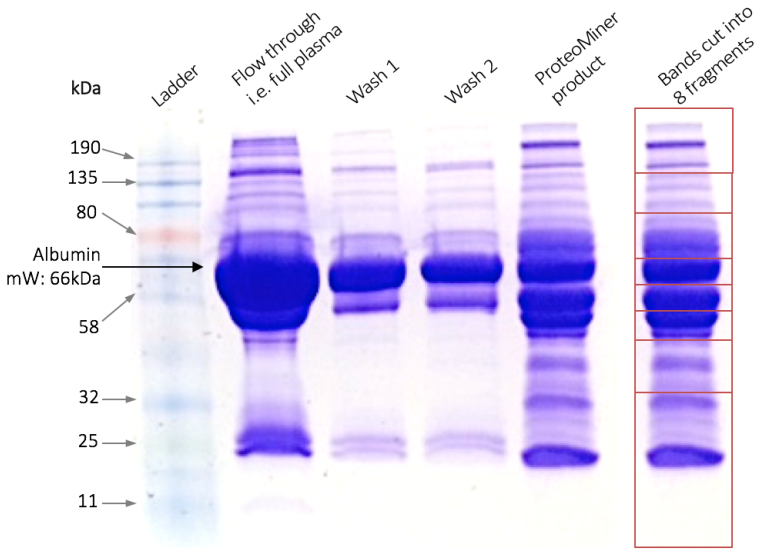

Figure 4.15 – SDS-PAGE gel for products from flow-through (FT), wash 1, and wash 2 of the ProteoMiner protocol. The final lane shows a lower complexity plasma sample than prior to ProteoMiner. The positions where the 8 fractions were cut from the gel is also shown.

4.2.2.2 Label-free UPLC-MS/MS analysis of plasma

For both individuals, each fraction generated a fragment ion spectrum file for input into Progenesis QIP. Sample runs from a single fraction were analysed separately to generate protein abundances and identifications in 8 different Progenesis QIP experiments. To enhance the protein coverage, all 8 experiments were merged and normalised to create the full profile of each sample. Normalisation occurred both within fractions and between fractions to reduce variability observed through systematic errors. Although there was some fluctuation in peptide number across the 8 fractions, in general, the peptide count was fairly consistent throughout (**figure 4.16**).

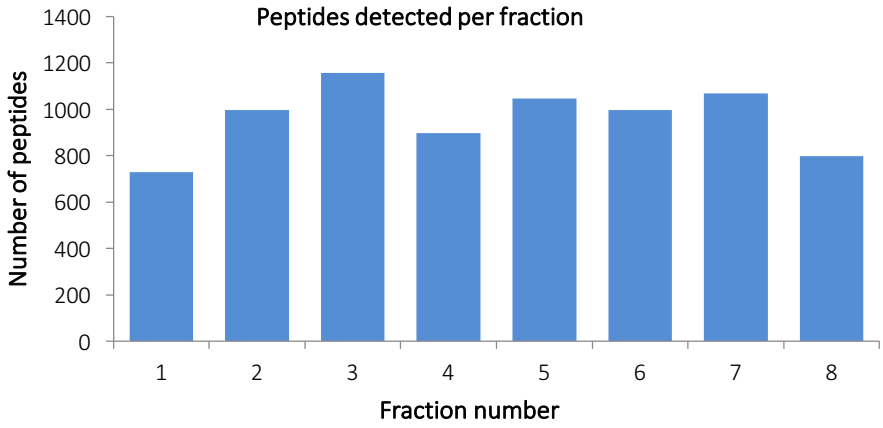

Figure 4.16 – The distribution of peptides from plasma over 8 fractions. Peptides are representative of both samples run in the Progenesis QIP experiment.

Overall, 237 proteins were identified, of which 222 were quantifiable through identification of at least 1 unique peptide (supplemental data, **table S.28**). 104 of these proteins had 3 or more unique peptides, whereas 75 proteins were identifiable by a single unique peptide. 6 magnitudes of abundance were recorded, where the majority had a relative abundance less than 1 (**figure 4.17**). These low abundance proteins may not have been identifiable without the ProteoMiner protein enrichment step. However, ProteoMiner does not completely remove the highly abundant proteins, as 3 different fibronectin chains, complement proteins, and serum albumin dominate the 10 most abundant proteins (**table 4.5**).

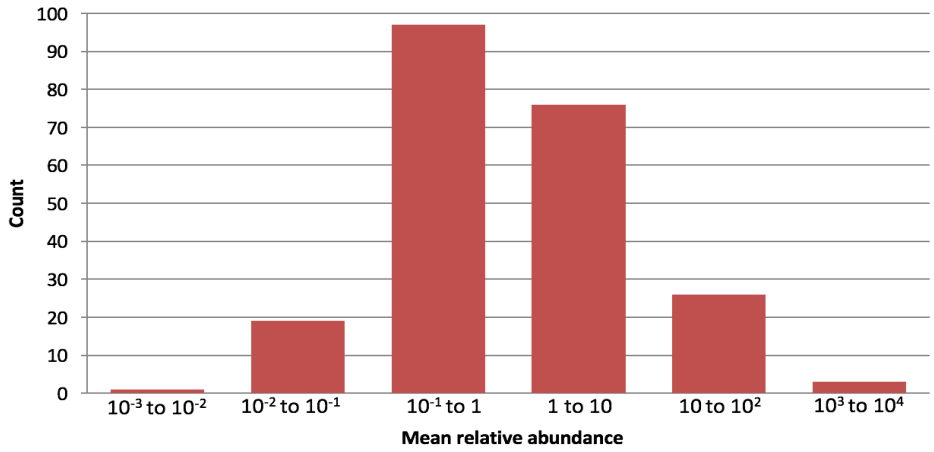

Figure 4.17 – Histogram of mean relative abundances for quantified proteins from plasma LC-MS/MS.

This method development investigation has demonstrated that protein enrichment methods and SDS-PAGE are suitable methods for reducing sample complexity. However, the circulating proteome is predicted to contain approximately 3400 proteins, which span up to 12 magnitudes

of concentration (Gold et al. 2010). Therefore, the 222 proteins observed in this experiment only made a small indent upon the full plasma proteome. Due to the limits of profiling plasma with mass spectrometry, this was technique was not pursued further. Therefore, a different profiling methodology was implemented that does not share the same limitations as mass spectrometry, to attempt a more comprehensive analysis of patient and control plasma.

Gene name	Protein gene	Peptide count	Unique peptides	Confidence score	Relative quantitation		
					BBS010	BBS011	Mean
FGG	Fibrinogen gamma chain	76	69	660.326	137.19	152.43	144.81
FGA	Fibrinogen alpha chain	170	168	1124.698	135.29	150.48	142.89
FGB	Fibrinogen beta chain	107	87	959.443	93.54	106.46	100.00
ALB	Serum albumin	104	82	702.502	112.60	34.29	73.45
APOA1	Apolipoprotein A 1	61	58	624.243	68.95	76.75	72.85
VTN	Vitronectin	47	43	331.224	67.04	78.24	72.64
IGHM	Ig mu chain C region	60	21	399.970	59.12	54.87	56.99
F2	Prothrombin	52	50	484.531	44.82	54.73	49.78
TESK2	Dual specificity testis-specific protein kinase 2	1	1	12.034	31.99	39.92	35.96
CLU	Clusterin	42	42	365.710	32.74	38.08	35.41

Table 4.5 – Top 10 most abundant proteins identified from mean of patient BBS010 and BBS011 plasma by LC-MS/MS.

4.2.3 Proteomic profiling by a targeted technology: SOMAscan

4.2.3.1 Targeted proteomic profiling of plasma

Plasma was taken from the same stock used for mass spectrometry analysis previously described. 15 BBS patient and 6 control plasma samples were assayed by UCL Genomics (UCL GOS Institute of Child Health) and the corresponding data outputs were sent to SOMAlogic (Boulder, CO, USA) for QC testing and quantitation. Unfortunately, patient BBS003 did not pass QC tests, therefore it was excluded from the analysis. Upon retrieval of the quantified data, samples were normalised to allow accurate comparison between samples.

The full capacity of the SOMAscan assay was to detect 1310 proteins (supplemental data, **table S.41**). In this experiment, 1243 proteins were detected from plasma (supplemental data, **table S.26**). The samples were quantified in relative fluorescent units (RFU), which is a relative measure of sample protein quantity. The quantitation ranged from 5.590 RFU up to 17.991 RFU. The mean 20 most and least abundant proteins are presented in **table 4.6**. In some cases, there are multiple gene names for a single protein target, for example Immunoglobulin G. This is present for proteins

with shared sequence homology, where specific aptamers cannot be synthesised to distinguish between closely related proteins.

Gene name	Protein name	Mean abundance (RFU)
APOE	Apolipoprotein E (isoform E2)	17.991
SERPINA3	Alpha-1-antichymotrypsin	17.811
APOE	Apolipoprotein E (isoform E4)	17.710
TF	Serotransferrin	17.707
CFP	Properdin	17.596
IGHG1/IGHG2/IGHG3 /IGHG4/IGK2/IGL2	Immunoglobulin G	17.512
APOE	Apolipoprotein E (isoform E3)	17.460
FGA/FGB/FGG	Fibrinogen	17.327
C3	C3a anaphylatoxin des Arginine	17.179
F2	Prothrombin	17.053
C4A/C4B	Complement C4	16.963
SERPINA5	Plasma serine protease inhibitor	16.635
C6	Complement component C6	16.632
SERPINC1	Antithrombin-III	16.632
F5	Coagulation Factor V	16.619
C3	Complement C3	16.336
ALPL	Alkaline phosphatase	16.332
C9	Complement component C9	16.281
CFH	Complement factor H	16.196
FN1	Fibronectin Fragment 4	16.183
KLK5	Kallikrein-5	6.889
IFNL2	Interferon lambda-2	6.874
	Protein Rev_HV2BE	6.816
IL10	Interleukin-10	6.813
NANOG	Homeobox protein NANOG	6.793
SIGLEC1	Sialoadhesin	6.784
PRDX5	Peroxiredoxin-5	6.761
MPL	Thrombopoietin Receptor	6.744
RPS6KA5	Ribosomal protein S6 kinase alpha-5	6.719
PON1	Serum paraoxonase/arylesterase 1	6.647
PTK2	Focal adhesion kinase 1	6.603
IL17B	Interleukin-17B	6.591
HCE004333	HCE004333	6.422
THPO	Thrombopoietin	6.415
RNASEH1	Ribonuclease H1	6.407

LDLR	Low-density lipoprotein receptor	6.379
LRRTM1	Leucine-rich repeat transmembrane neuronal protein 1	6.257
IL7	Interleukin-7	6.200
CTLA4	Cytotoxic T-lymphocyte protein 4	5.822
IFNA2	Interferon alpha-2	5.590

Table 4.6 – The 20 most abundant and least abundant proteins identified by SOMAscan from plasma. The average abundance is presented in relative fluorescence units (RFU).

Statistical analysis was completed to identify DE proteins between patients and control biological groups (supplemental data, **table S.25**, **table S.27**). 100 proteins were found to be statistically significant with a p-value <0.05. However, significance was not maintained after the multiple testing correction was applied. The top 25 differentially expressed proteins are presented in **table 4.7** and **figure 4.18**. Although not a significant finding, the protein with the smallest p-value was leptin, found to be upregulated in patients compared to controls (BBS mean = 13.22 RFU, 95% CI = 12.69, 13.75; control mean = 12.07 RFU, 95% CI = 11.25, 12.88; log2(FC)= 1.269; p-value= 0.0003; FDR= 0.250). This may be biologically interesting as BBS patients have been reported to exhibit leptin resistance (Feuillan et al. 2011). Of note, CRP, which is the current BBS biomarker adopted in clinics, is also present in the top 25 most significant proteins (BBS mean = 15.29 RFU, 95% CI = 14.86, 15.72; control mean = 14.01 RFU, 95% CI = 13.09, 14.93; log2(FC) = 1.280; p-value = 0.0094; FDR = 0.503).

Gene name	Target name	BBS mean (RFU)	Control mean (RFU)	Log2FC	p-value	FDR
LEP	Leptin	13.22 (12.69, 13.75)	12.07 (11.25, 12.88)	1.269	0.0003	0.250
PGAM1	Phosphoglycerate mutase 1	11.53 (10.90, 12.15)	9.57 (8.86, 10.27)	1.877	0.0006	0.250
OLR1	Oxidized low-density lipoprotein receptor 1	11.58 (11.34, 11.82)	10.72 (10.45, 10.98)	0.849	0.0006	0.250
HGF	Hepatocyte growth factor	10.41 (10.25, 10.58)	9.87 (9.74, 10.00)	0.549	0.0014	0.441
SOD2	Superoxide dismutase, mitochondrial	14.80 (14.64, 14.95)	15.35 (15.10, 15.61)	-0.548	0.0022	0.465
MMP13	Collagenase 3	8.68 (8.56, 8.80)	9.14 (8.94, 9.35)	-0.464	0.0022	0.465
CNTN5	Contactin-5	9.01 (8.88, 9.14)	8.58 (8.40, 8.77)	0.435	0.0028	0.480
F9	Coagulation factor IX	13.84 (13.75, 13.94)	13.48 (13.33, 13.64)	0.357	0.0031	0.480
ENO1	Alpha-enolase	13.26 (12.99, 13.54)	12.40 (11.88, 12.92)	0.831	0.0040	0.482

WFIKKN2	WAP, Kazal, Ig, Kunitz and NTR domain-containing protein 2	11.87 (11.67, 12.07)	12.42 (12.03, 12.81)	-0.513	0.0045	0.482
KLK3	Prostate-specific antigen	9.46 (9.05, 9.87)	10.23 (8.85, 11.61)	-0.903	0.0050	0.482
IGHG1/IGHG2/ IGHG3/IGHG4/ IGK2/IGL2	Immunoglobulin G	17.39 (17.24, 17.53)	17.81 (17.66, 17.95)	-0.424	0.0053	0.482
F9	Coagulation factor IXab	13.35 (13.26, 13.44)	13.02 (12.88, 13.17)	0.323	0.0054	0.482
IGFBP1	Insulin-like growth factor-binding protein 1	7.89 (7.31, 8.66)	9.65 (9.05, 10.26)	-1.603	0.0058	0.482
S100A9	Protein S100-A9	12.85 (12.51, 13.19)	11.78 (10.99, 12.58)	1.016	0.0061	0.482
DSG2	Desmoglein-2	15.52 (15.43, 15.62)	15.85 (15.70, 16.01)	-0.323	0.0064	0.482
APOM	Apolipoprotein M	12.42 (12.11, 12.73)	13.27 (12.87, 13.68)	-0.836	0.0066	0.482
FABP5	Fatty acid-binding protein, epidermal	9.69 (9.54, 9.84)	9.26 (9.06, 9.45)	0.420	0.0077	0.503
IL1RAP	Interleukin-1 Receptor accessory protein	14.24 (13.97, 14.50)	15.05 (14.49, 15.60)	-0.787	0.0082	0.503
IGHA1/IGHA2	Immunoglobulin A	11.24 (10.97, 11.51)	11.84 (11.65, 12.03)	-0.624	0.0087	0.503
IL1RN	Interleukin-1 receptor antagonist protein	11.91 (11.56, 12.26)	11.20 (11.11, 11.29)	0.744	0.0090	0.503
IL12RB2	Interleukin-12 receptor subunit beta-2	9.42 (9.24, 9.61)	10.34 (9.44, 11.24)	-0.920	0.0093	0.503
CRP	C-reactive protein	15.29 (14.86, 15.72)	14.01 (13.09, 14.93)	1.280	0.0094	0.503
BCAN	Brevican core protein	10.20 (10.01, 10.40)	10.70 (10.51, 10.89)	-0.483	0.0097	0.503
GDF11/MSTN	Growth differentiation factor 11/8	9.14 (8.92, 9.37)	9.54 (9.33, 9.76)	-0.437	0.0111	0.554

Table 4.7 – The top 25 DE proteins ranked by p-value identified from plasma by SOMAscan. Sample mean for BBS and control are also presented, where data is expressed as mean (95% CI). Note that significance was not maintained after multiple testing to an FDR <0.05.

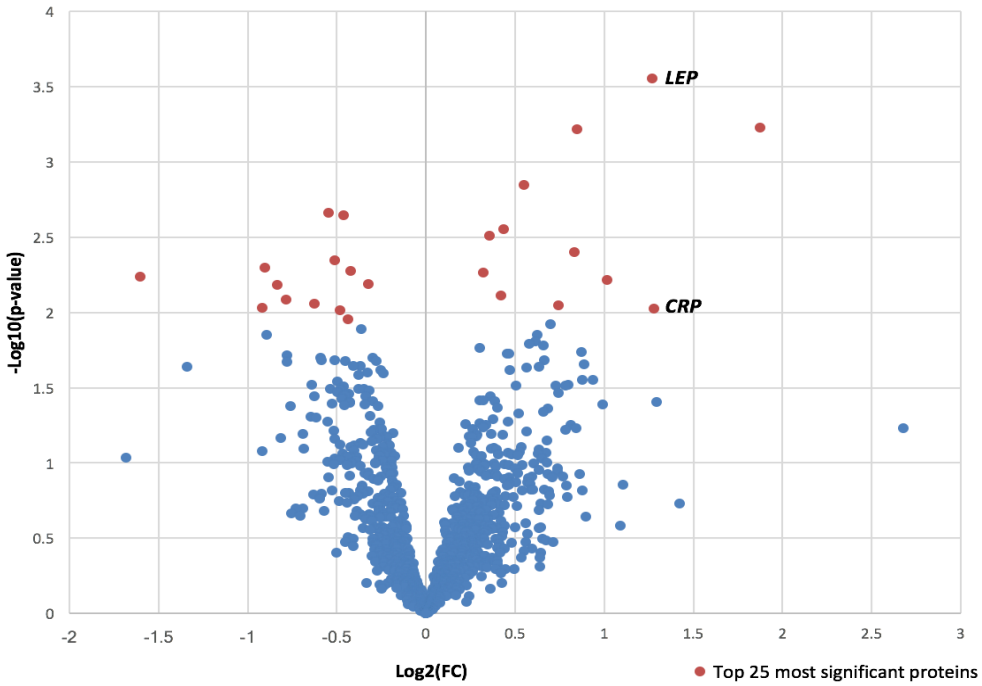

Figure 4.18 – Volcano plot displaying differential expression of proteins between patient and controls from plasma, plotted as a factor of fold change vs p-value. Red data points represent the top 25 most significant proteins in **table 4.7**.

4.2.3.2 Comparison of proteins by targeted and untargeted techniques

The protein identifications for untargeted mass spectrometry and targeted SOMAscan were compared to determine whether these techniques were able to recognise the same proteins. Out of 1243 SOMAscan findings and 222 proteins identified by mass spectrometry, 34 proteins were shared (**figure 4.19**; appendix, **table A.5**). This demonstrates that these different technologies may be suited to detect different types of proteins and suggests that they may be most effectively used in combination with one another.

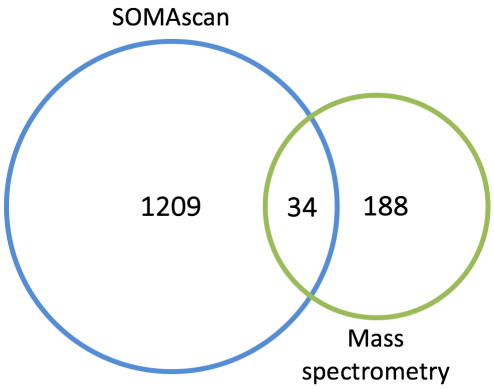

Figure 4.19 – The number of total protein identifications from plasma for SOMAscan, untargeted mass spectrometry, and shared between both technologies.

4.3 Discussion

In this chapter, urine and plasma samples from BBS patients were profiled using label-free and targeted proteomic techniques. To our knowledge, this is the first study where the chemical composition of BBS biofluids have been investigated.

4.3.1 Proteomic profiling of urine

For proteomic profiling of urine, UPLC coupled to the Synapt G2 Si mass spectrometer was utilised with HDMS^E capability. 726 quantifiable proteins were detected by identification of 7901 peptide ions across 15 BBS patient and 6 control samples. 5 magnitudes of abundance were uncovered by relative quantitation; however, low abundance proteins (<10 RQ) were not found. This could be due to high complexity of sample, limited instrument sensitivity, or undersampling. Although urine was considered to have relatively low complexity, especially compared to plasma, previous studies demonstrated that urine is actually a more complicated fluid than was anticipated. The current urinary proteome is known to exist as at least 2362 proteins, as determined by Kentsis and colleagues (2009). This high identification rate was possible due to extensive fractionation and separation techniques, including ultracentrifugation, SDS-PAGE, ion exchange chromatography and reverse phase chromatography, before mass spectrometry (Kentsis et al. 2009).

During analysis by Progenesis QIP, samples BBS004, BBS005, BBS008, BBS015, and CTRL026 were identified as having low quality fragmentation spectra after MS/MS. Alignment and normalisation of MS1 spectra, and use of ion abundance quantitation by Progenesis QIP, meant that this should not have negatively affected protein identification. However, the software could only quantify ions if they were also seen in another sample. Therefore, if there were ions unique to samples that were not detected, some proteins may have been underrepresented.

The principle component analysis of the experiment presented a method for quality control. 4 QC samples, which were representative of all samples, were run throughout the course of the experiment. The PCA plot demonstrated that the instrument sensitivity was maintained and thus the data was reliable for further analysis. The PCA plot also allowed a visual characterisation of the data. As a result, BBS004 was identified as a possible outlier. It was originally considered that this may be due to BBS004 not displaying a metabolic syndrome phenotype compared to other patients within the cohort. However, further investigation by heatmap analysis of the data

demonstrated that this may be due to poor sample quality, as some of the significantly regulated proteins were found only in low quantities in BBS004, including PEDF.

4.3.1.1 Statistical analysis of proteins from urine

In the first instance, DE proteins were calculated between 15 patients and 6 control samples. A single DE protein was identified, namely alpha-1-acid glycoprotein 1 (ORM1) (BBS mean log2RQ= 16.46, 95% CI= 15.95, 16.98; control mean log2RQ= 14.15, 95% CI= 13.31, 15.00; log2(FC)= 2.31; FDR= 0.039) (table 4.2). Upon removal of BBS004 from the analysis, the statistical significance of this protein increased (BBS mean log2RQ= 16.56, 95% CI= 16.05, 17.08; control mean log2RQ= 14.15, 95% CI=13.31, 15.00; log2(FC)= 2.36; FDR= 0.029) and the FDR of an additional 7 proteins decreased to less than 0.05, all of which were upregulated in patients. Of these, 4 proteins have been previously linked to aspects of metabolic syndrome, such as diabetes (PEDF and carbonic anhydrase 1) or CVD (ORM1 and titin) (table 4.3) (Yamagishi et al. 2006; Christiansen et al. 2002; Hou et al. 2014). These proteins may be indicative of onset of type II diabetes or cardiovascular complications in the BBS patients.

4.3.1.2 Identification of urinary protein biomarkers

Pigment epithelium-derived factor (PEDF)

PEDF was recognised as the most significantly differentially expressed protein between patient and control biological groups, after removal of BBS004 (BBS mean log2RQ= 9.36, 95% CI= 8.86, 9.85; control mean log2RQ= 7.03, 95% CI= 6.60, 7.46; Log2(FC) = 2.56; FDR = 0.015). PEDF is a glycoprotein belonging to the family of serine protease inhibitors. It is a multifunctional protein that is synthesised in adipose tissue and secreted to negatively regulate angiogenesis (Dawson et al. 1999). Studies have reported PEDF to be both pro-inflammatory (Crowe et al. 2009; Chavan et al. 2012) and anti-inflammatory (Nakamura et al. 2009), eliciting the assumption that this protein may exhibit its inflammatory properties in a tissue specific or temporal manner. PEDF has been identified as a serum biomarker in previous studies relating to metabolic syndrome, as well as renal and retinal dysfunction related to diabetes (Nakamura et al. 2009; Yamagishi et al. 2006; Crowe et al. 2009; Jenkins et al. 2008). It was shown to be upregulated in patients with metabolic syndrome and thought to contribute to insulin resistance. Until now, there was no previous evidence that PEDF could also be a urinary biomarker.

The significant upregulation within BBS patients compared to controls may be due to PEDF being one of the most abundantly secreted proteins from adipocytes (Famulla et al. 2011). Famulla et

al. (2011) determined that PEDF expression significantly increased throughout adipogenesis, was further augmented by treatment of insulin, and was suppressed by induction of hypoxia. PEDF was also demonstrated to induce insulin resistance in adipocytes and mice *in vivo*, as well as activate inflammatory signalling (Famulla et al. 2011; Crowe et al. 2009). The mechanism for this process, however was not elucidated.

Interestingly, PEDF is not only a key player in adiposity, and has demonstrated other roles relevant to BBS. PEDF was first shown to be expressed in cultured retinal pigmented epithelium (RPE) cells (Tombran-Tink et al. 1991). *In vivo*, RPE cells exist as a pigmented layer behind the photoreceptor layer in the retina. RPE functions to support and maintain photoreceptor and neural retina function (German et al. 2008). PEDF was shown to be secreted from RPE cells, where it exhibited neurotrophic activity to promote photoreceptor and retinal neuronal cell survival (Steele et al. 1993; Cayouette et al. 1999). Furthermore, Cayouette and colleagues (1999) prolonged the survival of photoreceptors in a mouse model of retinal degeneration (*rd*) through intraocular injection of PEDF. It has not yet been established whether RPE cells increase expression of PEDF in response to photoreceptor degeneration in models of retinal dystrophy. However, it is worth investigating whether PEDF is upregulated in BBS patients compared to controls as a consequence of photoreceptor death. The observation that a protein from the ocular system is found in the urine is not unlikely, as Kentsis et al. (2009) uncovered proteins from urine that are uniquely expressed in the nervous system.

Alpha-acid glycoprotein 1 and 2 (ORM1 and ORM2)

ORM1 and ORM2 were both upregulated in patients compared to control with log2(FC) of 2.36 (BBS mean log2RQ= 16.56, 95% CI= 16.05, 17.08; control mean log2RQ= 14.15, 95% CI= 13.31, 15.00; FDR = 0.029), and log2(FC) of 2.35 (BBS mean log2RQ= 15.23, 95% CI= 14.66, 15.79; control mean log2RQ= 13.01, 95% CI= 12.34, 13.68; FDR = 0.042), respectively. These paralogues are both acute inflammatory circulatory glycoproteins. ORM1 was reported as a urinary biomarker for chronic heart failure (CHF) (Hou et al. 2014), which was upregulated in CHF patients compared to normal controls. CVD is a secondary phenotype of BBS, linked to the syndrome's obesity phenotype. Indeed, patients in this cohort exhibit risk factors or features of CVD, such as hypertension (BBS002, BBS003, BBS005), aortic valve complications (BBS013), and angina pectoris (BBS005). In a gene expression experiment, upregulation of *ORM1* was recorded in circulating blood of metabolic syndrome patients, but not in patients with coronary artery disease (Grayson et al. 2011). This may suggest that ORM1 is an indicator of obesity-related

cardiovascular abnormalities. ORM2 was shown to be significantly elevated in plasma from colorectal cancer (CRC) patients (Zhang et al. 2012). As yet, there has been no connection between BBS and an increased risk of CRC. With the exception of Grayson et al. (2011), the studies discussed here utilised clinically healthy individuals as controls. Thus, it is possible the significant changes in expression may be due to general inflammatory responses and not specific to CHF or CRC.

Other relevant findings

After investigation of the significant proteins at an FDR<0.05, the confidence interval was widened to 90%, with the knowledge that this increased the chance of false positive identifications. This identified 5 more markers associated with type II diabetes, namely G-protein coupled receptor-C group 5B (BBS mean log2RQ= 8.49, 95% CI= 7.76, 9.13; control mean log2RQ= 10.49, 95% CI= 10.00, 10.97; log2(FC)= -1.72; FDR= 0.064), serotransferrin (BBS mean log2RQ= 14.60, 95% CI= 14.15, 15.06; control mean log2RQ= 13.23, 95% CI= 12.89, 13.57; log2(FC) = 1.59; FDR = 0.064), Heat Shock Protein 12B (BBS mean log2RQ= 13.81, 95% CI= 13.16, 14.45; control mean log2RQ= 11.93, 95% CI= 11.64, 12.22; log2(FC) = 2.37; FDR = 0.064), Leucine rich alpha-2-glycoprotein (BBS mean log2RQ= 12.36, 95% CI= 11.33, 13.38; control mean log2RQ = 8.99, 95% CI= 7.20, 10.78; log2(FC) = 2.5; FDR=0.077) and Apolipoprotein(A) (BBS mean log2RQ= 13.21, 95% CI= 12.76, 13.67; control mean log2RQ= 14.36, 95% CI= 14.02, 14.69; log2(FC) = -0.95; FDR=0.097) (Soni et al. 2013; Soggiu et al. 2012; Ruotolo et al. 1991). Although these proteins may have a genuine specificity to BBS-related metabolic syndrome, it is important to consider that they may be an artefact of disease associated inflammation. To test this, further experimentation will be required with larger BBS and control groups. To evaluate whether these differentially expressed proteins are specific to BBS, a control group composed of metabolic syndrome patients could be used. A greater improvement would be to compare the proteomes of BBS patients that manifest obesity with patients that have a normal BMI. Following identification of possible biomarkers, it is imperative that verification through targeted analysis, such as ELISA, is performed.

4.3.1.3 Cellular components of urine

Overrepresentation analysis by PANTHER of all 726 quantifiable proteins identified an enrichment of proteins associated with extracellular vesicles and exosomes GO terms (p<0.0001). Furthermore, STRING protein-protein interaction analysis identified 18 proteins out of 39 proteins with FDR<0.1 associated with extracellular exosomes. Urinary exosomes are shed from

renal epithelial cells and contain RNAs, lipids, and proteins that characterise the pathophysiological state of their host cells. Previous studies have shown that urinary exosomes can act as valuable vehicles for biomarker detection (Dear et al. 2013; Fang et al. 2013; Gonzales et al. 2009; van Balkom et al. 2011). Indeed, one large-scale study identified 1132 proteins specific to urinary exosomes, 177 of which were associated with disease, including primary cilia proteins ARL6, PKD1, and PKD2 (Gonzales et al. 2009).

In this project, proteins were separated from urine by ultracentrifugation with a 3kDa molecular weight cut-off. This removed low molecular weight molecules such as metabolites, salts, and detergents, however the resulting mixture would have retained its complexity of proteins. For future investigations, the complexity of the sample could be reduced by sequencing exclusively exosomes. This may provide a greater insight into disease-specific intracellular interactions, as other secretions will be removed. To extract exosomes, further refinement of the protocol is required, adding sucrose separation and size-exclusion chromatography steps (Dear et al. 2013). The will result in a less complex, more specific mixture; however, this may contain more sample variation due to extensive handling.

4.3.2 Proteomic profiling of plasma

4.3.2.1 Investigation of plasma proteome by UPLC-MS/MS

Plasma samples from BBS010 and BBS011 were profiled for a method development experiment by mass spectrometry. As plasma is a high complex biofluid, consisting mostly of highly abundant proteins such as albumin and fibrinogen (Anderson and Anderson 2002), plasma samples required a greater number of processing steps compared to urine analysis. First, low abundance proteins were enriched by use of the ProteoMiner column, followed by the separation of proteins into 8 fractions by SDS-PAGE. Samples were profiled using UPLC coupled to the MALDI-QToF Premier mass spectrometer with MS^E capability, where 222 proteins were quantified over 6 logs of concentration.

ProteoMiner was shown to successfully reduce the complexity of the sample, where SDS-PAGE illustrated that the quantity of highly abundant proteins, such as albumin, had reduced. However, of the 222 proteins identified, the top 4 most abundant findings consisted of 3 fibrinogen chains (gamma, alpha, and beta chains) and serum albumin (**table 4.5**). The ProteoMiner column is composed of a dynamic library of hexapeptides that should have affinity for all proteins in a sample. Once a ligand is oversaturated, surplus proteins are washed through the column,

resulting in a less complex protein sample for mass spectrometry (Millioni et al. 2011). The greatest issue with this technique is that it relies on the specificity of the column, where some non-specific proteins may be washed through. Moreover, this method coincidently alters the protein quantity, thereby creating ambiguous results upon quantitation. Depletion of high abundance proteins through an immunodepletion column may be more reliable. This technique uses specific antibodies for abundant proteins, such as albumin and fibrinogen, to retain these on the column and allow other proteins to be washed through (Jaros et al. 2013). However, these are expensive kits that were not obtainable for use in this study.

For plasma proteomic analysis, the MALDI-QtoF Premier instrument was used rather than its high-performance Synapt G2 Si counterpart. The difference between the machines is presented at the MS/MS level, where the Synapt G2 Si utilises $HDMS^E$ to enhance ion separation within the ToF mass analyser. By contrast, the Premier mass spectrometer does not have the high definition capability, so has reduced sensitivity. This difference in performance has been shown to dramatically decrease the number of protein identifications by up to half (Tonge et al. 2011). Fractionation is able to enhance the identification potential to the same level as an unfractionated run on a UPLC-MS/MS system (Plumb et al. 2006). In this study, protein was separated offline by SDS-PAGE into 8 distinct fractions. The benefit of SDS-PAGE is that proteins can be visualised on the gel so approximately comparable amounts of protein can be separated in each fraction. This can also be valuable as a quality control step to identify degradation or missing proteins. Furthermore, the gel acts as a cleaning step as it removes salts and contaminants before analysis (Jafari et al. 2012). On the other hand, the gel can only accurately resolve proteins over a molecular weight of 11kDa. Therefore, very small proteins or polypeptides may be lost from the bottom of the gel.

Despite separating the proteins into 8 fractions, the technique described in this project was only able to identify ~200 proteins, which represents only 6.5% of the reported circulating proteome (Anderson and Anderson 2002). Proteins may have been lost through sample preparation, undersampling or remained undetected by the mass spectrometer. This is supported by 76 proteins only being identified through a single unique peptide (supplemental data, **table S.28**), indicating that many peptides were missing. The small sample size for this method development would not have helped identification. The use of Progenesis QIP analysis software can increase the number of proteins identified through alignment of ion abundance spectra, but this requires a large sample size.

The hope for this project was to complete profiling by mass spectrometry for all 15 patients and 6 control plasma samples. However, due to the limitations of this experiment, such as long preparation protocol and low sensitivity of the mass spectrometer, which led to limited identifications in the method development, it was decided to proceed with a different technique for plasma profiling.

4.3.2.2 Protein profiling using a targeted technique: SOMAscan

The plasma of 15 BBS patients and 6 controls were analysed by the targeted SOMAscan assay. Of the 1310 probes available, 1243 targets were measured over 12 relative fluorescent units. Differentially expressed proteins were calculated between patients and control samples, however no statistically significant comparisons with FDR<0.05 were identified. This suggests that for the 1310 proteins on the SOMAscan panel, proteins detected within the plasma samples were expressed similarly between patient and controls.

Although not statistically significant, there were BBS relevant findings within the top 25 most significant proteins with uncorrected p-value <0.05. The protein that had the lowest p-value was leptin, which was upregulated in patients compared to controls (BBS mean RFU= 13.22, 95% CI= 12.69, 13.75; control mean RFU= 12.0, 95% CI= 11.25, 12.88; log2(FC) = 1.269; p=0.0003). This finding supports studies where obese BBS patients were shown to have leptin resistance and elevated circulatory leptin (Feuillan et al. 2011). Leptin is a hormone secreted by adipocytes in response to energy intake and acts to regulate energy homeostasis through activation of leptin receptors on hypothalamic neurons (Rosenbaum and Leibel 2014). Secreted leptin levels correlate with adiposity (Han and Lean 2016), thus elevated leptin is not specific to BBS patients. However, in a study conducted by Seo et al. (2009), obese control mice could be rescued by leptin treatment, whereas BBS mice did not respond to leptin supplementation, suggesting that resistance was intrinsic and not secondary to obesity. As leptin elevation is known to exist in BBS patients, it is surprising that a significant difference is not observed between patient and control plasma samples. This may be attributed to the *BBS1* p.M390R mutation, where circulatory leptin levels were shown to be less than in patients with mutations in other BBS genes (Feuillan et al. 2011).

CRP was identified to be upregulated in patients compared to controls (BBS mean RFU = 15.29, 95% CI= 14.86, 15.72; control mean RFU = 14.01, 95% CI= 13.09, 14.93; log2(FC) = 1.28; p = 0.009).

Despite being used as a biomarker in clinic, this protein was narrowly within the top 25 most significantly DE proteins. Other inflammatory markers indicative of the acute phase response were also present in the top 25 DE proteins, such as immunoglobulins and interleukin receptors. Apolipoprotein M (ApoM) was downregulated compared to controls (BBS mean RFU= 12.42, 95% CI= 12.11, 12.73; control mean RFU= 13.27, 95% CI= 12.87, 13.68; log2(FC) = -0.836; p = 0.006). ApoM is a component of high density lipoprotein (HDL) particles that function to scavenge lipids, such as cholesterol, phospholipids and triglycerides, from the circulatory system (Mahley, 1988). Decreased ApoM is associated with diabetes and metabolic syndrome, thought to be caused by insulin resistance (Dullaart et al. 2009). Zheng et al. (2014) demonstrated that increased circulating fatty acid levels in rats decreased expression of ApoM and reduced insulin sensitivity (Zheng et al. 2014). This finding may be relevant, as BBS patients exhibit high cholesterol and hypertriglyceridemia (Feuillan et al. 2011). The decrease in ApoM expression may be highlighting the metabolic abnormalities within these patients and could be a sign of insulin resistance. Further validation would be required to determine whether the differential expression of ApoM is a true positive.

4.3.3 Comparison of targeted and untargeted proteomic approaches

The SOMAscan assay was a novel technology only recently commercialised for use in research centres. This project was among one of the first to test this technology at UCL, therefore sample preparation and analysis were in the early stages of development. SOMAscan was optimised for use with plasma, and thus the panel consisted mostly of soluble proteins commonly secreted into blood. As a targeted panel, this technique is limited by the number and affinity of the SOMA aptamers. New aptamers may be available in the future, as the success rate for generating aptamers specific to proteins is high at 85% (Gold et al. 2010). On the other hand, aptamers do not yet have the capability of always recognising proteins with very similar sequence homology. For example, in this study, immunoglobulin G (IgG) was identified from plasma, which may refer to 6 different homologues of IgG (**table 4.7**). At the same time, however, the assay was able to detect at least 3 different isoforms of apolipoprotein E (**table 4.6**).

SOMAscan is highly sensitive and only requires a small sample volume (65µl). This makes it beneficial for clinical analysis, especially for patients where only small volumes of blood can be acquired (e.g. infants). Protease inhibitors are added upon sample collection, which acts to protect samples against degradation during transport from the clinic and sample storage. By contrast, protease inhibitors cannot be utilised for mass spectrometry, as they would impede

trypsinisation and proteocleavage. During sample preparation, minimal processing steps are required prior to analysis, which reduces experimental variation. As a targeted technique, mixture complexity is not a limiting factor for SOMAscan, as each probe is specific and cannot be masked by other proteins with similar chemical properties.

Due to the complexity of proteins and their different chemical characteristics, a single technique may not be able to provide a complete proteomic solution. To investigate this, protein identifications for targeted and untargeted plasma proteomics were compared. Only 34 proteins were seen in both SOMAscan and mass spectrometry experiments (**figure 4.19**). This highlights how different technologies may be more suited for detection of certain proteins, and that a combination of untargeted and targeted approaches would be beneficial to resolve the total profile of a sample. Future elucidation would be required to determine if data can be directly compared across different assays.

4.3.4 Considerations for future biomarker experiments

Another key point to make is that a single sample of urine or plasma reflects only a narrow time point of an individual's proteome. Serial sampling over a long period can enhance the confidence of a true biomarker or can be used to assess disease progression (Jaffe et al. 2006). Moreover, the proteomic composition can be heavily influenced by the time of day, diet, exercise, or medication, and therefore can be variable, even for samples from the same individual. Protein abundance can also be influenced by hydration levels, particularly for urine samples (Dear et al. 2013). An important aspect of this project has been to devise standardised protocols to reduce variability in sample collection and preparation. Inconsistencies derived through storage errors may enhance artificial significances and mask true positives during analysis. In the experiments presented here, samples were treated equally, and randomisation was used where possible. The greatest inconsistency would arise from time taken to transport samples from clinic to the lab, which was uncontrollably varied. This is an aspect that would need to be managed more effectively in the future.

4.3.4.1 Further work for biomarker discovery and validation

In terms of morbidity and mortality in BBS, renal dysfunction and secondary complications of obesity are the most common contributing factors (Guo and Rahmouni 2011; Forsythe and Beales 2013). Therefore, a biomarker for both of these phenotypes would aid patient management and treatment. Although patients BBS003, BBS005, BBS007, and BBS014 had structural renal

abnormalities, it was not possible within this experiment to stratify patients by this feature due to lack of statistical power. Further biomarker discovery studies would need to be implemented with larger groups of BBS patients, where individuals with renal dysfunction could be compared to those without. This investigation would benefit from a longitudinal study with serial sampling to follow the progression of the disease.

Within this chapter, the most significant findings were related to obesity, diabetes, or CVD. The most promising identification was PEDF, which was found to be statistically upregulated in BBS patients compared to controls, and previously linked to insulin resistance (Famulla et al. 2011) and photoreceptor development (Steele et al. 1993; Cayouette et al. 1999). Other results in this chapter support previously known relationships to BBS, such as elevated leptin and CRP (Feuillan et al. 2011). Furthermore, the finding of diminished ApoM in BBS patients may provide insight into hypertriglyceridemia and hypercholesterolemia commonly found in BBS. All identifications within this project must be further verified by a different proteomic technique, such as ELISA, and validated in a larger sample size. To determine whether any of these findings are specific to BBS, patients should be compared to BMI-matched controls. If proteomic profiling techniques were to be implemented again, samples should be assayed by mass spectrometry in replicate after multistep fractionation to increase the reliability and sensitivity of the findings. A targeted proteomic approach could be used to help validate findings and identify proteins not detected by mass spectrometry.

Chapter 5: Exploration of cellular pathways in BBS patient fibroblasts through implementation of complete mRNA transcriptomic and deep proteomic profiling

5.1 Background

5.1.1 Regulatory pathways for control of ciliogenesis

The ciliome is predicted to consist of up to 2500 cilia-interacting genes (Inglis et al. 2006; Lai et al. 2011; van Dam et al. 2017). In recent years, the identification of new proteins in the ciliome has dramatically complicated our view of cilia, but simultaneously has provided fascinating insight into this complex organelle and the pathology of ciliopathies (Davis and Katsanis 2012). The primary cilium is now known to be linked to cellular processes, including cell cycle control, cytoskeletal regulation, multiple signalling cascades, proteostasis, vesicle transport and endocytic recycling, cell motility, and autophagy (Kim et al. 2010; Christensen et al. 2008; Schneider et al. 2005; Takacs and Proikas-Cezanne 2016; Wheway et al. 2015; Boldt et al. 2016). These processes are thought to be involved in regulation and maintenance of ciliogenesis and cellular homeostasis, though it is currently uncertain whether these roles influence disease phenotype within the ciliopathy spectrum.

5.1.1.1 The relationship between cilia and the cell cycle

The processes involved in cell cycle and ciliogenesis are tightly interconnected. This is due to the dual function of the centrosome. In the cell cycle, the centrioles act as a microtubule organising centre (MTOC), thus controlling mitotic spindle formation and chromosome segregation during mitosis (Nigg and Stearns 2011). After cell division, the centrioles migrate to the apical cell surface where the older, mother centriole nucleates to form the basal body of the primary cilium. As a result of this multifaceted role, ciliogenesis and mitosis cannot occur simultaneously, where the transition between cell cycling and cilia assembly must be tightly regulated. Cilia are not a necessity for the cell cycle to occur, as there are cell types that do not ciliate, yet still proliferate, such as lymphocytes (Plotnikova et al. 2009). Studies in *Drosophila melanogaster* and mammalian cells have demonstrated that functional centrosomes are not necessary for mitotic spindle formation and mitosis (Mahoney et al. 2006; Uetake et al. 2007; Basto et al. 2006), however the centriole is an essential requirement for cilia assembly (Santos and Reiter 2008; Plotnikova et al. 2009).

The balance between ciliogenesis and cell cycle exit/entry is controlled by centrosome proteins, such as CCP110 and CEP97, and the regulation of kinases, including PLK1, NEK2, and TTBK2 (Izawa et al. 2015). CCP110 and CEP97 form a complex that is crucial for mitotic spindle function and subsequent cytokinesis. Depletion of either protein results in polyploidy, but also induces premature ciliogenesis in proliferating cells, demonstrating that CCP110-CEP97 acts as an inhibitory complex against cilia assembly (Spektor et al. 2007). CCP110 is thought to impede ciliogenesis through sequestration of CEP290, which thereby prevents recruitment of the small GTPase RAB8A to the basal body (Tsang et al. 2008); a step essential for cilia biogenesis. CCP110 is inactivated via phosphorylation by TTBK2, which removes it from the mother centriole, thus initiating ciliogenesis (Tsang and Dynlacht 2013). Ciliogenesis is also controlled through manipulation of microtubule dynamics. PLK1 and NEK2 induce cilia disassembly through phosphorylation of kinesins, KIF2A and KIF24, respectively (Miyamoto et al. 2015; S. Kim et al. 2015). Activation of these kinesins stimulates microtubule depolymerisation and induces deacetylation of tubulin. NEK2 and KIF24 also act to prevent reassembly of the cilium as the cell progresses into late G1/S phase (Kobayashi et al. 2011; Izawa et al. 2015). Furthermore, key signalling pathways, such as sonic hedgehog (SHH) pathway, can promote cell cycle progression and cilia disassembly to induce proliferation (Kenney et al. 2003). HH signalling upregulates expression of cyclin proteins, which activates cyclin-dependent kinases (CDKs), promoting entry into G1/S phase (Kenney et al. 2003; Ruiz I Altaba et al. 2002).

There have been many studies that have linked proteins crucial to cilia function with additional roles in cell cycle regulation. For example, IFT88 localised to the centrosome during cell cycle where it was shown to regulate G1/S transition in non-ciliated cells. Furthermore, overexpression of IFT88 prevented progression into S phase and induced apoptosis (Robert et al. 2007). Dual capabilities have also been demonstrated by the centrosome protein, pericentrin (encoded by *PCNT*), which is required for correct localisation of IFT proteins (Jurczyk et al. 2004), but also for mitotic spindle formation and entry into mitosis (Zimmerman et al. 2004). Moreover, Kim et al. (2004) elucidated that BBS4 is essential for recruitment of cargo to the centrioles and subsequent microtubule nucleation. Ablation of *BBS4* resulted in mislocalisation of pericentriolar material protein PCM1, disruption of the cell cycle, and increased apoptosis (Kim et al. 2004). This role was likely to be in addition to its capacity within the BBSome.

In cell culture, regular replenishment of growth factors and availability of space for expansion means that cells can cycle continuously until confluence at a higher density is reached. Therefore, in order to induce cell cycle arrest, cells in culture can be serum starved, which consequently stimulates ciliogenesis (Tucker et al. 1979). Maximal ciliation (>90% of cells) has been reported to occur after 24-48 hours of serum starvation in fibroblasts (Schneider et al. 2005; Christensen et al. 2013). Cilia start to form in early G1 phase, and are most abundant in G0. Resorption of the primary cilia occurs over 24 hours, with complete disassembly by the time the cell has committed to DNA replication and centrosome duplication, prior to mitosis (Tucker et al. 1979) (figure 5.1).

5.1.1.2 Regulation by the actin cytoskeleton

The cytoskeleton refers to three structural protein filaments; microtubules, intermediate filaments, and actin. In addition to microtubule dynamics, the actin cytoskeleton is thought to regulate cilia formation (Malicki and Johnson 2017). The evidence arises from research showing that ablation of ciliary proteins increased actin stress fibre formation, shortened cilia, and reduced cell motility (Hernandez-Hernandez et al. 2013). Furthermore, a web of actin filaments has been demonstrated to act as a docking site for the basal body at the apical surface of mouse primary culture epithelial cells (Pan et al. 2007). This interaction was dependent on the small GTPase RHOA, which regulates actin dynamics and formation of focal adhesions between cells. BBS proteins were shown to control RHOA activity, as ablation of *Bbs4* and *Bbs6* in mouse renal cells upregulated *RhoA* in the GTP active form, resulting in cytokinesis and motility defects (Hernandez-Hernandez et al. 2013).

RHOA activated actin remodelling has been proposed to act as a regulatory event for ciliogenesis, where actin acts to recruit the basal body to the apical surface (Pan et al. 2007). However, actin has also demonstrated an inhibitory role, where its presence prevents ciliary assembly. Silencing of the ARP2/3 complex required for actin branching, diminished actin nucleation and enhanced ciliogenesis, even in the absence of serum starvation (Kim et al. 2010; Cao et al. 2012). On the other hand, upregulation of actin destabilisation protein, gelsolin, enhanced cilia formation (J. Kim et al. 2015; Kim et al. 2010). It is thought that actin may inhibit ciliogenesis by operating as a mechanical barrier to prevent trafficking of vesicles to the basal body (Malicki and Johnson 2017; Kim et al. 2010). This suggests that actin aids basal body targeting to the apical membrane, and also prevents early onset of ciliogenesis before the basal body is attached.

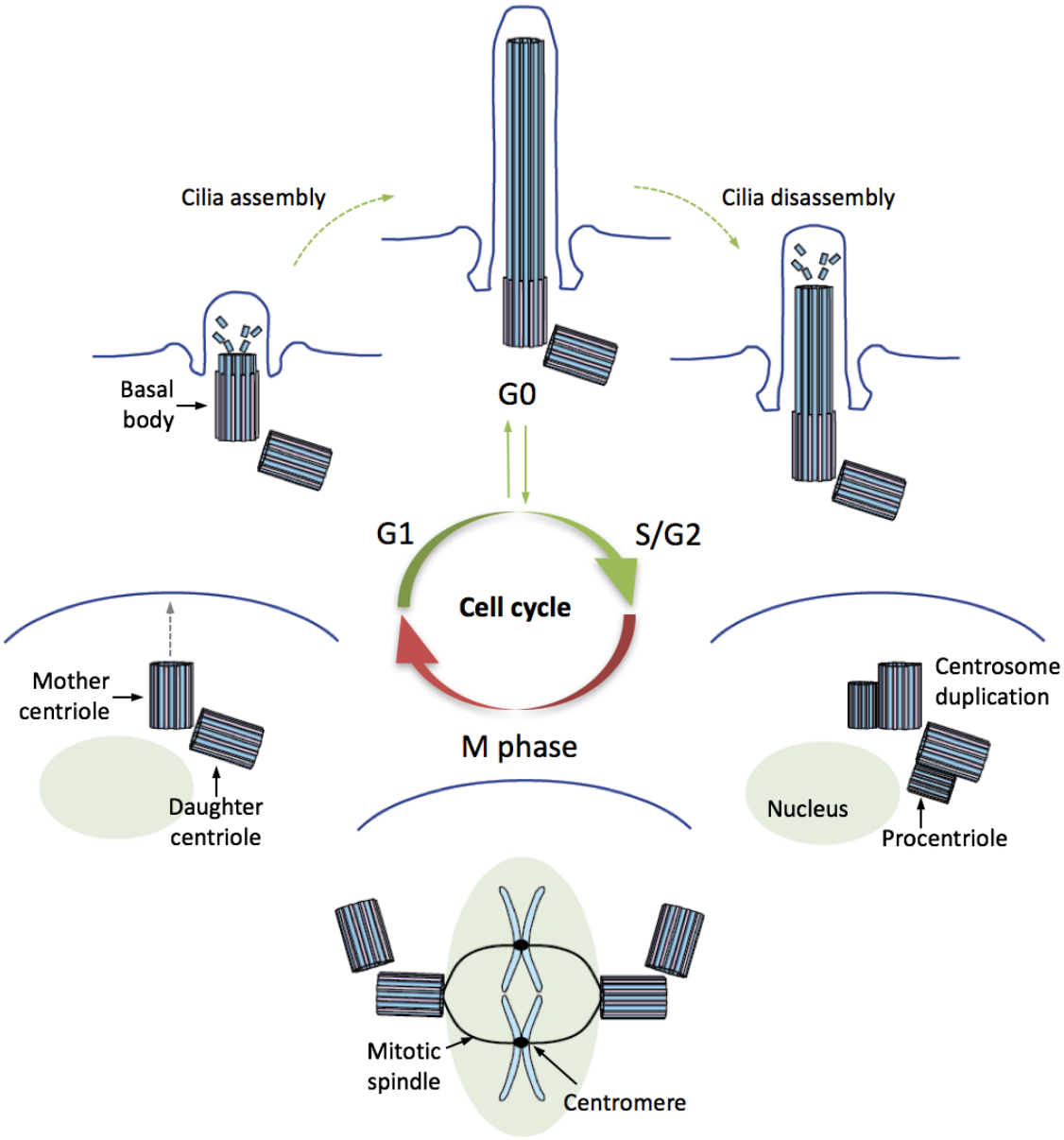

Figure 5.1 – Schematic representation of the centrosome's dual function linking the cell cycle with ciliogenesis. During early G1 of the cell cycle, the centrosome translocates to the apical membrane where the mother centriole nucleates to form the basal body of the primary cilium. Cilia reach their full length and are most abundant in G0 when cells are in a quiescent state. Upon re-entry into the cell cycle, the cilium disassembles prior to S phase entry. As DNA replicates in S, the centrioles duplicate, where the daughter centriole from the previous cycle matures to become the mother centriole for the next cycle. During mitosis, the mature centrosomes form mitotic spindles at each pole, which segregates the chromosomes for cytokinesis.

5.1.1.3 The ciliary function in proteostasis

The regulation of proteins by proteostasis is essential for cellular function, signalling, and maintenance. The primary cilium is a fundamental player in orchestrating the 26S proteasome, a

complex of 43 protein subunits that exhibit proteolytic activity. The proteasome consists of the 19S regulatory subunit and 20S catalytic subunit, which together creates the 26S proteasome, responsible for degradation of ~80-90% of a cell's proteins (Lilienbaum 2013). In addition to degradation of surplus or defective proteins, the proteasome also acts as a modulator for signalling pathways, such as HH and Wnt signalling. The proteasome is responsible for proteolytic cleavage of Gli2/3 from an active to a repressor form (Wang et al. 2007), in addition to switching Wnt signalling to β-catenin independent pathways by degradation of β-catenin (Aberle et al. 1997). Proteins are marked for degradation through addition of ubiquitin molecules by a three-step cascade, requiring respective activity of E1 ubiquitin-activating enzymes, E2 ubiquitin-conjugating enzymes, and E3 ubiquitin ligases (Lilienbaum 2013).

Members of the 19S regulatory proteasome localise to the basal body in wild type mouse epithelial fibroblasts (MEFs), whereas Psma5 of the 20S subunit, localised to the ciliary base and along the axoneme, indicating an association with IFT machinery. Furthermore, Psmd2, a member of the 19S regulatory complex, exhibited a direct interaction with Rpgrip1l, which was necessary for robust proteasome activity (Liu et al. 2014). Cilia genes, including *BBS4, OFD1, BBS7,* and *RPGRIP1L*, have been reported to be required for proper localisation of proteasome subunits at the centrosomes and basal body (Liu et al. 2014; Gascue et al. 2012; Gerdes et al. 2007; Gerhardt et al. 2015). Loss of *BBS4* resulted in aberrant proteasome targeting to centrosomes and the basal body, and an increase in cytoplasmic β-catenin (Gerdes et al. 2007; Liu et al. 2014). This close interaction with ciliary proteins and the proteasome may assist in determining a mechanism to explain perturbed planar cell polarity in BBS mutants (Ross et al. 2005; Cui et al. 2013). In relation to proteostasis regulation, the ubiquitination machinery is also thought to closely regulate ciliogenesis and cilia function (Kasahara et al. 2014; Wheway et al. 2012).

5.1.2 Strategy for pathway identification in BBS

The evidence from the literature presented so far within this chapter suggests that the pathways involved in ciliary function and maintenance are highly dynamic, requiring continuous interactions and signalling events with the rest of the cell. Investigation of ciliary pathways in cellular models of ciliopathies have rarely been studied on a global scale, and seldom with use of patient derived cells. Within this study, fibroblasts from 15 BBS patients manifesting the *BBS1* p.M390R variant have been utilised to identify cellular processes that are differentially regulated in response to serum starvation and by comparison with fibroblasts obtained from 6 healthy controls. This has been achieved through two methods of large-scale data acquisition; via RNA-

seq to attain the global mRNA transcriptome, and by isobaric tagging for relative and absolute quantitation (iTRAQ) mass spectrometry to achieve deep proteomic profiling.

5.1.2.1 iTRAQ proteomics

In chapter 4, untargeted label-free proteomics was utilised to identify proteins within plasma and urine samples. Label-free proteomics is a low cost quantitation method with rapid turnaround times. This makes it advantageous for large scale biomarker discovery projects, which requires multiple validation steps following candidate identification (Griffin et al. 2010). However, as each sample is separately processed and analysed, experimental variability may occur during sample preparation steps. Isotopically-labelled proteomics provides a method of reducing variation, as samples are distinctly labelled and pooled prior to processing. Furthermore, isotopic labelling offers a robust quantification method (Wang et al. 2012). The accuracy of label-free mass spectrometry relies on chromatographic reproducibility, and the sensitivity of the instrument, to measure the same ion precisely across different samples. In contrast, pooling of labelled samples means that each ion is identically assayed and quantified within the exact same experimental parameters. This reduces any external variability and dramatically increasing accuracy (Wiese et al. 2007; Timms and Cutillas 2010). Within this chapter, iTRAQ mass spectrometry was utilised for deep proteomic profiling.

iTRAQ labels samples with peptide-based tags that are chemically identical, but can be detected by mass spectrometry due to their different fragmentation patterns during MS/MS (Timms and Cutillas 2010). Peptides undergo derivatisation with the iTRAQ tags at amine groups of the N-terminal and lysine residues. Within the iTRAQ tag, there is a balancer group and a reporter ion component, where the two together have a total mass of 305.1 Da. During MS/MS fragmentation, the balancer group is lost, as it has a neutral charge, and the remaining reporter group is released (Aggarwal and Yadav 2016). Within this study, an 8-plex iTRAQ methodology was implemented, featuring 8 different tags with reporter groups that generate ions at mass to charge ratios (m/z) of 113, 114, 115, 116, 117, 118, 119, and 121, respectively, upon MS/MS fragmentation. As a result, both peptide identification and quantitation occurred at the MS/MS level. Since every peptide of each sample is labelled with their respective tag, relative quantitation occurs by direct comparison of ion intensity for the reporter groups of each peptide (figure 5.3).

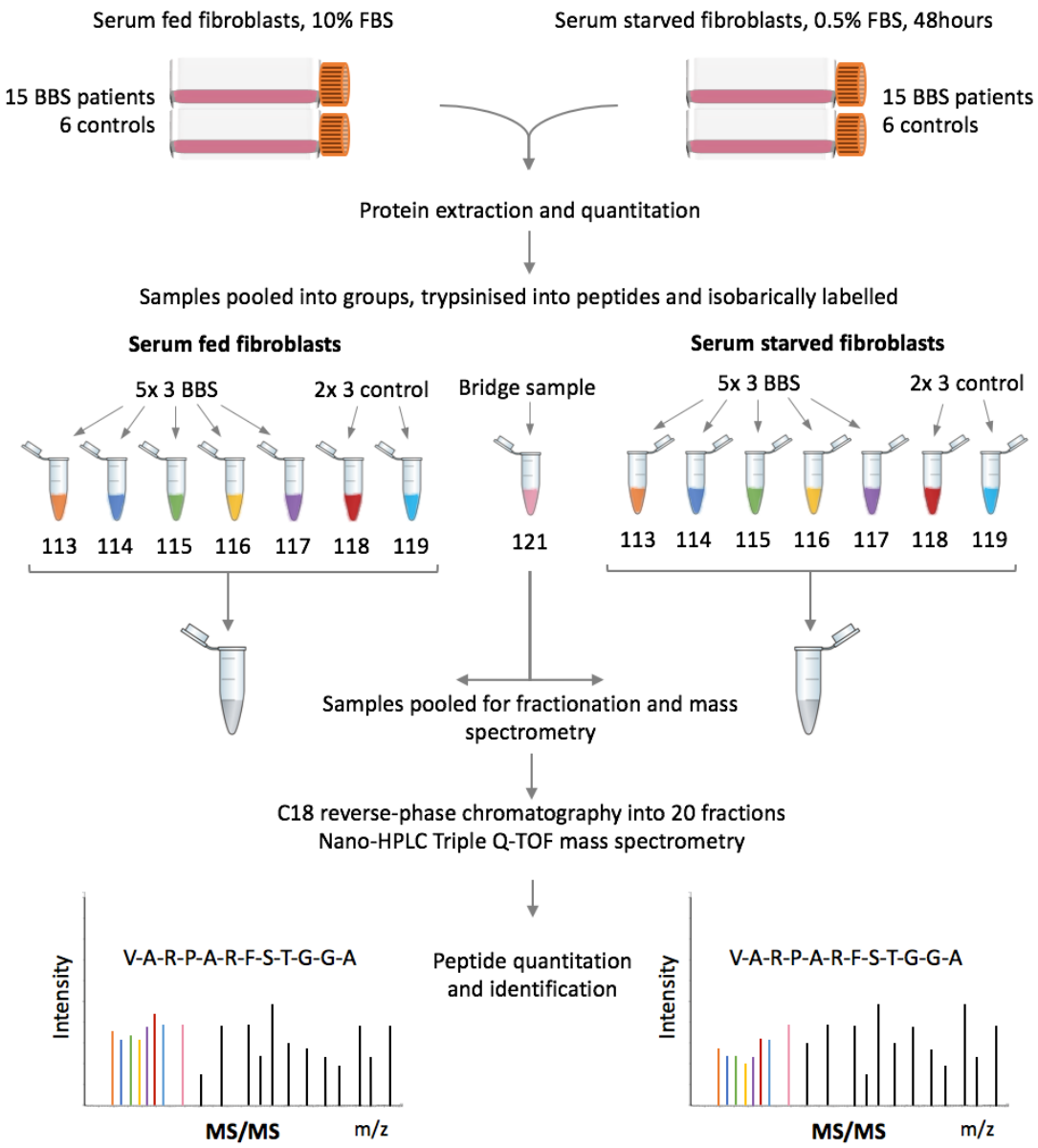

Figure 5.2 – Experimental and methodological design for iTRAQ mass spectrometry. Protein from fibroblasts was extracted and quantitated before pooling into 7 groups of 3 samples, each for serum fed and serum starved samples. An iTRAQ reagent tag was assigned to each pool. The bridge sample contained a complete pool of all samples and was added as the 8 sample to both runs. All tagged samples were pooled by serum fed or serum starved experiments for fractionation and mass spectrometry. Peptides were identified and quantified from MS/MS spectra. The peptide fragmentation ions, represented by the black peaks, were used for peptide identification. iTRAQ tags produced a fragment ion, represented by the coloured peaks, with a specific m/z. Peptides from a single sample were quantifiable by the intensity of the respective iTRAQ tag. The bridge sample was used to normalise the two iTRAQ runs for comparison between fed and starved groups.

As an 8-plex iTRAQ kit was utilised, in order to profile all 42 samples, they had to be pooled into groups of 3 and separated into two iTRAQ experiments. One 8-plex experiment was comprised of 21 serum starved fibroblast samples (15 patients, 6 controls) and the other for 21 fed fibroblasts samples. The 8th sample on each iTRAQ run consisted of a total pool of all samples, named a bridge sample. During analysis, this sample was used to normalise across the two experiments so that the serum starved experiment could be directly compared with the fed experiment. Deep proteomic profiling was ensured by fractionation into 20 fractions by reverse phase chromatography prior to high performance liquid chromatography (HPLC) coupled to triple quadrupole time of flight (Q-ToF) mass spectrometry (**figure 5.3**).

5.1.3 Aims of investigation

The 15 patients in this study all carry the p.M390R mutation in *BBS1*, the pathobiology of which is generally unknown. Investigation of the crystal structure of the BBS1:ARL6^{GTP} interaction demonstrated that p.M390R reduces the BBS1's affinity for ARL6, which is required for BBS1 recruitment to the basal body (Mourão et al. 2014). However, outside of the BBSome, the function of BBS1 remains elusive. Various ciliopathy proteins, including those that function within the BBSome, have additional cellular duties (Hernandez-Hernandez et al. 2013; Gascue et al. 2012; Liu et al. 2014). With this knowledge, this study aimed to identify differentially regulated pathways that may help determine the pathology associated with the *BBS1* p.M390R mutation. To our knowledge, this was the first study of its kind to globally identify differential expression and altered pathways in such patient-derived cultures.

Fibroblasts from 15 BBS patients and 6 controls were cultured under 2 conditions; to promote cell cycle progression (serum fed with 10% fetal bovine serum (FBS)), and to promote ciliogenesis (serum starved with 0.5% FBS for 48 hours). The purpose of this was to identify differentially expressed genes and proteins that would be up or downregulated as a consequence of ciliogenesis. These could also be compared between disease and control groups. The analysis strategy was divided into three sections, which aimed to identify differentially expressed candidates, significantly enriched gene sets, and differentially regulated pathways, both for transcriptomic and proteomic datasets (**figure 5.4**). Further to this, different datasets were integrated to find pathways that were interconnected between transcriptomic and proteomic datasets, thereby enhancing the power of the analysis. These aims were implemented through use of various computational databases and analytical tools. Of note, unpublished experimentation from our lab has revealed no phenotypic and morphological dissimilarities

between fibroblast cultures obtained from controls and patients carrying the *BBS1* p.M390R variant.

The study design allowed a 3-way comparison strategy. Firstly, serum starved BBS cells were compared with starved control cultures to identify differences in ciliated cells; secondly, serum fed BBS fibroblasts were analysed against fed control cells to find differences in proliferating cells; and finally, BBS starved cells were compared to BBS fed cells, with an equivalent comparison of control cells. In this way, the study isolates BBS specific findings in response to starvation. Overall, this analytical strategy has instigated the identification of pathways and gene sets linked to BBS pathology. This has included disease relevant findings related to obesity, as well as processes involved in regulation of the primary cilium such as centrosome function, RHO GTPase activity, and cilium assembly.

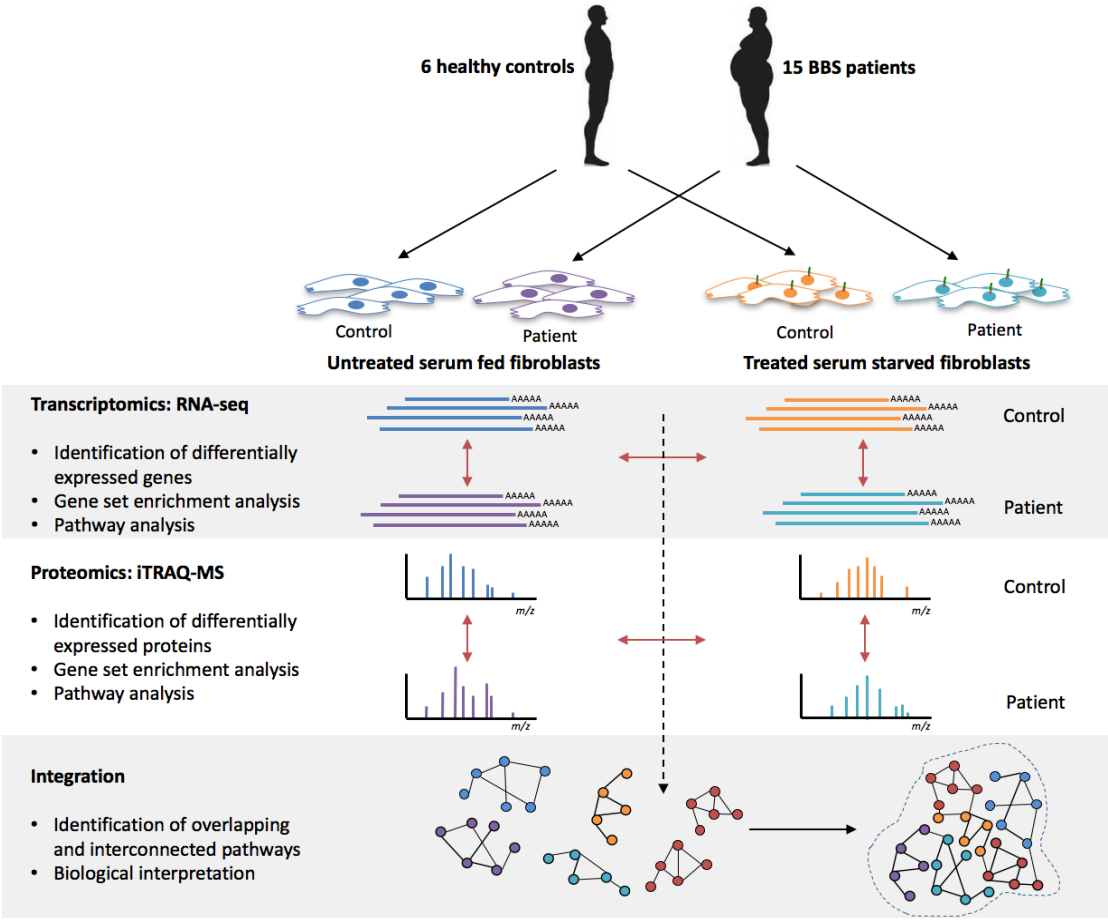

Figure 5.3 – Experimental design and analysis strategy for chapter 5. RNA and protein were extracted from serum fed and serum starved fibroblasts of 15 BBS patients and 6 controls. RNA and protein were compared for differentially expressed transcripts/proteins between treatment groups and disease status i.e. patient vs control, serum starved vs serum fed cells. Pathway and gene set enrichment analyses were utilised for each comparison. Finally, transcriptomic and proteomic data from BBS cells were integrated to identify interconnecting differentially regulated networks.

5.2 Results

5.2.1 Experimental design and quality control

5.2.1.1 Ciliation of fibroblasts

In cultured cells, ciliogenesis was induced via serum starvation by decreasing media FBS content from 10% to 0.5% for 48 hours. The response to starvation was first tested through qPCR and immunocytochemical staining to ensure that there was a sufficient difference to implement the experimental design. Ciliogenesis studies were carried out using 5 primary fibroblast cultures, randomly selected from the cohort. Gene expression analysis was completed by qPCR, which established that serum starvation significantly increased expression of *BBS1* (relative fold change (RFC)= 7.35; fed mean ΔC_T= 6.81, 95% CI = 5.95, 7.67; starved mean ΔC_T= 3.93, 95% CI= 3.40, 4.46; p= 0.0005), *BBS2* (RFC= 5.16; fed mean ΔC_T= 7.22, 95% CI= 6.51, 7.92; starved mean ΔC_T= 4.85, 95% CI= 4.22, 5,48; p=0.0012), *BBS8* (RFC= 3.64; fed mean ΔC_T=6.58, 95% CI= 5.77, 7.38; starved ΔC_T=4.72, 95% CI= 3.94, 5.49; p= 0.0115), and *CEP290* (RFC= 4.81; fed mean ΔC_T= 8.53, 95% CI= 7.78, 9.28; starved mean ΔC_T= 6.26, 95% CI= 5.53, 7.00; p= 0.0028) transcripts (**figure 5.4A**).

Protein analysis of serum fed and starved cells was determined through immunocytochemical staining of acetylated tubulin. There was an increase in ciliated cells after 48 hours serum starvation (**figure 5.4B**). Furthermore, the percentage of ciliated cells for fed and starved fibroblasts was established using an ImageJ macro for cilia counting. The frequency of observed cilia per image was divided by the number of nuclei. 30 images were taken over 5 cultures for each treatment group and the mean calculated (**figure 5.4C**). There was a significant increase in ciliated cells after serum starvation compared serum fed cells (starved mean= 78.0%, 95% CI= 70.6, 85.3; fed mean= 33.8%, 95% CI= 25.2, 42.4; p<0.0001). Despite this large difference, fed cells demonstrated a greater number of ciliated cells than expected. It was noted that the majority of ciliated cells in fed cultures were found in fibroblasts grown to a higher confluency, which would also have induced cell cycle arrest. Therefore, following this method development experiment, serum fed cell cultures used for RNA-seq and iTRAQ proteomic experiments were controlled for cell cycling by fixation at 70-80% confluency.

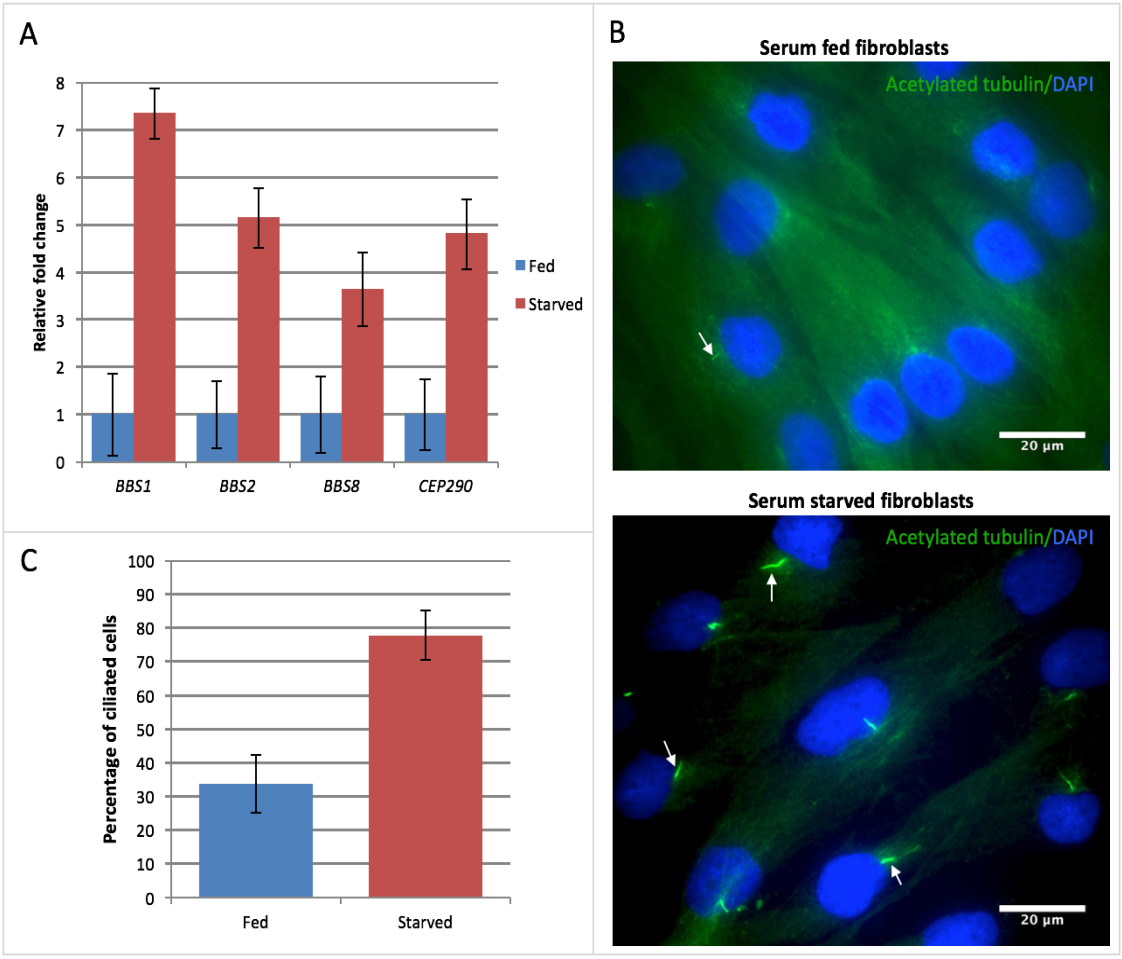

Figure 5.4 – Increase in ciliogenesis induced by serum starvation. A) Relative fold change of gene expression of basal body genes *BBS1, BBS2, BBS8* and *CEP290* upon starvation of fibroblasts. Error bars show 95% confidence intervals (CI) established from 5 biological repeats. Samples were normalised to reference gene *ACTB*. B) Immunocytochemical analysis of cilia by staining with acetylated tubulin (white arrows) for serum fed (top) and serum starved cells (bottom), where nuclei were stained with DAPI. C) The percentage of ciliated fibroblasts with and without starvation, averaged over 5 fibroblasts cultures. Error bars show 95% CI, calculated from 30 images per treatment over 5 cell cultures.

5.2.1.2 Quantification and quality control of RNA

Prior to RNA-seq, total RNA was assessed for quality and concentration using the Agilent 4200 TapeStation. An RNA integrity number (RIN) was provided by the TapeStation output as a numerical assessment of the quality of RNA. A score of 10 indicates intact RNA whereas 1 is completely degraded; a recommended threshold for good quality is a RIN score above 8. All samples that were sequenced had an RIN of 8 or above. The results featuring RNA extracted from serum starved (treated) and fed (untreated) fibroblasts from 15 patients and 6 controls are displayed in **table 5.1**.

Patient ID	Condition	Concentration (ng/µl)	Yield (µg)	RIN	28S/18S
BBS001	Starved (treated)	36.8	1.104	9.4	3.2
BBS002	Starved (treated)	66	1.98	9.8	2.2
BBS003	Starved (treated)	53.1	1.593	10	2.4
BBS004	Starved (treated)	57.4	1.722	8.9	2.8
BBS005	Starved (treated)	45.6	1.368	9.5	3.1
BBS006	Starved (treated)	64.4	1.932	9.9	2.5
BBS007	Starved (treated)	98.6	2.958	9.9	3.4
BBS008	Starved (treated)	63	1.89	10	3
BBS009	Starved (treated)	66.1	1.983	10	3.1
BBS010	Starved (treated)	83.3	2.499	10	3.1
BBS011	Starved (treated)	31.4	0.942	9.4	2.4
BBS012	Starved (treated)	123.0	3.69	9.9	2.6
BBS013	Starved (treated)	46.8	1.404	9.3	2.8
BBS014	Starved (treated)	49.6	1.488	8.9	3.3
BBS015	Starved (treated)	88.0	2.64	10	2.6
CTRL019	Starved (treated)	75.5	2.265	9.8	3.1
CTRL020	Starved (treated)	98.6	2.958	9.8	3.3
CTRL031	Starved (treated)	50.2	1.506	8.9	3.1
CTRL032	Starved (treated)	306	9.18	10	3.2
CTRL033	Starved (treated)	19.4	0.582	9.4	2
CTRL030	Starved (treated)	138	4.14	10	2.8
BBS001	Fed (untreated)	79.8	2.394	10	2.8
BBS002	Fed (untreated)	60.2	1.806	10	2.7
BBS003	Fed (untreated)	55.2	1.656	9.4	3.8
BBS004	Fed (untreated)	74.4	2.232	9.4	2.7
BBS005	Fed (untreated)	89.2	2.676	10	2.9
BBS006	Fed (untreated)	19.94	0.5982	9.4	3.3
BBS007	Fed (untreated)	47.6	1.428	9	3.7
BBS008	Fed (untreated)	114	3.42	10	3
BBS009	Fed (untreated)	21.2	0.636	9.5	2.7
BBS010	Fed (untreated)	20.8	0.624	8.6	2.1
BBS011	Fed (untreated)	153.0	4.59	10	3.1
BBS012	Fed (untreated)	181.0	5.43	10	3.1
BBS013	Fed (untreated)	135.0	4.05	10	3
BBS014	Fed (untreated)	95.6	2.868	10	3.6
BBS015	Fed (untreated)	354.0	10.62	8.3	2.2
CTRL019	Fed (untreated)	116	3.48	9.9	3.3
CTRL020	Fed (untreated)	218	6.54	10	3
CTRL031	Fed (untreated)	68.4	2.052	8	3.1
CTRL032	Fed (untreated)	127	3.81	9.9	3.9
CTRL033	Fed (untreated)	267	1.716	9.4	3.1
CTRL030	Fed (untreated)	129	3.87	10	3.5

Table 5.1 – Sample quantitation and quality control results from RNA extracted from serum fed or serum starved fibroblasts.

RNA quality was additionally assessed by the ribosomal 28S/18S ratio. This was calculated by electrophoretic separation of the RNA within the TapeStation chip and comparison of the molecular weights against the ladder. There should be two distinct peaks on the read out at ~2000nt and ~4000nt, which represent the 18S and 28S ribosomal subunits, respectively. The ribosomal ratio is calculated from the area of the 28S peak divided by the area of the 18S peak. Small RNAs can also be detected with a peak ~200nt. An electropherogram representative of RNA with a high integrity score is presented in **figure 5.5A**. Ideally, the base of the 28S peak should be twice the width of 18S and have greater signal intensity. Less distinct or shifted peaks suggest RNA degradation **(figure 5.5B)**. The electrophoresis gel image for each sample is displayed in **figure 5.6**.

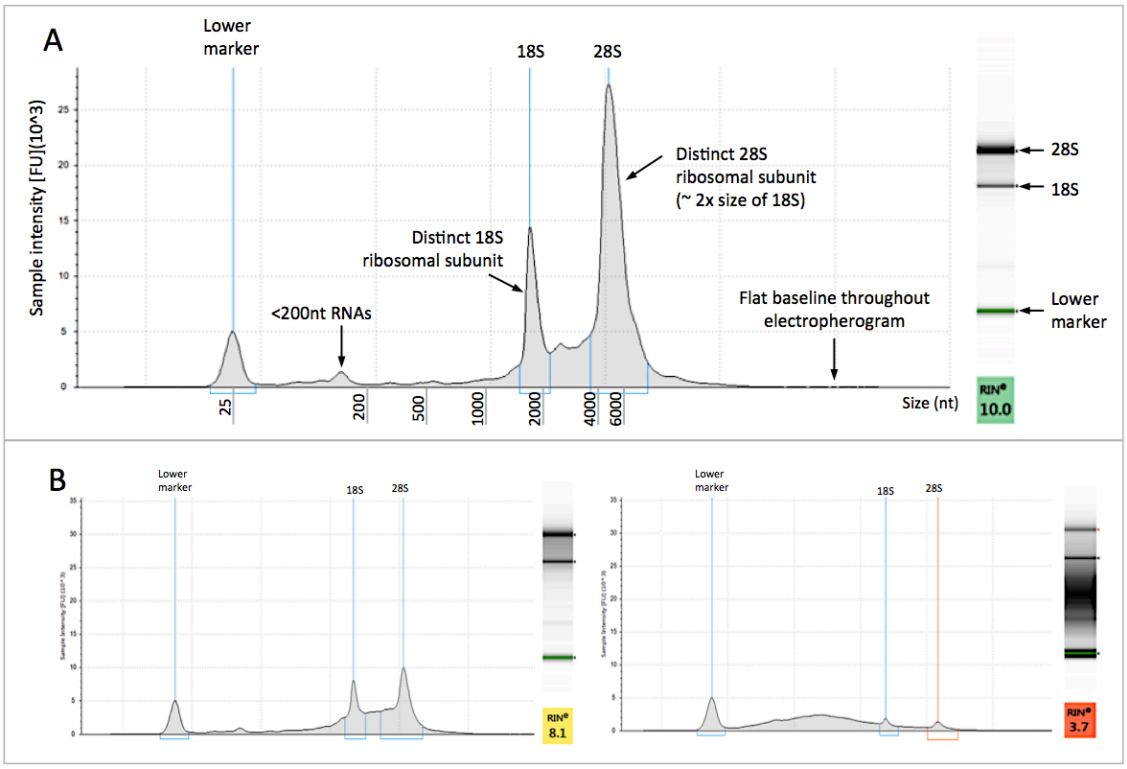

Figure 5.5 – Electropherogram traces of RNA run on the TapeStation chip. A) RNA with high integrity score (RIN 10), demonstrating how to recognise quality from the electropherogram. B) Illustration of RNA with reduced integrity. A sample with RIN 8.1 can still be sequenced, whereas a score of 3.7 shows significant degradation.

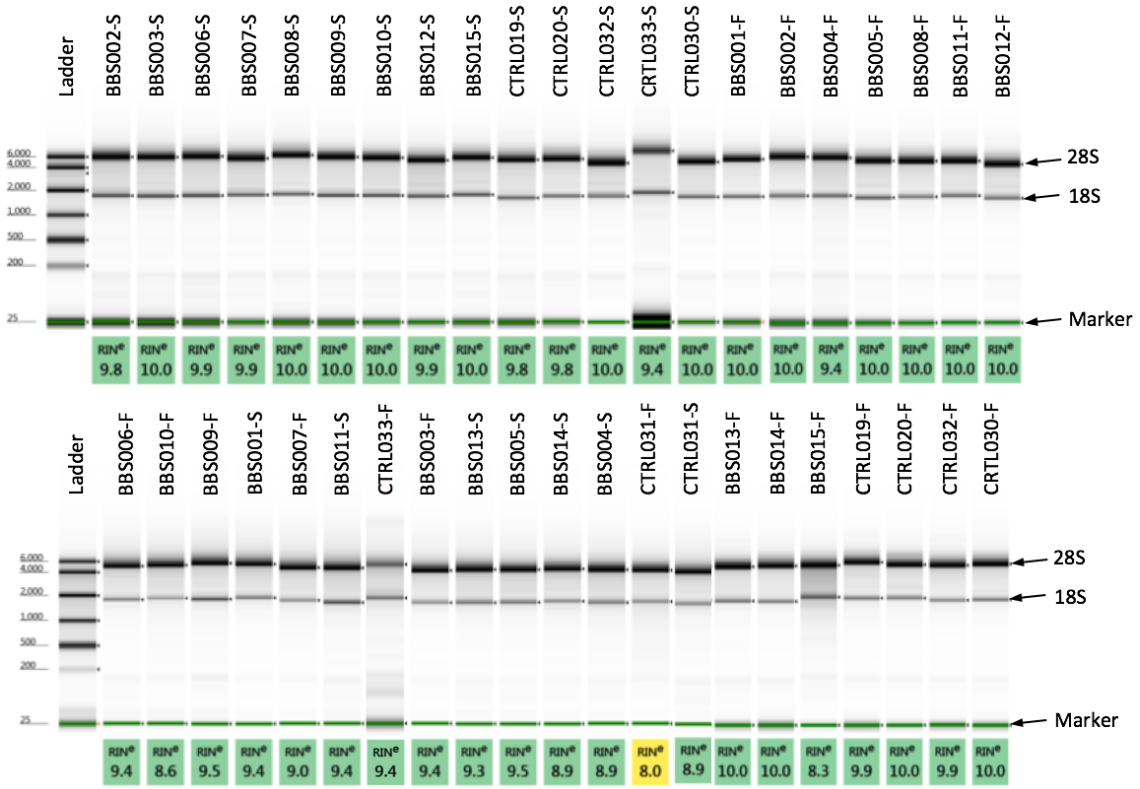

Figure 5.6 – Single lane electrophoresis outputs from TapeStation showing the bands of ribosomal subunits, 28S and 18S, and corresponding RIN for each sample (where S and F indicate starved and fed samples, respectively).

5.2.1.3 Quantitation of protein from fibroblasts

Serum starved and serum fed fibroblasts from 15 patients and 6 controls sent to Beijing Genome Institute (Hong Kong, China) for iTRAQ mass spectrometry. Protein extracted from fibroblasts by homogenisation was quantitated by Bradford assay, where absorbance was measured at 595nm and a standard curve was generated from BSA standards (**figure 5.7**). The quantity of protein for each sample was calculated using the equation of the standard curve (**table 5.2**). Samples were then pooled into sets of 3, forming 7 pools for the serum starved experiment, and 7 pools for the serum fed experiment. A complete pool of all samples made the 8th sample, which was run on both mass spectrometry runs. This sample was used to normalise samples across the two runs to ensure they were comparable.

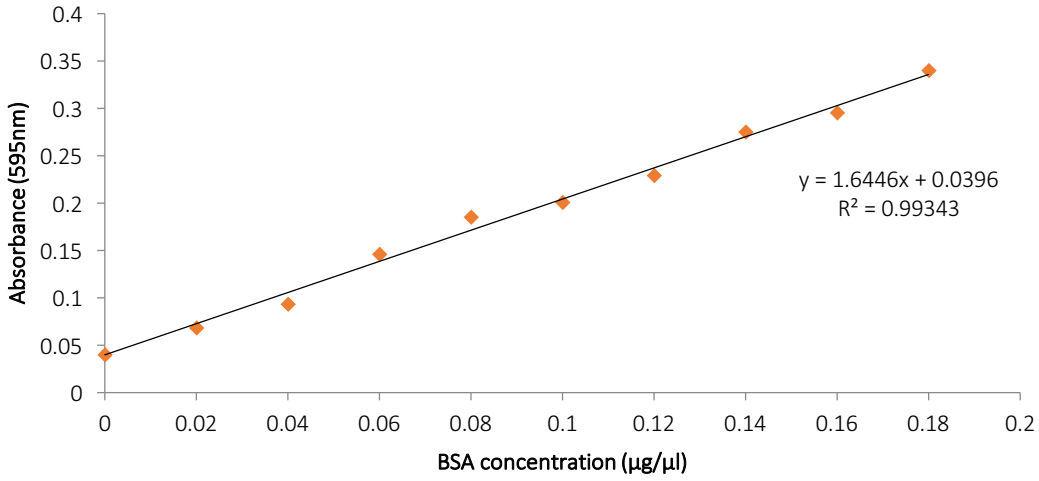

Figure 5.7 – Standard curve generated by Bradford assay that was used for quantitation of fibroblast-derived protein.

Sample Name	Sample name analysis	Concentration (µg/µl)	Yield (µg)	Pooling info	Run info
BBS001	Starved (treated)	1.49	447.82	Pool 1	Run 1
BBS002	Starved (treated)	1.48	443.45	Pool 1	Run 1
BBS003	Starved (treated)	1.74	520.55	Pool 1	Run 1
BBS004	Starved (treated)	1.87	559.82	Pool 2	Run 1
BBS005	Starved (treated)	1.52	456.73	Pool 2	Run 1
BBS006	Starved (treated)	2.29	686.00	Pool 2	Run 1
BBS007	Starved (treated)	2.08	624.00	Pool 3	Run 1
BBS008	Starved (treated)	1.44	430.55	Pool 3	Run 1
BBS009	Starved (treated)	1.54	463.27	Pool 3	Run 1
BBS010	Starved (treated)	1.09	326.18	Pool 4	Run 1
BBS011	Starved (treated)	1.39	416.00	Pool 4	Run 1
BBS012	Starved (treated)	1.68	502.55	Pool 4	Run 1
BBS013	Starved (treated)	2.46	736.73	Pool 5	Run 1
BBS014	Starved (treated)	1.68	502.55	Pool 5	Run 1
BBS015	Starved (treated)	1.42	426.55	Pool 5	Run 1
CTRL019	Starved (treated)	1.50	450.91	Pool 6	Run 1
CTRL020	Starved (treated)	1.34	402.00	Pool 6	Run 1
CTRL030	Starved (treated)	1.18	355.27	Pool 6	Run 1
CTRL031	Starved (treated)	1.20	361.09	Pool 7	Run 1
CTRL032	Starved (treated)	1.47	440.36	Pool 7	Run 1
CTRL033	Starved (treated)	1.33	400.36	Pool 7	Run 1
BBS001	Fed (untreated)	1.34	401.64	Pool 1	Run 2
BBS002	Fed (untreated)	1.26	376.55	Pool 1	Run 2
BBS003	Fed (untreated)	1.24	370.55	Pool 1	Run 2
BBS004	Fed (untreated)	1.51	453.64	Pool 2	Run 2

BBS005	Fed (untreated)	1.58	474.36	Pool 2	Run 2
BBS006	Fed (untreated)	1.53	459.09	Pool 2	Run 2
BBS007	Fed (untreated)	1.44	430.55	Pool 3	Run 2
BBS008	Fed (untreated)	1.46	438.55	Pool 3	Run 2
BBS009	Fed (untreated)	1.63	488.00	Pool 3	Run 2
BBS010	Fed (untreated)	1.53	458.73	Pool 4	Run 2
BBS011	Fed (untreated)	1.52	457.09	Pool 4	Run 2
BBS012	Fed (untreated)	1.54	461.82	Pool 4	Run 2
BBS013	Fed (untreated)	1.58	472.55	Pool 5	Run 2
BBS014	Fed (untreated)	1.47	440.36	Pool 5	Run 2
BBS015	Fed (untreated)	1.60	480.18	Pool 5	Run 2
CTRL019	Fed (untreated)	1.60	478.55	Pool 6	Run 2
CTRL020	Fed (untreated)	1.25	374.18	Pool 6	Run 2
CTRL030	Fed (untreated)	1.70	508.91	Pool 6	Run 2
CTRL031	Fed (untreated)	1.57	471.82	Pool 7	Run 2
CTRL032	Fed (untreated)	1.22	365.45	Pool 7	Run 2
CTRL033	Fed (untreated)	1.02	365.45	Pool 7	Run 2
Bridge Protein	Bridge Protein			Pool 8 – all samples	

Table 5.2 – Protein concentrations and yields from sample quantitation by Bradford assay, where pooling and run information for each sample is also provided.

5.2.2 Analysis of RNA-seq data

5.2.2.1 Identification of differentially expressed genes

From RNA-seq of 42 fibroblast samples, a total of 16,336 transcripts were detected (supplemental data, **table S.2, table S.3**), of which 14,573 were protein-coding genes. These were analysed to determine differentially expressed (DE) transcripts between comparisons based on treatment, starved vs. fed for BBS fibroblasts and starved vs. fed for control fibroblasts (**figure 5.8 A-B**); and comparisons based on disease status, BBS vs. control for starved fibroblasts and BBS vs. control for fed fibroblasts (**figure 5.8 C-D**). DE transcripts were determined through T-test with Benjamini-Hochberg adjustment for multiple correction, which generated a false discovery rate (FDR) for each transcript. Transcripts were considered differentially expressed with an FDR less than 0.05 (supplemental data, **tables S.4-S.8**).

The number of up and downregulated transcripts are found in **table 5.3**. Comparisons between starved and fed fibroblasts yielded the greatest number of DE transcripts, with 5968 and 2167 for BBS and control comparisons, respectively. By contrast, there were far fewer DE transcripts upon comparison between BBS and control cells; only 29 DE transcripts were identified for starved BBS vs control and 817 for fed BBS vs control.

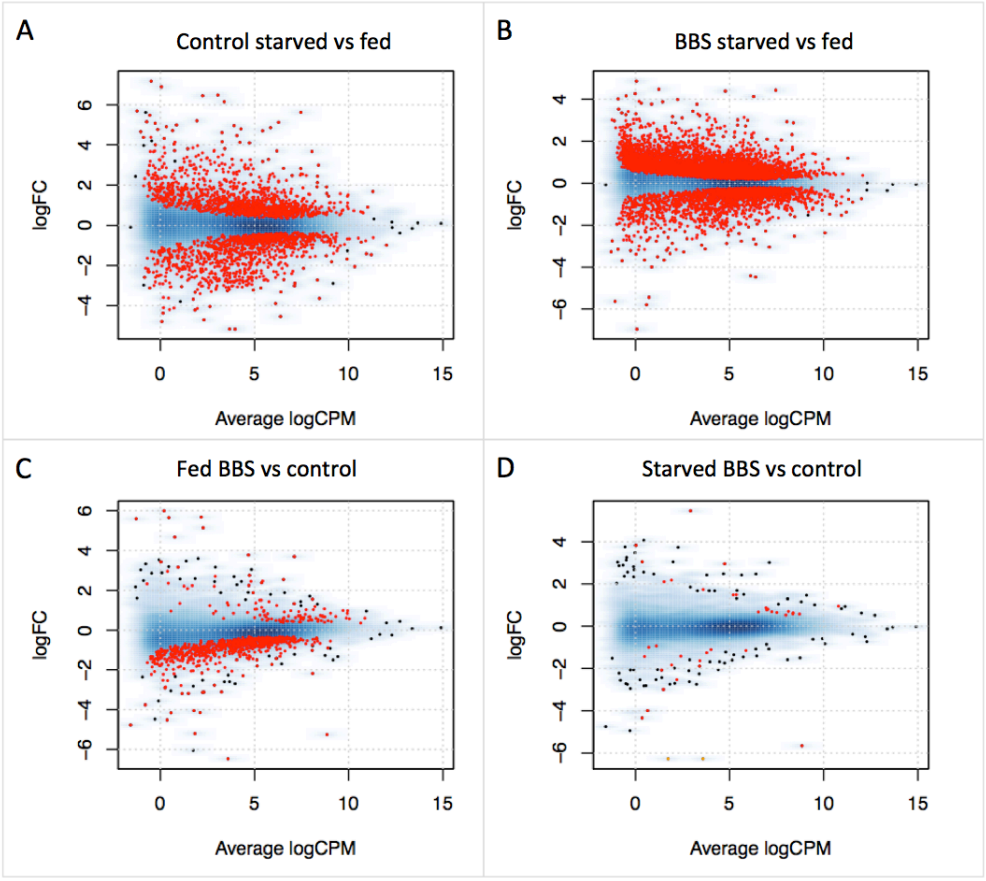

Figure 5.8 – MA plots displaying significantly differentially expressed transcripts (red) for each comparison. Transcripts are plotted by the average normalised read counts per million (CPM) in log2 form against the log2 of fold change (FC). Transcripts were defined as differentially expressed with FDR <0.05. A-B) Distribution of DE genes within treatment comparisons. C-D) Spread of DE transcripts within disease comparisons.

Table 5.3	Downregulated	Not differentially expressed	Upregulated	TOTAL	TOTAL DE transcripts
Control starved vs fed	1003	14,169	1164	16,336	2167
BBS starved vs fed	1852	10,368	4116	16,336	5968
Fed BBS vs control	690	15,522	124	16,336	817
Starved BBS vs control	9	16,307	20	16,336	29

Table 5.3 – The number of transcripts detected as differentially expressed for each experimental comparison.

5.2.2.2 Gene set enrichment analysis

The first challenge with a large dataset, such as the one generated in this study, was to reduce the size of the data into more manageable groups. This is achievable through pathway analysis, which not only allows simplification, but also uses knowledge driven analysis to rearrange the

data into biologically meaningful processes (Khatri et al. 2012). Gene set enrichment analysis (GSEA) with Hallmark gene sets was used to probe the data. The advantage of this technique is two-fold. Firstly, the Hallmark gene sets encompass a broad range of biological, non-overlapping processes, allowing a refined and concise analysis. Secondly, the input for GSEA is the entire transcriptome dataset, which is not restricted by a p-value threshold. Therefore, it can reliably be used on experiments such as starved BBS vs control, where only a handful of genes were identified after multiple correction was applied.

GSEA was applied to all protein-coding genes within the RNA-seq experiment, pre-ranked by fold change. 48 Hallmark gene sets were identified that were considered significantly enriched with FDR<0.25 in at least one comparison (supplemental data, tables S.17-S.24). Gene sets that may be of interest for further investigation are highlighted by yellow or blue boxes in the heatmap showing significantly positively or negatively enriched gene sets (figure 5.9). Although the Hallmark gene sets were derived to allow a broad overview of general pathways, there were a handful of significantly enriched gene sets that were relevant to primary cilia or BBS pathology. Two gene sets were found to be negatively enriched in starved BBS vs control, namely notch signalling and cholesterol homeostasis (normalised enrichment score (NES)=-1.336, FDR=0.1415; NES=-1.928, FDR=0.0006, respectively). Notch signalling is a pathway, that like hedgehog (HH) and Wnt signalling, is thought to be partly orchestrated by the primary cilium (Ezratty et al. 2011; Leitch et al. 2014). The cholesterol homeostasis gene set may be worth investigation as hypercholesterolemia is commonly present in BBS patients (Forsythe et al. 2015). An additional finding related to obesity in BBS was a positive enrichment of the adipogenesis gene set in the fed BBS vs control comparison (NES=1.682, FDR=0.0149). This was also positively enriched in starved BBS vs control, but this finding was not significant to FDR<0.25 (NES=1.083; FDR=0.6442) (figure 5.10). This suggests that a shift to genes that are upregulated during adipogenesis genes may be specific to serum fed conditions.

The statistically enriched gene sets of adipogenesis, notch signalling, and cholesterol homeostasis were investigated in more detail. The genes that contribute to the enrichment of the gene set is called the leading edge. These are all of the genes that are present before (if positively enriched) or after (if negatively enriched) the peak of the enrichment score graph (figure 5.10). The gene expression levels (log2 counts per million (CPM)) were plotted for all genes of the leading edge for adipogenesis, notch signalling, and cholesterol homeostasis (figure 5.11).

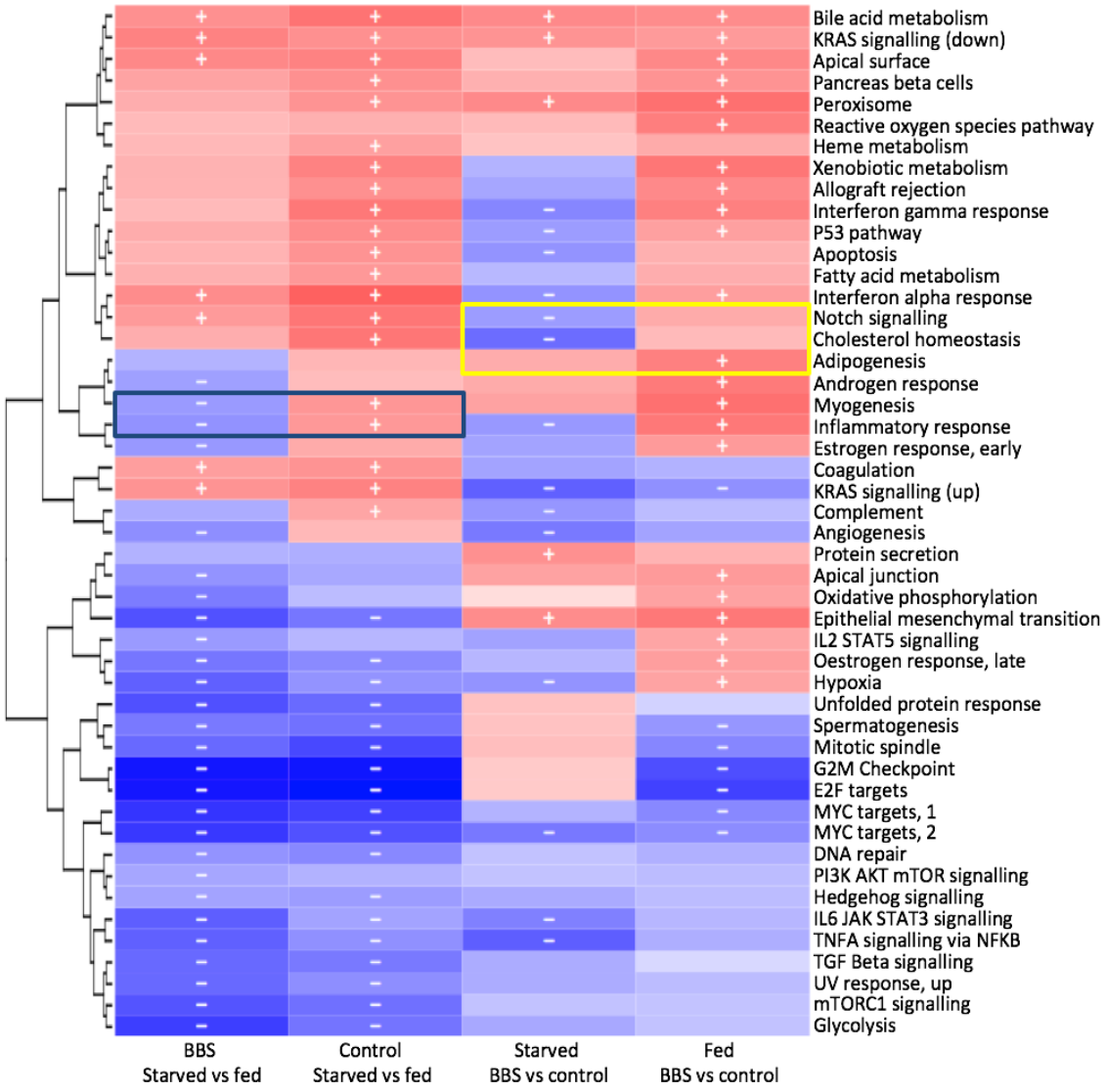

Figure 5.9 – Heatmap featuring results of enriched Hallmark gene sets generated by GSEA. Significant negatively or positively enriched pathways where FDR<0.25 are marked with – or +, respectively. As the FDR value decreases (i.e. becomes more significant) by each factor of 10, the colour intensity increases. The yellow box illustrates gene sets that are biologically relevant to primary cilia or BBS pathology. The blue box highlights gene sets that are discordantly enriched between different starved vs fed experiments.

79 genes contributed to the leading edge of the adipogenesis gene set in fed BBS vs control experiment (**figure 5.11A**). Of interest, the gene with the greatest fold change was *LEP* (log2(FC)= 2.38), which encodes the satiety hormone known to be elevated in BBS, leptin. Other genes upregulated in the leading edge include those required for adipocyte differentiation, *CD36, PPARG, SLC27A1,* and *MTCH2*, and fatty acid biosynthesis, *SLC25A1, ECH1, ACAA2, ACADS*, and *SCP2*. Of note, *PPARG* and *ECH1* overlap with the cholesterol homeostasis gene set, which is negatively enriched in starved BBS vs control cultures. Other cholesterol homeostasis genes that

are downregulated in starved BBS fibroblasts include *SREBF1, ACAT2, CYP51A1*, and *LDL*, involved in lipid metabolism (**figure 5.11B**). This may suggest a mechanism for high adiposity in BBS patients, where there is an imbalance between adipose tissue biogenesis and lipid metabolism. Finally, there were 8 genes associated with a negative enrichment of Notch signalling, including downstream effectors *(HEYL* and *HES1*) and processors of Notch signalling (*DLL1, PSEN2* and *DTX4*) (**figure 5.11C**).

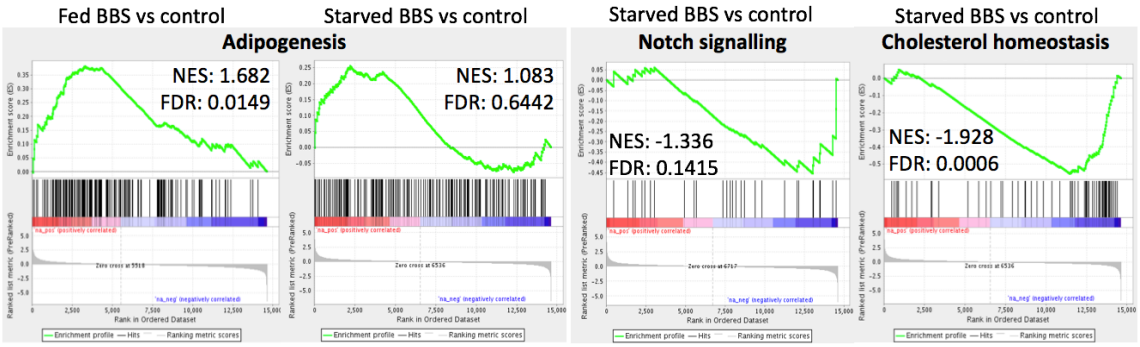

Figure 5.10 – GSEA profiles for enriched Hallmark gene sets that may link to BBS pathobiology. Enrichment is considered significant when FDR<0.25.

5.2.2.3 Comparison of RNA-seq pathways between starved vs fed

The next step in the analysis was to identify gene sets that were discordantly enriched between starved vs fed experiments. It was hypothesised that any similarities would be in response to starvation and cell cycle arrest, and any differences uncovered may be due to the *BBS1* defect in the patient cells. There were two differentially enriched pathways; myogenesis (BBS starved vs fed: NES = -1.360, FDR = 0.0519; control starved vs fed: NES = 1.341, FDR = 0.0731), and inflammatory response (BBS starved vs fed: NES = -1.412, FDR = 0.0456; control starved vs fed: NES = 1.385, FDR = 0.0555), which were both negatively enriched in BBS starved vs fed and positively enriched in control starved vs fed (**figure 5.9**). A link between myogenesis and BBS pathobiology has not yet been reported, so this may require further experimentation to determine the significance of this finding. Components of the acute phase inflammatory response, such as C-reactive protein, are raised in BBS patients as a consequence of secondary responses to obesity (Mirhafez et al. 2016). Therefore, future research would be needed to determine whether there is a biological explanation for a negative enrichment of inflammatory response in starved BBS cells, compared to fed.

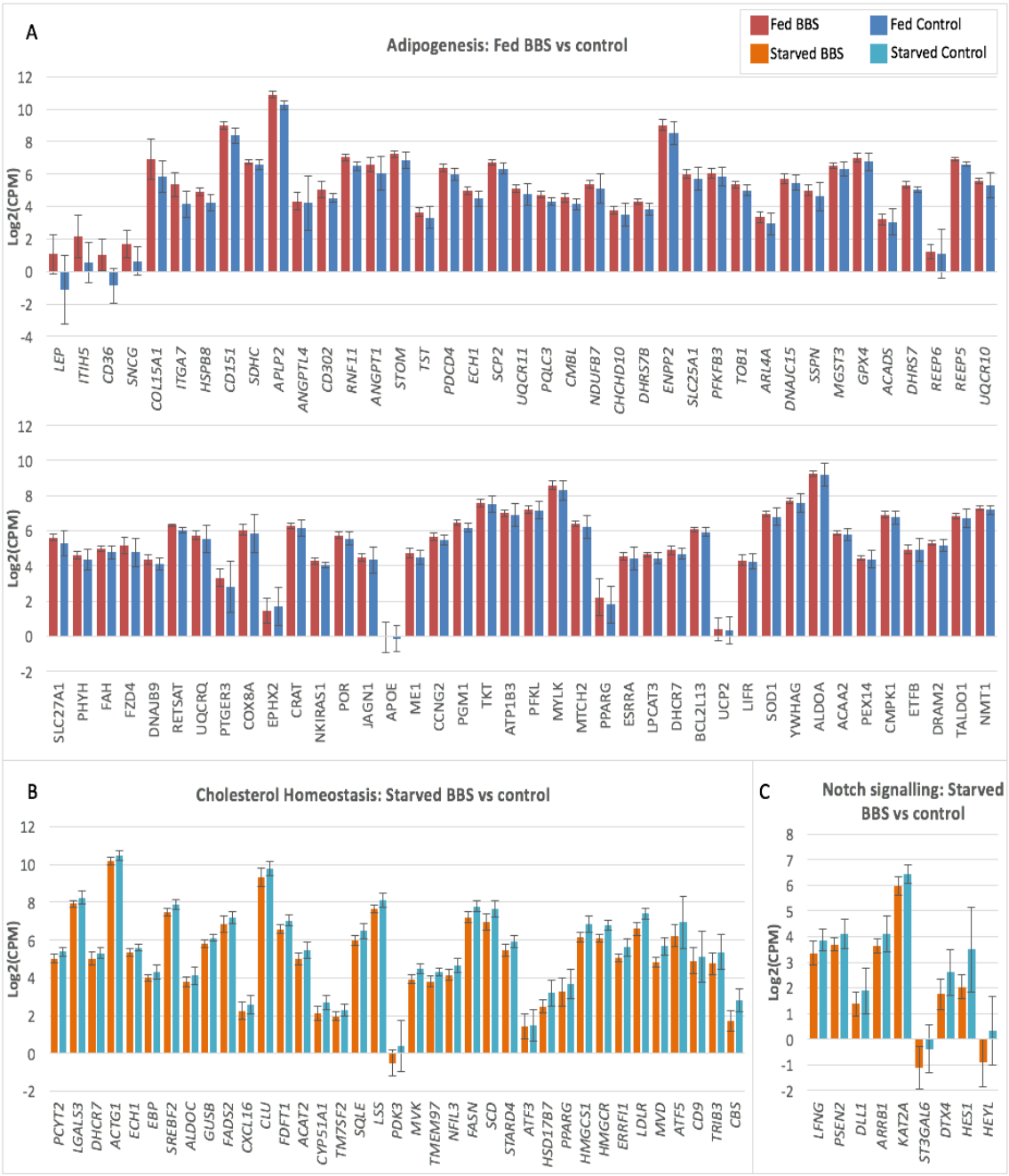

Figure 5.11 – Gene expression for each gene of the leading edge for disease relevant findings from GSEA. Genes are plotted on the x-axis from highest to lowest log2(FC), against log2(CPM). As a log2 function, genes with negative expression levels have CPM less than 1. Error bars represent 95% CI.

Many of the highly significant enriched gene sets correlate between starved vs. fed comparisons, suggesting that these were enriched as a consequence of starvation. For example, the strongest enrichment was a negative correlation with G2M checkpoint genes, with a NES of -3.129 and -3.163 for BBS and control, respectively (FDR<0.0001 for both) (**figure 5.12**). This finding appears consistent with induction of cell cycle arrest. Other gene sets that support this observation include significant negative enrichment of mitotic spindle (**figure 5.12**), hypoxia, and E2F targets.

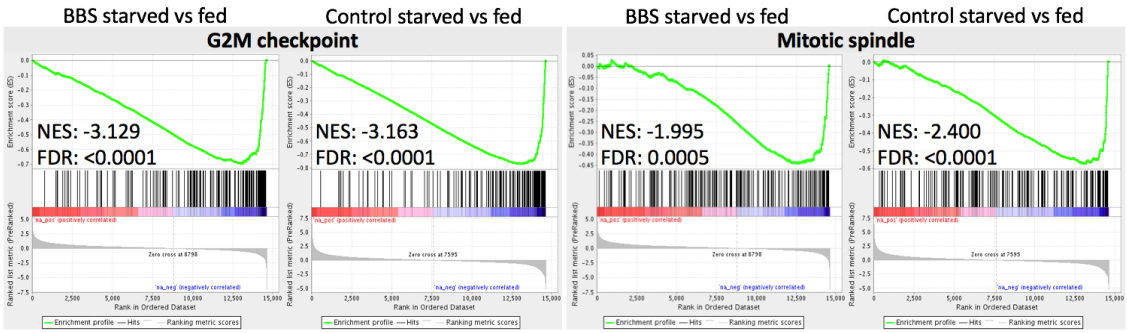

Figure 5.12 – GSEA profiles for treatment comparisons. Significant negatively enriched Hallmark gene sets for BBS starved vs fed and control starved vs fed analyses. Normalised enrichment score (NES) and false discovery rate (FDR) are displayed for each gene set. Enrichment is considered significant when FDR<0.25.

5.2.2.4 Comparison of treatment with respect to disease

As the aim of this study was to capture differences in ciliated cells between disease and control cultures, the most biologically valuable comparison was anticipated to be between serum starved BBS vs controls. However, there were only 29 DE transcripts for this comparison, making it challenging to draw any inferences from such a small statistically significant set of genes. Therefore, a different method for analysis was adopted, where BBS starved vs fed and control starved vs fed experiments were directly compared with one another. For this investigation, the PANTHER statistical enrichment test was implemented with gene sets from the Reactome pathways database. Compared to the Hallmark gene sets, the Reactome database features a vast range of biological processes, which may increase biological understanding to complement GSEA. PANTHER statistical enrichment test was completed with all protein-coding genes, where log2(FC) for each protein was inputted as enrichment criteria. Pathways were considered statistically enriched after Bonferroni correction with adjusted p<0.05 (**figure 5.13**).

The resulting pathways were compared for BBS starved vs fed and control starved vs fed experiments, selecting pathways that were unique or shared for the comparisons. As expected in response to cell cycle arrest by starvation, negative enrichment of cell cycle and mitotic pathways were identified, including “mitotic anaphase”, “mitotic prometaphase”, “cyclin E association events during G1/S transition”, “APC:Cdc20 mediated degradation of cell cycle proteins”, and “cell cycle”. This negative enrichment of cell cycle components was concordant with GSEA. There was also evidence of reduced cytoskeletal remodelling in response to starvation, indicated by negative enrichment of actin regulation, such as “RHO GTPase effectors”, “cooperation of prefoldin and TriC/CCT in actin and tubulin folding”, and “cytoskeletal regulation by Rho GTPase”.

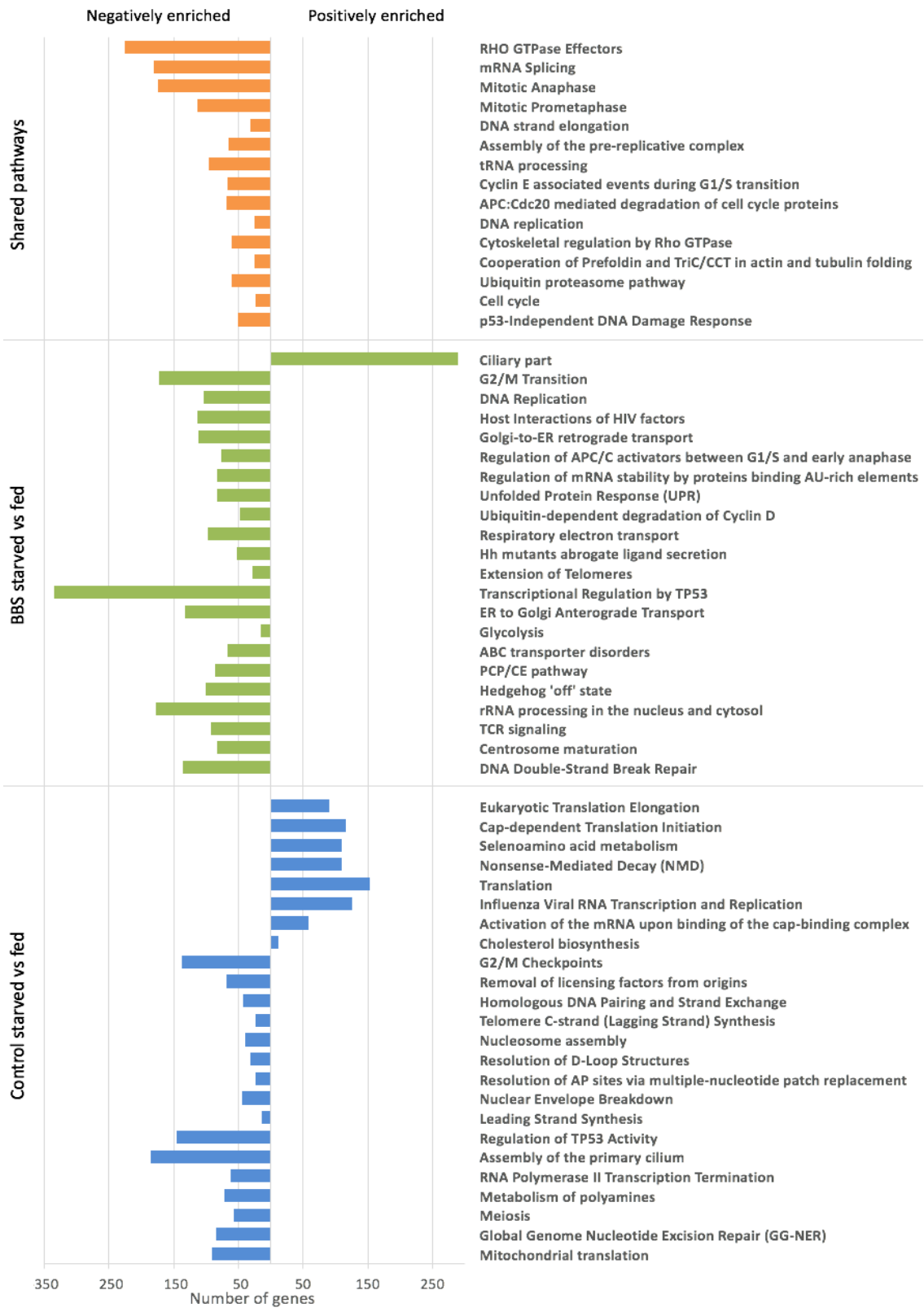

Figure 5.13 – Reactome pathways significantly positively or negatively enriched by PANTHER enrichment test in starved vs fed cultures. Statistically significant pathways were compared to identify those unique to BBS starved vs fed (green), control starved vs fed (blue), or shared (orange). Pathways were statistically significant with adjusted $p<0.05$, ranked in the figure with most significant at the top of each section.

There were some terms that are similar but not identical between BBS starved vs fed and control starved vs fed, highlighting the overlapping nature of the Reactome gene sets. For example, "G2/M transition" in BBS and "G2/M checkpoints" in control, or "transcriptional regulation by TP53" in BBS and "regulation of TP53 activity" in control. Although these respective pathways have different connotations, there were genes shared across the respective processes. Unique pathways in BBS starved vs fed included "PCP/CE pathway", "HH 'off' state", and "centromere maturation", all of which are negatively enriched. The HH and PCP pathways may be of interest, as they are both known to utilise the primary cilia as a signalling hub (Goetz and Anderson 2010). Interestingly, the pathway "assembly of the primary cilium" is negatively enriched in control starved vs fed, yet "ciliary part" is positively enriched in BBS starved vs fed. This discordant finding was investigated further through analysis of DE genes that interact in the pathway "assembly of the primary cilium" in BBS starved vs fed and control starved vs fed.

The assembly of the primary cilium is a well-orchestrated pathway that requires a functional switch from centrosome genes required for mitosis to basal body and IFT associated genes (Izawa et al. 2015). In order to investigate whether differential regulation occurred within an interconnected biological network, protein-protein interactions were mapped within STRING (figure 5.14). In total, there were 186 genes associated with the term "assembly of the primary cilium", of which 80 were significantly differentially expressed in BBS starved vs fed, and 43 differentially expressed in control starved vs fed. 30 DE genes were shared across both BBS and control comparisons.

The protein-protein interaction map revealed at least 4 key interacting networks, defined as IFT machinery, BBS-causing, tubulin, and transition zone/centrosome genes. There was also a group of genes involved in vesicular trafficking, though this did not present itself as a complex network. There were 5 TriC/CCT chaperone genes, *CCT2*, *CCT3, CCT5, CCT8*, and *TCP1*, that connected the BBS-causing genes to the rest of the network through interactions with the tubulin complex. Of note, the transition zone/centrosome genes were separated into 2 subnetworks depending on their direction of expression. For example, genes such as *TCTN1, TMEM216, CEP290, TMEM67, AHI1, B9D1*, and *CC2D2A* were upregulated, whereas *ALMS1, PLK1, PLK4, CDK1, NEK2*, and *CENPJ* were downregulated. These are roughly assembled into upregulated transition zone proteins, and downregulated gene products that localise to the centrosome for mitotic spindle formation, centriole duplication, or kinases responsible for the negative regulation of ciliogenesis.

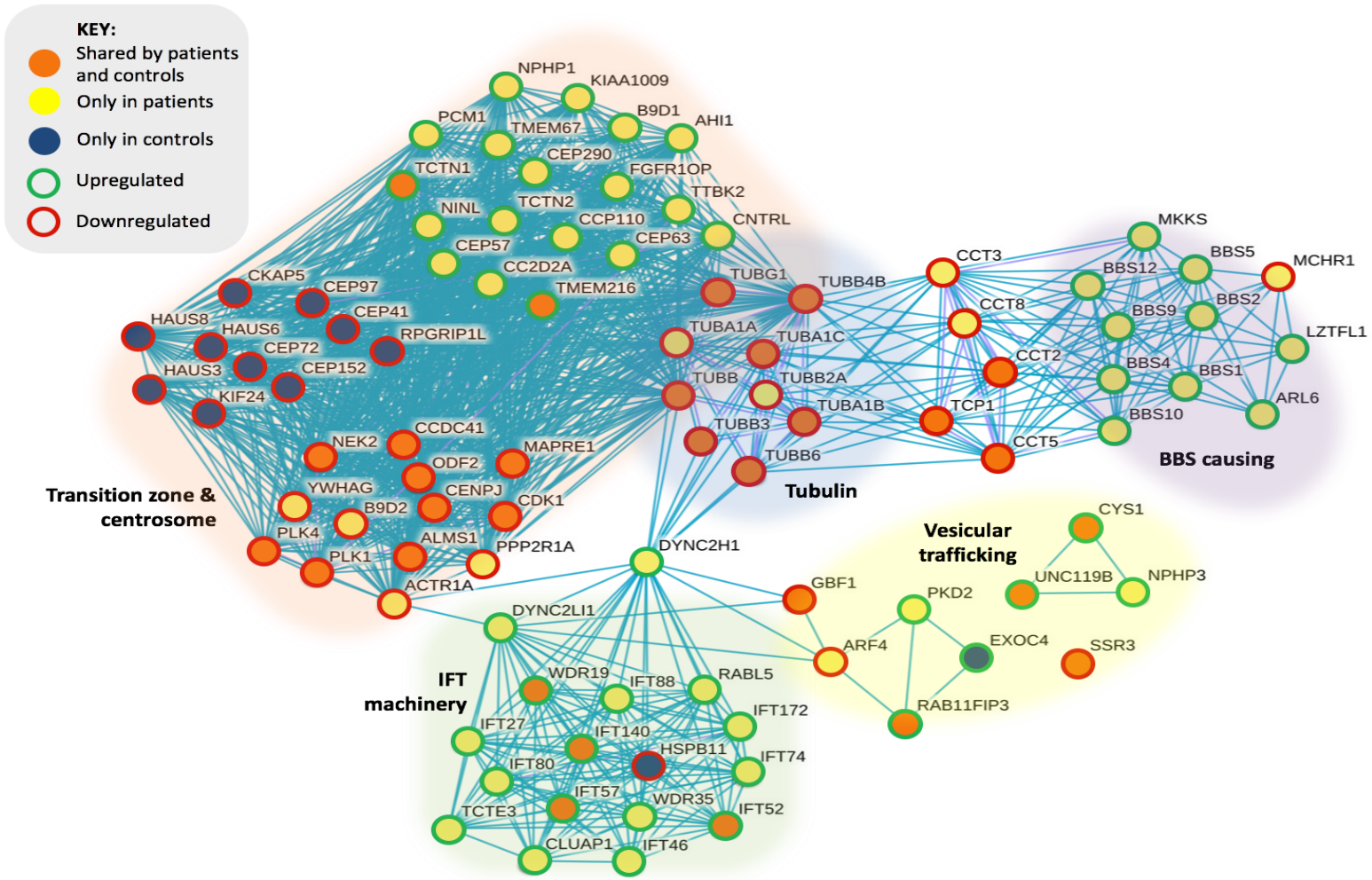

Figure 5.14 – Interactions between DE genes from the Reactome pathway "assembly of the primary cilium", featuring genes shared (orange) or unique to patients (yellow) and controls (blue) from starved vs fed comparisons. Up and downregulated genes are distinguished by green or red bordered nodes, respectively. Gene interactions were plotted using STRING using evidence obtained from curated databases (represented by blue edges).

There were a larger number of DE genes for cilia assembly within the BBS starved vs fed comparison compared to the control group. It is possible that this is a result of having a larger group of BBS patients than controls, creating a statistical artefact. Therefore, making inferences from the comparison of BBS compared to control may be biased. However, if there were differences observed in controls that were not in BBS, this may be a real finding. For example, there were 12 DE genes in control starved vs fed that were not found in BBS. Of these, there was a cluster of 10 genes within the transition zone/centrosome complex that were all significantly downregulated in control starved vs fed, but not BBS (figure 5.15).

Research of this group of genes found that they are involved in centrosome duplication (*CEP72* and *CEP152*), mitotic spindle formation (*CKAP5, HAUS3, HAUS6*, and *HAUS8*), or are negative regulators of ciliogenesis (*CEP97* and *KIF24*) (Spektor et al. 2007; Kobayashi et al. 2011). Moreover, *HAUS3, HAUS6,* and *HAUS8* form part of a hetero-octamer complex named Homologous to Augmin subunits (HAUS), and are required for mitotic spindle assembly and maintenance of centrosome integrity (Lawo et al. 2009). Therefore, it would be expected that expression of these genes would downregulate in response to cell cycle arrest. Further investigation of expression data for these genes demonstrated that read counts were not significantly different between starved BBS and starved control fibroblasts, but 4 out of the 10 genes were significantly different between fed BBS and fed control (figure 5.15; table 5.4). Furthermore, expression was significantly higher in fed control compared to starved control for all genes, whereas this was not the case for BBS comparisons. Expression of *HAUS3* (BBS fed mean= 3.87, 95% CI= 3.65, 4.26; control fed mean= 4.67, 95% CI= 4.29, 5.06; FDR= 0.0157)*, CEP41* (BBS fed mean= 3.98, 95% CI= 3.87, 4.10; control fed mean= 4.54, 95% CI= 4.28, 4.80; FDR= 0.0251)*, CEP152* (BBS fed mean= 3.54, 95% CI= 3.18, 3.89; control fed mean= 4.70, 95% CI= 4.04, 5.36; FDR= 0.0478), and *CEP97* (BBS fed mean= 4.01, 95% CI= 3.79, 4.23; control fed mean= 4.69, 95% CI= 4.24, 5.14; FDR= 0.0478) was significantly lower in BBS fed compared to control fed.

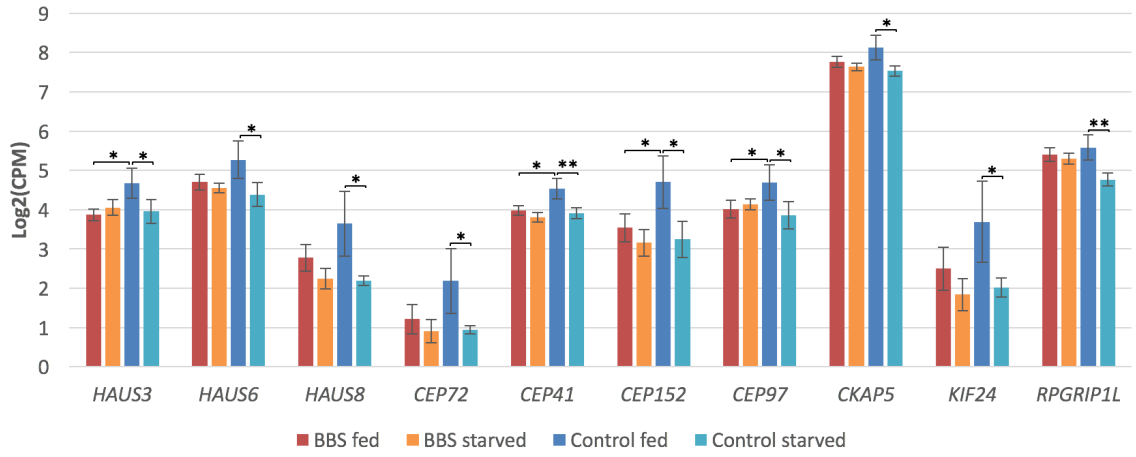

Figure 5.15 – Gene expression for a cluster of 10 genes from the Reactome pathway "assembly of the primary cilium" that were significantly downregulated in control starved vs fed, but not BBS starved vs fed. Error bars represent 95% CI, where * FDR<0.05 and ** FDR<0.005 for control starved vs fed or fed BBS vs control.

Gene	BBS starved	BBS fed	Control starved	Control fed	BBS starved vs fed FDR	Control starved vs fed FDR	Fed BBS vs control FDR	Starved BBS vs control FDR
HAUS3	4.06 (3.85, 4.26)	3.87 (3.71, 4.02)	3.96 (3.65, 4.26)	4.67 (4.29, 5.06)	0.3613	0.0380	0.0157	0.8610
HAUS6	4.55 (4.44, 4.67)	4.70 (4.50, 4.90)	4.39 (4.09, 4.68)	5.27 (4.80, 5.74)	0.4931	0.0108	0.1418	0.7381
HAUS8	2.24 (1.98, 2.50)	2.78 (2.44, 3.11)	2.19 (2.07, 2.31)	3.65 (2.82, 4.47)	0.0915	0.0092	0.1338	0.9540
CEP72	0.90 (0.61, 1.19)	1.22 (0.84, 1.59)	0.94 (0.84, 1.05)	2.19 (1.36, 3.01)	0.4233	0.0453	0.0978	0.9870
CEP41	3.81 (3.68, 3.94)	3.98 (3.87, 4.10)	3.91 (3.77, 4.04)	4.54 (4.28, 4.80)	0.1974	0.0048	0.0251	0.9404
CEP152	3.15 (2.81, 3.50)	3.54 (3.18, 3.89)	3.25 (2.79, 3.70)	4.70 (4.04, 5.36)	0.2712	0.0157	0.0478	0.9892
CEP97	4.13 (3.99, 4.28)	4.01 (3.79, 4.23)	3.86 (3.51, 4.21)	4.69 (4.24, 5.14)	0.6033	0.0260	0.0478	0.6878
CKAP5	7.63 (7.54, 7.73)	7.77 (7.62, 7.91)	7.54 (7.40, 7.67)	8.13 (7.81, 8.45)	0.3760	0.0154	0.1758	0.8036
KIF24	1.84 (1.43, 2.24)	2.50 (1.95, 3.04)	2.02 (1.78, 2.27)	3.69 (2.66, 4.72)	0.1269	0.0329	0.6384	0.9746
RPGRIP1L	5.30 (5.17, 5.44)	5.41 (5.23, 5.58)	4.77 (4.60, 4.93)	5.58 (5.27, 5.90)	0.5545	0.0020	0.1485	0.1869

Table 5.4 – Data for each biological group for the 10 genes from the Reactome pathway "assembly of the primary cilium" that were significantly downregulated in control starved vs fed, but not BBS starved vs fed. Figures are expressed as sample mean (log2(CPM)) and 95% CI (upper, lower limit), as well as the FDR for each comparison.

5.2.3 Identifications of proteins from iTRAQ-MS/MS

The spectra obtained from iTRAQ mass spectrometry runs for serum fed and serum starved samples were analysed for identifications separately, and implemented by the BGI. Peptide identifications were made using MascotPercolator software version 2.3.02 (Käll et al. 2007), against a human proteome database of 134,169 sequences. Matching peptides required an FDR <0.01 to be considered a true positive for protein identification. 30,590 peptides from 360,137 spectra were identified from serum starved fibroblasts, which corresponded to 5439 proteins. Of these peptides, 29,554 were unique. The experiment from serum starved fibroblasts identified 30,587 peptides from 357,751 spectra. 5633 proteins were identified, where 29,517 peptides were unique. Only proteins with at least 1 unique peptide were kept for quantitation. MS/MS spectra were used for protein quantitation by Iquant (Wen et al. 2014), which normalised and quantified findings. Data was provided as a protein ratio between each isobaric tag, where only proteins with FDR<0.01 were included for further statistical analysis (supplemental data, **tables S.34-S.39**).

The runs were compared to determine how many quantifiable proteins were overlapping. Between the 2 runs, 65.7% of proteins were shared, whereas 1050 (15.7%) were only detected in serum fed fibroblasts, and 1244 (18.6%) were only identified by serum starved cells. Therefore, for statistical analysis between the 2 experiments, only 4389 proteins could be compared.

5.2.3.1 Differentially expressed proteins from iTRAQ experiments

Statistical analysis with T-test and multiple correction for FDR was applied to comparisons of protein ratios between biological groups (supplemental data, **tables S.30-S.33**). The same analytical design that compared treatment and disease data in RNA-seq experiments were also utilised here. Proteins were considered differentially expressed with an FDR less than 0.05 (**figure 5.16**). The numbers of differentially up or downregulated proteins for each comparison are presented in **table 5.5**. As many proteins were statistically significant, proteins with a greater fold change difference (-0.5< log2(FC) >0.05) were also counted. The greatest number of DE proteins were identified between starved BBS vs control, where 1194 proteins were downregulated and 1279 proteins were upregulated. Fed BBS vs control also exhibited a significant number of DE proteins, where 1267 and 961 were down and upregulated, respectively. Comparison of BBS starved vs fed identified 773 downregulated and 1313 upregulated proteins, whereas control starved vs fed resulted in fewer DE proteins, with 211 downregulated and 373 upregulated proteins. The comparison with the greatest DE proteins with log2(FC) less than -0.5 or greater than 0.5 was BBS starved vs fed, with 118 downregulated and 349 upregulated genes.

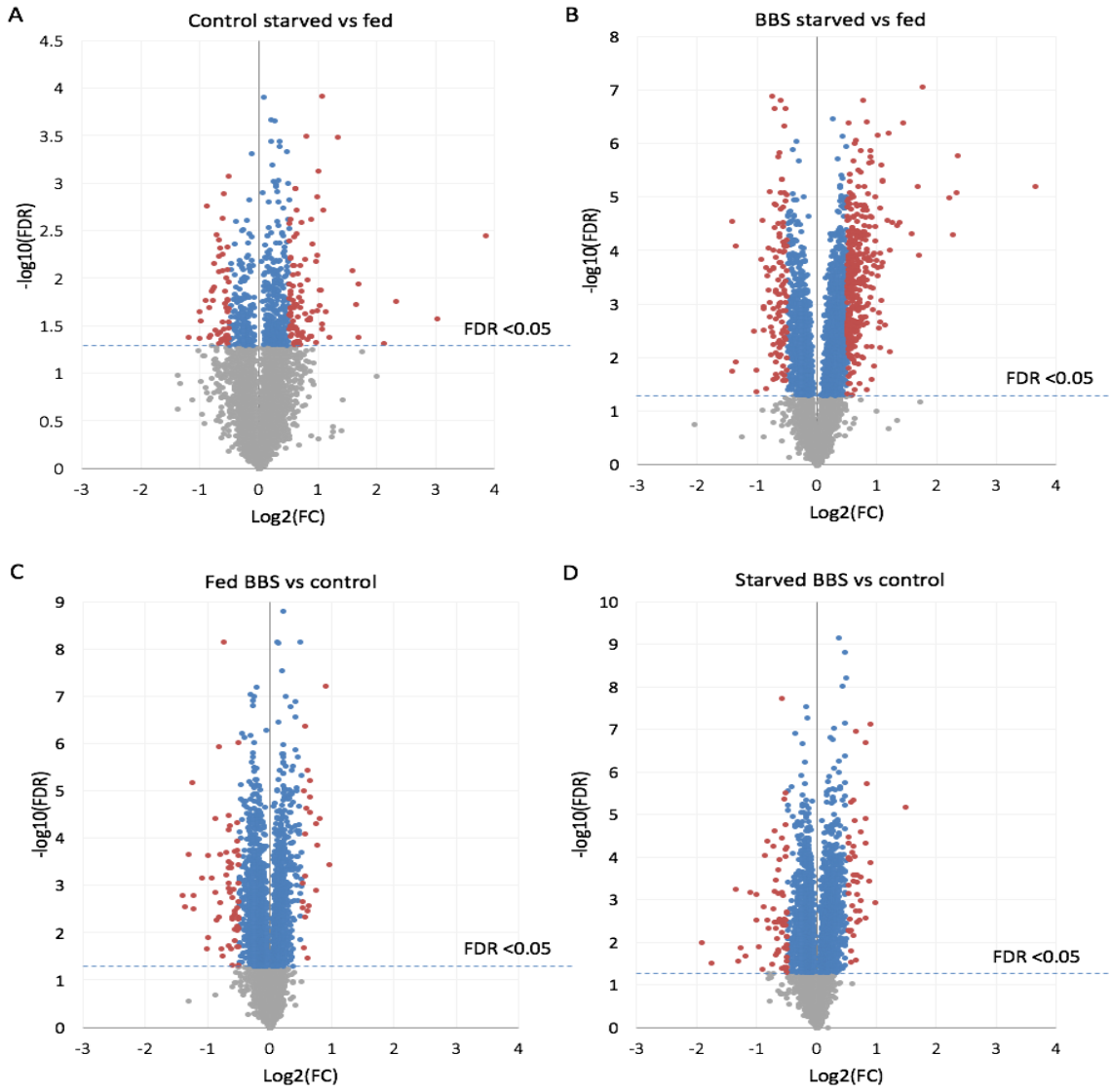

Figure 5.16 – Volcano plots displaying differentially expressed proteins (blue) for each comparison. Proteins are plotted by log2 fold change against –log10(FDR). Proteins were defined as differentially expressed with FDR <0.05. DE proteins with log2(FC) >0.5 or <-0.5 are highlighted in red. A-B) Distribution of DE proteins within treatment comparisons. C-D) Spread of DE proteins within disease comparisons.

Table 5.5	Down regulated	Not differentially expressed	Up regulated	Log2FC <-0.5	Log2FC >0.5	Total	Total DE Proteins
Control starved vs fed	211	3805	373	52	96	4389	584
BBS starved vs fed	773	2303	1313	118	349	4389	2086
Fed BBS vs control	1267	3211	961	62	23	5439	2228
Starved BBS vs control	1194	3160	1279	66	47	5633	2473

Table 5.5 – The number of proteins detected as differentially expressed for each experimental comparison. DE proteins were defined as FDR<0.05. The number of genes with -0.5 < log2(FC) > 0.5 are also recorded in the table.

5.2.3.2 Gene set enrichment analysis with iTRAQ proteomics data

The GSEA methodology was applied to the iTRAQ proteomics data to simplify the data and obtain gene sets that were enriched at the protein level. 26 Hallmark gene sets were returned as significantly enriched in at least one comparison with FDR<0.25 (**figure 5.17**, supplemental data, **tables S.9-S.16**). The significantly enriched gene sets were explored to find disease-relevant findings for starved BBS vs control, fed BBS vs control, and between BBS and control for starved vs fed comparisons. Interestingly, adipogenesis was found to be positively enriched in fed BBS vs control cultures, in accordance with the RNA-seq GSEA. However, no gene sets were identified as significantly enriched in opposing directions for BBS starved vs fed and control starved vs fed.

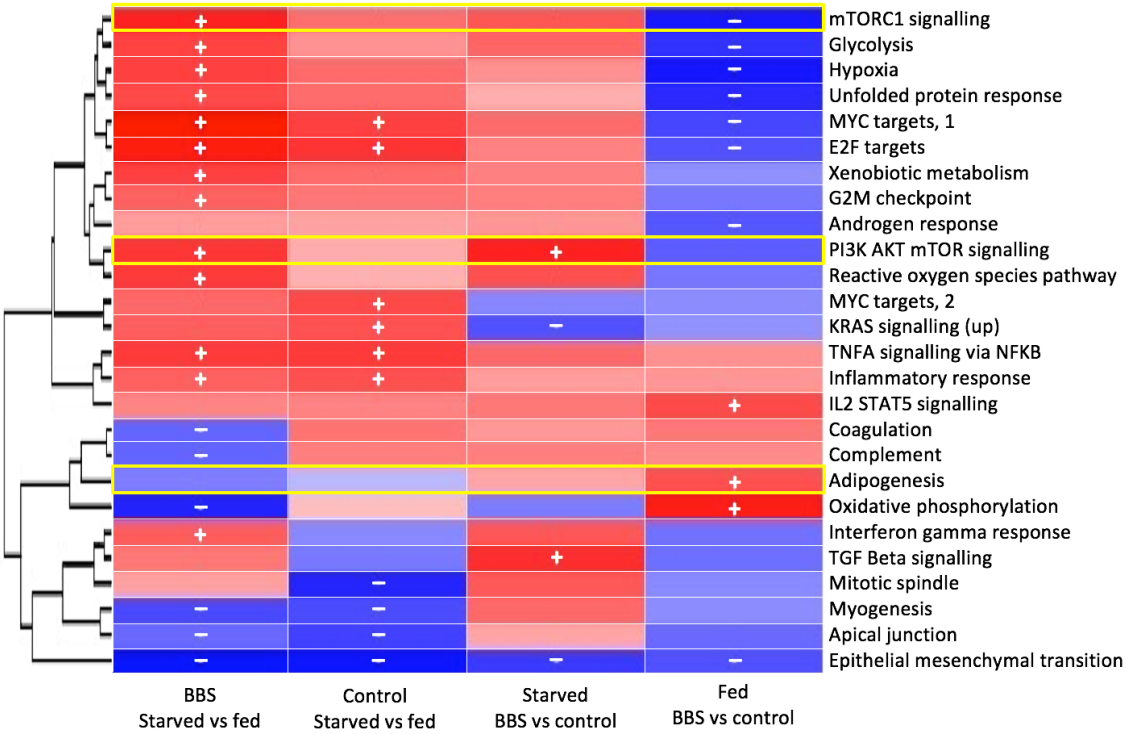

Figure 5.17 – Heatmap featuring enriched Hallmark gene sets generated by GSEA with iTRAQ proteomics data. Significantly negatively or positively enriched pathways where FDR<0.25 are marked with – or +, respectively. As the FDR value decreases (i.e. becomes more significant) by each factor of 10, the colour intensity increases. Gene sets that were identified as biologically relevant are highlighted with yellow boxes.

Of interest to cilia function, the mTOR signalling pathway was established as significantly enriched. This pathway functions to regulate cellular growth in response to cellular energy levels. Dysregulation of the mTOR pathway has been shown to reduce ciliary length, and is implicated in polycystic kidney disease (Basten and Giles 2013; Ibraghimov-Beskrovnaya and Natoli 2011). mTOR signalling pathway is covered by two different Hallmark gene sets, namely PI3K AKT mTOR and mTORC1. Both of these sets were found in be significantly enriched in BBS comparisons, but not in control starved vs fed (**figure 5.18**). mTORC1 signalling was significantly negatively enriched

in fed BBS vs control comparison (NES=-1.911; FDR=-0.0045), whereas starved BBS vs control was positively enriched for PI3K AKT mTOR signalling (NES=1.843; FDR=0.0142). Both pathways were significantly positively enriched in BBS starved vs fed comparisons. This suggests that the mTOR pathway is induced in response to starvation in BBS fibroblasts, but not in controls.

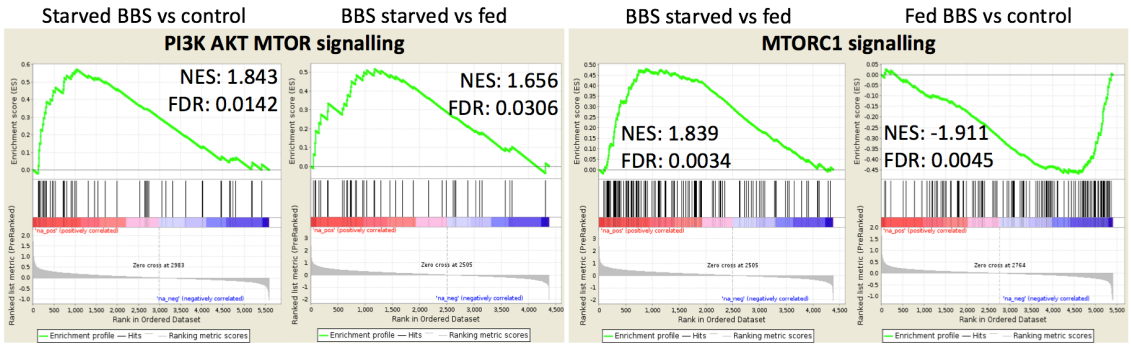

Figure 5.18 – GSEA profiles for enriched Hallmark gene sets related to mTOR signalling from proteomics data. Enrichment is considered significant when FDR<0.25.

5.2.3.3 Pathway analysis for iTRAQ BBS vs control experiments

In contrast to the RNA-seq experiment, there was a large set of DE proteins for starved BBS vs control. As it was not possible to investigate this comparison in detail through analysis of transcripts, the opportunity was taken for the iTRAQ dataset. To investigate Reactome pathways that were significantly enriched, all proteins identified from starved BBS vs control were entered into PANTHER statistical enrichment test, where log2(FC) for each protein was used as enrichment criteria. Pathways were considered statistically enriched after Bonferroni correction with adjusted p<0.05 (**figure 5.19**).

17 pathways were found to be significantly positively enriched and a single pathway was significantly negatively enriched. From this analysis, there were 5 pathways that were recognised as biologically interesting due to association with ciliary maintenance, signalling, or function. These were “assembly of the primary cilium”, “hedgehog ‘off’ state”, “RHO GTPase effectors”, “centrosome maturation”, and “cytoskeletal regulation by RHO GTPase”. As these have some overlapping functions, it was hypothesised that proteins within these pathways would form a rich highly connected network. This was tested by creating a STRING protein-protein interaction map, based on knowledge from curated databases, and featuring the DE proteins from the 5 relevant pathways. It was also investigated whether these pathways may be implicated with BBS1 function, thereby forming a link between protein dysfunction caused by the *BBS1* p.M390R variant and the upregulation of pathways found in the starved BBS vs control experiment. To

model these interactions, proteins that were known to interact with BBS1 were also added to the network, which featured BBS1, BBS2, ARL6, BBS4, BBS5, BBS5, BBS7, TTC8, BBS12, BBIP1, LZTFL1, and RAB3IP (note that with the exception of LZTFL1, these proteins were not detected by the iTRAQ experiment). The resulting network is presented in **figure 5.20**.

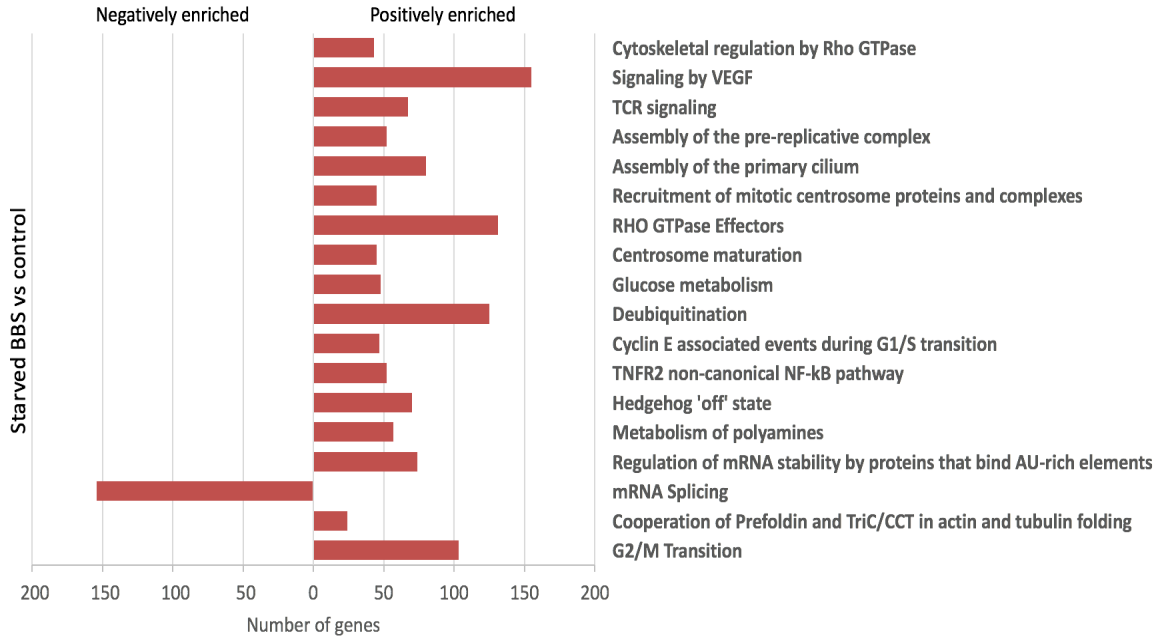

Figure 5.19 – Reactome pathways found to be significantly positive or negatively enriched by PANTHER statistical enrichment test for starved BBS vs control comparison from iTRAQ proteomics data. Pathways were considered statistically enriched with adjusted $p<0.05$, ranked in the figure with most significant at the top.

Upon closer review of DE proteins within each pathway, there were large overlaps. Of the 27 DE proteins found in "centrosome maturation", many overlapped with "assembly of the primary cilium", which consisted of 45 DE proteins. These included CEP78, CEP290, FGRF1OP, and ACTR1A, and emphasised the dual function of many centrosome proteins. These 2 pathways also shared some proteins with "RHO GTPase effectors", including CLASP1, MAPRE1, YWHAE, YWHAG, and tubulins. Unsurprisingly, there was also a large overlap in DE proteins between "RHO GTPase effectors" and "cytoskeletal regulation by RHO GTPases", which had 73 and 28 DE proteins, respectively. Finally, "hedgehog 'off' state" was associated with 40 DE genes. 27 of these were subunits of the proteasome, including alpha subunits (PSMA1-7), beta subunits (PSMB1-8), proteasome ATPases (PSMC1-4,6), non-ATPases (PSMD1-14), and activator subunits (PSME1,3). Components of the proteasome also interacted with proteins of other pathways, including many from "RHO GTPase effectors" through PSMA7 subunit. The presence of the proteasome in the HH pathway is suggestive of GLI protein processing into the repressor state (Gerhardt et al. 2016).

Overall, assembly of these DE proteins into a protein-protein interaction map uncovered a vast interconnected network.

Next, the protein network map was used to model the putative involvement of BBS1, where mutation of p.M390R may have led to the upregulation of these 5 pathways in starved cultures. BBS1 is incorporated into the network through interactions with RAB3IP, also known as RABIN8, SMO, and TriC/CCT chaperone proteins that partake in the assembly of the BBSome. The interaction with these intermediates is shown to connect BBS1 with proteins from all 5 pathways. For example, BBS1 incorporates into "Hedgehog 'off' state" through direct interaction with SMO, a HH effector that controls HH signalling depending on its localisation with respect to the primary cilium. The interaction of the BBSome with SMO is thought to allow trafficking of SMO into the cilium for HH signalling, an interaction also involving LZTFL1 (Seo et al. 2011), which was upregulated in starved BBS fibroblasts compared to starved controls. SMO is also predicted to interact with the cAMP-dependent protein kinase, PRKACB, which in turn networks with members of the "RHO GTPase effectors" and "cytoskeletal regulation by RHO GTPase" pathways. Furthermore, BBS1 is linked to components of "assembly of primary cilium" and "centrosome maturation" through RAB3IP. Although RAB3IP was not identified by the iTRAQ proteomics experiment, it can provide an insight into how the presence of the *BBS1* p.M390R variant impacts these pathways. The interaction with BBS1 and RAB3IP is required for transport of cargo to the ciliary membrane (Nachury et al. 2007). Unfortunately, further investigation into expression levels of these proteins could not be completed due to pooling of samples; a limitation that was understood upon conception of the experimental design.

5.2.4 Integration of transcriptomic and proteomic datasets

Due to the volume of information and low availability of sophisticated analytical techniques, data observed from a single omics technology may contain a high number of false positive and false negative findings (Ge et al. 2003). Although statistical analysis attempts to reduce these, there is still a 5% chance of a false finding with an FDR threshold of 0.05. To increase the confidence and reliability of a true biological finding, multiple independent omic techniques can be integrated. However, due to sample heterogeneity and often low correlation between different types of technology, this is not a trivial task (Gomez-Cabrero et al. 2014). This challenge is exacerbated when attempting to integrate data at the individual gene/protein level, as there are far too many differentially expressed findings to consider. Therefore, within this study, datasets have been integrated at the pathway level.

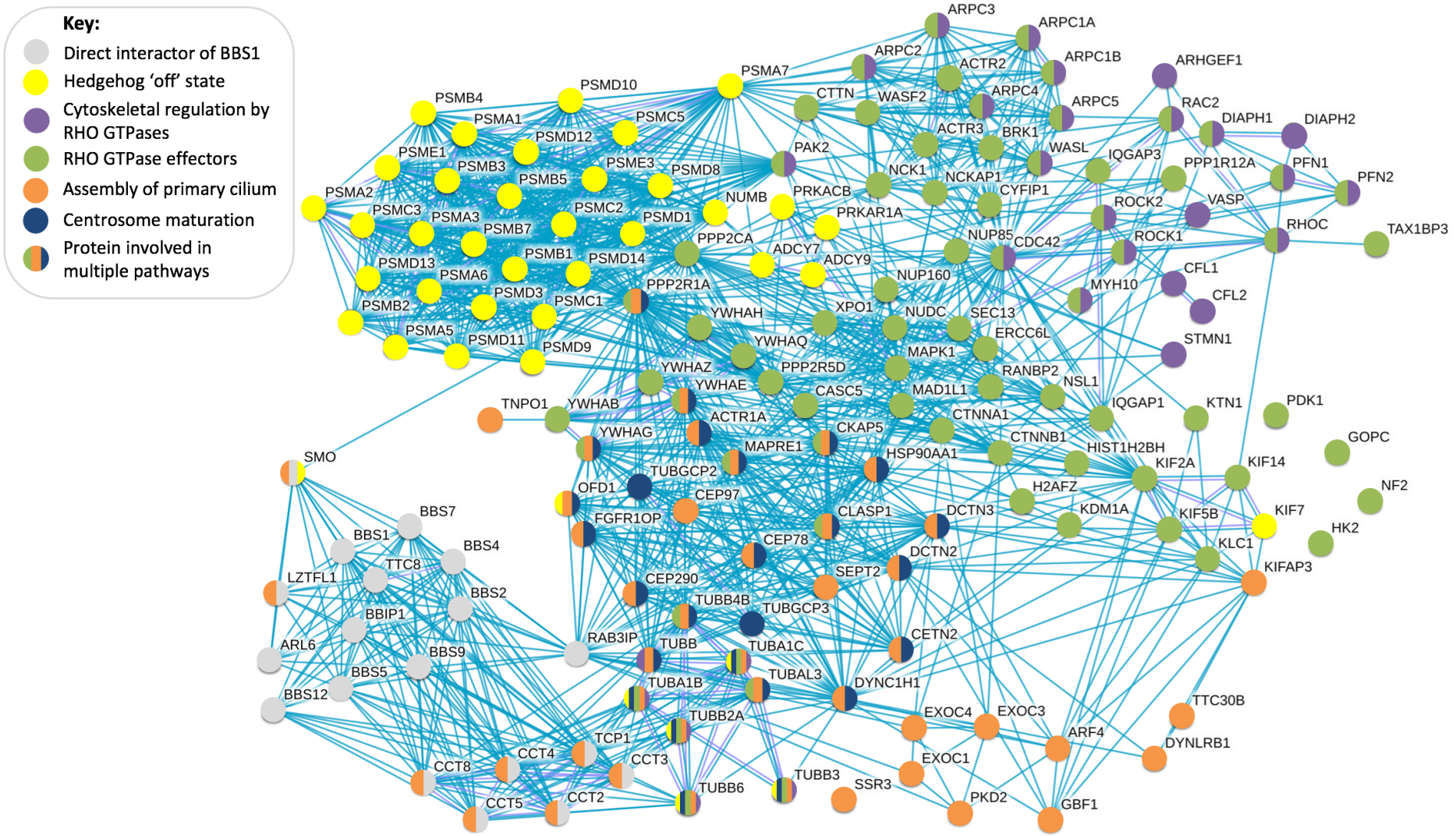

Figure 5.20 – Protein-protein interaction network of upregulated pathways from starved BBS vs control comparison (figure legend on next page)

Figure 5.20 (on previous page) – Protein-protein interaction network featuring DE proteins from 5 pathways significantly positively enriched by PANTHER statistical enrichment test. DE proteins from the 5 different pathways are coloured differently, where proteins may be coloured multiple colours if involved in more than one pathway. Grey nodes were added manually to model the interconnectivity of the pathways with BBS1. These were not identified in the iTRAQ experiment, unless labelled with additional colours. Protein interactions were plotted using STRING using evidence obtained from curated databases (represented by blue edges).

The analysis completed with RNA-seq and iTRAQ proteomics data by GSEA and PANTHER statistical enrichment testing were assessed to determine whether there were overlapping findings. In total, 17 pathways were identified as significantly enriched in both RNA and protein datasets (**table 5.6**). Starved BBS vs control found 2 pathways from GSEA, fed BBS vs control had 7 shared pathways from GSEA, and there were 8 pathways found from PANTHER enrichment test that were present in BBS starved vs fed but not control starved vs fed. These pathways were highlighted for dysregulation, where it was not important whether pathways were consistently positively or negatively enriched, or contained the exact same genes, across different omic technologies. Interestingly, epithelial mesenchymal transition was identified for both starved BBS vs control and fed BBS vs control, and were both positively enriched in transcriptome and negatively enriched at the protein level. This suggests that it may be a dysfunctional pathway in BBS cells, which, as a novel finding, may be tissue specific. Out of the GSEA pathways, for fed BBS vs control, there were many overlapping pathways which were correlated in the same direction. This may illustrate that starvation increases cellular heterogeneity, making it more challenging to make statistically significant findings.

A finding that was well supported from the different omics datasets, and is biologically relevant to BBS, is adipogenesis. This gene set was identified as positively enriched in both transcriptome and proteomic studies in fed BBS vs control experiments. The genes from these pathways were investigated to determine whether the leading edge genes/proteins were interrelated. A STRING network was compiled of all leading edge genes, including 79 from RNA-seq analysis and 40 from iTRAQ proteomic analysis, 21 of which were shared. 65 of these were predicted to interact with information from databases, experimental data, or from text mining. The resulting network is presented in **figure 5.21**.

Experiment	Pathway/ gene set description	Enrichment transcripts	Enrichment protein	Method
Starved BBS vs control	KRAS signalling (up)	Negative	Negative	GSEA
	Epithelial mesenchymal transition	Positive	Negative	GSEA
Fed BBS vs control	MYC targets, 1	Negative	Negative	GSEA
	E2F targets	Negative	Negative	GSEA
	IL2 STAT5 signalling	Positive	Positive	GSEA
	Adipogenesis	Positive	Positive	GSEA
	Oxidative phosphorylation	Positive	Positive	GSEA
	Hypoxia	Positive	Negative	GSEA
	Epithelial mesenchymal transition	Positive	Negative	GSEA
BBS starved vs fed that were not in control starved vs fed	G2/M Transition	Negative	Positive	PANTHER
	Hedgehog 'off' state	Negative	Positive	PANTHER
	HH mutants abrogate ligand secretion	Negative	Positive	PANTHER
	Regulation of APC/C activators between G1/S and early anaphase	Negative	Positive	PANTHER
	Regulation of mRNA stability by proteins binding AU-rich elements	Negative	Positive	PANTHER
	Respiratory electron transport	Negative	Positive	PANTHER
	rRNA processing in the nucleus and cytosol	Negative	Positive	PANTHER
	Ubiquitin-dependent degradation of Cyclin D	Negative	Positive	PANTHER

Table 5.6 – Significantly enriched pathways or gene sets that were identified in both analysis of RNA-seq and iTRAQ, either by GSEA or PANTHER statistical enrichment test using Reactome pathways database.

The protein-protein interaction map output from STRING was highly complex, with few well-structured complexes or networks. Therefore, interacting genes from the respective transcriptome and proteome datasets were highlighted and reanalysed to find more manageable clusters of protein-protein interactions. This grouped the nodes into 5 complexes depending on their molecular function. The first identified network within the transcriptome dataset was adipocyte differentiation, which featured genes that are regulators of adipogenesis, such as *PPARG, CD36, LEP*, and *SLC27A1*. Another complex from the RNA-seq analysis was glucose metabolism, featuring components of the pentose phosphate pathway, *TALDO1* and *TKT*, and glycolysis, such as *PFKL, PFKFB3*, and *PGM1*. Within the proteomic set there were rich interactions involving mitochondrial components, including those involved in the respiratory chain (also known as oxidative phosphorylation), UQCR10, UQCRQ, NDUF subunits, ATP1B3, COX6A1, and ATP5O. Nodes were also found for both transcriptomic and proteomic datasets to be involved in fatty acid metabolism, including the genes *SLC25A1, EPHX2, DHCR7, ACAA2*, and *ACADS*, and proteins HADH, ACOX1 and ACO2. Although these were represented as two distinct networks, they were interconnected by shared nodes, PEX14 and ECH1. This demonstrated that individual networks could be reintegrated via the shared nodes back into the larger, complex network.

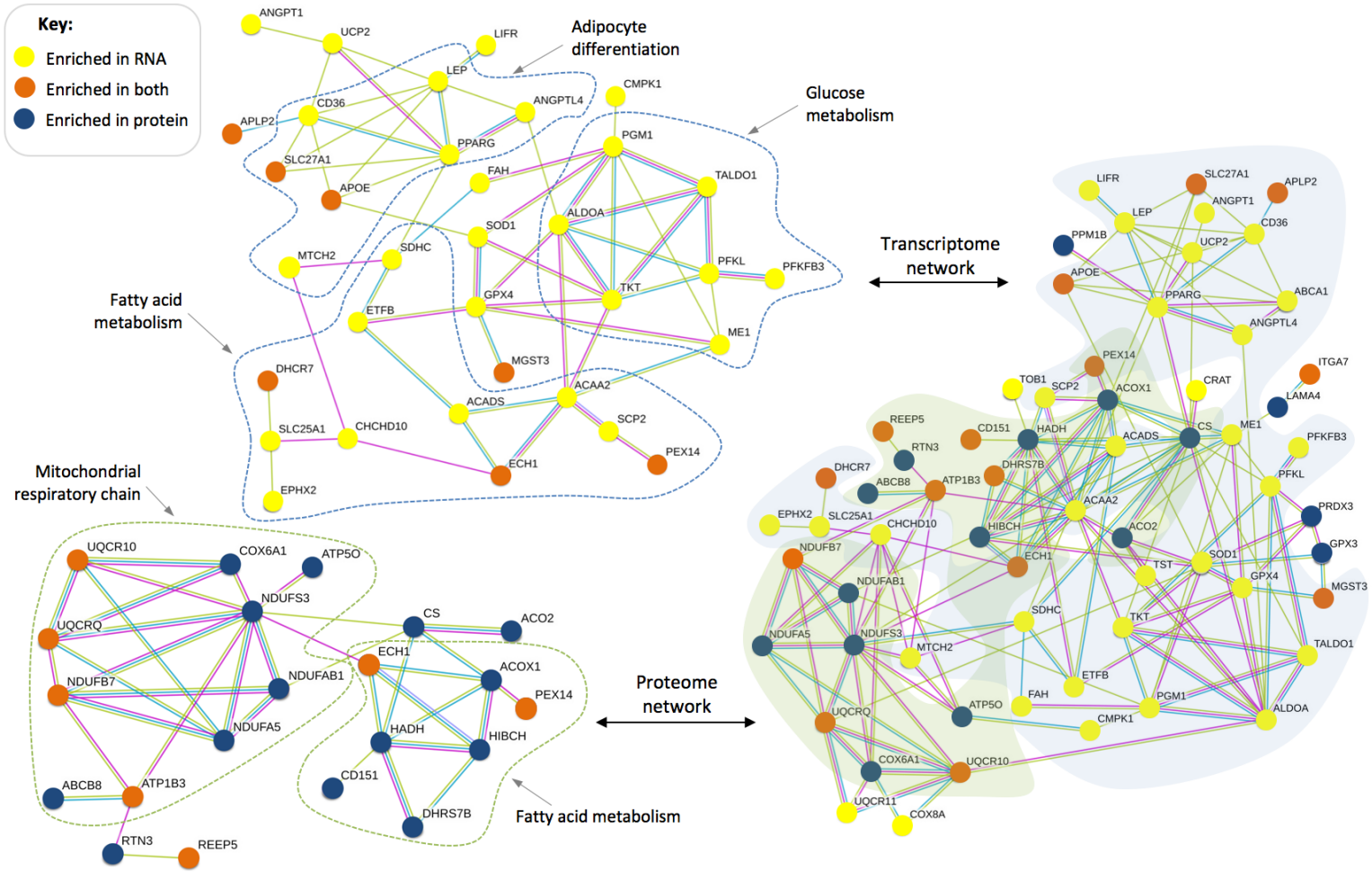

Figure 5.21 – Protein-protein interaction map of the Hallmark adipogenesis gene set identified as statistically positively enriched in fed BBS vs control by GSEA of 2 omic technologies (right). Nodes were coloured by genes/proteins in the GSEA leading edge from RNA-seq (yellow) or proteomic (blue) analyses, where orange nodes are found in both. The integrated map was separated to identify functional complexes found for genes (blue overlay, top left) or proteins (green overlay, bottom left), which are denoted by the broken lines. Protein interactions were plotted using STRING, using information from curated databases (blue edges), experimental data (pink edges), or text mining (green edges).

5.3 Discussion

Previous studies investigating the cellular effects of BBS-causing variants have mainly focussed on gene expression and protein localisation of a handful of markers related to cilia function. Therefore, many of the intricate pathways that are required for ciliary maintenance have been overlooked. To our knowledge, this was the first study to utilise BBS patient derived cells to obtain a global cellular profile, which allowed identification of dysregulated pathways beyond the cilium.

This investigation aimed to identify dysregulation of cellular pathways in BBS fibroblasts compared to controls to elucidate the pathobiology of *BBS1* p.M390R. BBS and control fibroblasts were probed at both mRNA transcriptomic and proteomic levels, unveiling over 15,000 differentially expressed transcripts and proteins combined, and over 100 statistically enriched gene sets and pathways. Furthermore, integration of omic technologies at the pathway level refined and enhanced reliability of findings, identifying 17 pathways across the different experimental comparisons that were found in both transcript and protein datasets. The experimental design presented a novel approach for analysis of cellular components that were differentially regulated in conditions to promote cell cycle progression or to promote ciliogenesis. These were then compared between disease and control groups. The consequence was identification of processes associated with obesity, upregulation of signalling pathways, and downregulation of genes involved in centrosome integrity. Although there were additional significant findings, these were identified as most biologically relevant to BBS pathobiology and prioritised for further investigation. They will be the focus of this discussion.

5.3.1 Enrichment of pathways associated with adipogenesis in serum fed BBS cultures

GSEA presented a knowledge driven and methodical approach to probing the datasets against 50 concise, yet relatively general, Hallmark gene sets. The analysis identified the gene set adipogenesis, as statistically positively enriched for the fed BBS vs control comparison. This was a highly interesting and relevant finding, as obesity is a cardinal feature of BBS. Increased confidence of this result was achieved through positive enrichment of both transcriptomic and proteomic datasets (RNA-seq NES= 1.682, FDR= 0.0149; protein NES= 1.447, FDR= 0.1780), making it a convincing candidate as a dysregulated pathway in proliferating BBS fibroblasts.

Adipogenesis describes the differentiation of a mesenchymal stem cell (MSC) through to a mature adipocyte (Ali et al. 2013). The role of the adipocyte is to maintain lipid homeostasis and control the energy balance through the storage of triglycerides. In a period of nutritional deficit,

triglycerides are metabolised by lipases into fatty acids, which can be converted into glucose by the liver via the TCA cycle. The primary regulator of adipogenesis is the transcription factor PPARγ (encoded by *PPARG*), which is essential for adipogenic induction and activation of C/EBP proteins, required for terminal adipocyte differentiation (Rosen et al. 2002). PPARγ expression is switched on early during adipocyte differentiation and is highly expressed in adipose tissue (Tontonoz et al. 1994). Interestingly, impromptu differentiation of murine fibroblasts to adipocytes has been demonstrated by overexpression of both PPARγ and C/EBPα (Tontonoz et al. 1994; Freytag et al. 1994). Signalling pathways Wnt and HH have also been shown to orchestrate adipogenesis and are potent inhibitory factors (Kopinke et al. 2017; Marion et al. 2009). For example, disruption of Wnt signalling was sufficient to induce spontaneous differentiation of myoblasts to adipocytes (Ross et al. 2000).

Genes that induce adipocyte differentiation were upregulated within the GSEA leading edge of the transcriptomic data for fed BBS vs control experiments, including *PPARG, CD36, SLC27A1*, and *LEP* (Tontonoz et al. 1994; Qiao et al. 2008; Challa et al. 2015). To determine how these key adipogenic genes interacted with the rest of the enriched adipogenesis gene set, gene names from the GSEA leading edge from RNA-seq and proteomics data were incorporated into a protein-protein interaction map (**figure 5.21**). This revealed a complex network, that when simplified, identified components of glucose metabolism, mitochondrial respiratory chain, and fatty acid metabolism, in addition to adipocyte differentiation. Many components of the network were involved in mitochondrial biogenesis and function. This included a number of proteins involved in the mitochondrial respiratory chain, such as subunits of complex I (NDUFB7, NDUFA5, NDUFAB1, and NDUFS3), complex III (UQCR10 and UQCRQ), complex IV ATP synthase (ATP5O), and cytochrome c oxidase (COX6A1). Mitochondria have an important role in the regulation of adipocyte differentiation, as during adipogenesis, increased oxygen consumption and mitochondrial biogenesis has been observed (Lu et al. 2010; Shi et al. 2008). Concurrently, inhibition of mitochondrial function results in diminished ability to store triglycerides and prevented differentiation of MSCs into adipocytes (Zhang et al. 2013). The upregulation of mitochondrial function in fed BBS vs control cells is also supported by positive enrichment of oxidative phosphorylation by GSEA in both transcriptomic and proteomic datasets (**table 5.6**).

A network containing components required for fatty acid metabolism was enriched in both RNA-seq and proteomic adipogenesis gene sets. These mostly included those involved in fatty acid β-oxidation, such as *ACAA2, ACADS, ECH1, HADH, ACOX1*, and *SLC25A1*. Fatty acid β-oxidation is

the mitochondrial generation of acetyl-CoA from triglycerides for energy production, and is particularly high in brown adipose tissue for thermogenesis (Festuccia et al. 2011). A significant increase in brown, as well as white adipose tissue was demonstrated in the *Bbs1* M390R knock-in mouse (Davis et al. 2007). Overall, there were many genes within the positively enriched adipogenesis gene set related to mitochondrial function. As fatty acid metabolism and multiple other mitochondrial components were featured in the leading edge of both datasets, it may suggest that upregulation of adipogenesis genes in fed BBS cultures was driven by unbalanced mitochondrial activity and/or fatty acid homeostasis.

MSCs can give rise to both fibroblasts and adipocytes (Friedenstein 1976), and reports suggest that morphologically, fibroblasts can be difficult to distinguish from preadipocytes (Ali et al. 2013). Furthermore, dermal fibroblasts can be reprogrammed into adipocytes, either by overexpression of *PPARG* (Tontonoz et al. 1994) or by exogenous adipogenic induction through cell culture (Marion et al. 2009). With this knowledge and upregulation of proadipogenic factors within this study, it is tempting to speculate that serum fed BBS fibroblasts exhibit a gene expression signature similar to differentiating adipocytes. Specific markers of adipocytes, such as *FABP4, PEPCK, CEBPA, GLUT4*, and *LPL* (Moseti et al. 2016), weren't detected in transcriptomic and proteomic data (supplemental data, table S.5, table S.32), and fibroblast markers such as fibroblast specific protein 1 (FSP1, encoded by *S100A4*) and collagen genes (Strutz et al. 1995) were highly expressed, demonstrating that cultures were fibroblastic in nature. However, it is possible that BBS fibroblasts exhibited a basal level of proadipogenic expression, demonstrated by upregulation of *PPARG* and *LEP* and mitochondrial processes. This finding would support other studies that suggest obesity is not exclusively a consequence of leptin insensitivity, but may also be linked to aberrant cellular processes leading to increased adiposity (Marion et al. 2009; Marion et al. 2012).

Further experiments would be required to determine whether BBS patients' fibroblasts have a predisposition to adipogenesis. Firstly, validation of markers, such as *PPARG* and *LEP*, in fibroblasts would be essential in a larger cohort to determine whether there is a significant increase in gene expression compared to controls. Secondly, it would be an interesting experiment to induce differentiation of BBS fibroblasts into adipocytes and serially follow gene expression of adipocyte markers, with the hypothesis that BBS fibroblasts have a higher differentiation efficiency than controls. The finding of dysregulated mitochondrial biogenesis and fatty acid metabolism could also be validated within the same experiment. Similar studies have

previously demonstrated that ablation of *BBS10* and *BBS12* increased leptin and *PPARG* expression in differentiating adipocytes, which was coupled to downregulation of antiadipogenic signalling pathways, such as Wnt and HH (Marion, Mockel, et al. 2012). Experiments such as this have yet to be performed with patient-derived cells carrying the *BBS1* p.M390R variant, and thus may provide further insight into high prevalence of obesity in patients.

5.3.2 Analysis of the effects of serum starvation on BBS and control fibroblasts

The experimental design for this study enabled comprehensive analysis of fibroblasts in either cell cycling or ciliated states. It also provided an opportunity to detect differences between starved and fed cells, thereby identifying putative cellular defects that may present in response to serum starvation. Serum starvation is a common method for induction of ciliogenesis in cell culture, as it stimulates cell cycle arrest and cellular quiescence (Tucker et al. 1979; Uetake et al. 2007). It is also used routinely to reduce basal cellular signalling activity (Pirkmajer and Chibalin 2011). GSEA revealed many parallels between BBS starved vs fed and control starved vs fed cultures, demonstrating cellular similarities in regulation as a response to starvation. These included a highly significant negative enrichment of processes associated with the cell cycle, including mitotic spindle, G2M checkpoint, and E2F targets, and downregulation of other key pathways, such as HH, mTORC1, hypoxia, and unfolded protein response. Two differences were identified between BBS and control for starved vs fed comparisons, namely myogenesis and inflammatory response, which were both negatively enriched in BBS and positively enriched in control experiments. Further elucidation is required to determine the significance of these findings, as there is little evidence of cilia-specific dysregulation of these processes in the literature.

GSEA provided a broad overview of enriched cellular processes, but did not identify biologically interesting differences between BBS and control experiments. Therefore, an alternative pathway analysis was applied, PANTHER enrichment test using the Reactome pathways database. This identified that “ciliary part” was upregulated in BBS starved vs fed, but “assembly of the primary cilium” was downregulated in control starved vs fed. As serum starvation intended to induce ciliary assembly, it was unexpected that this pathway was downregulated in controls, and therefore was explored further by investigation of complexing interactions for DE genes. Annotation of differential regulation for DE genes found in patients, controls, and both, revealed clear complexes of genes that were up and downregulated in response to starvation for induction of ciliogenesis. This highlighted a limitation of pathway analysis, where the quality of the gene set

determines the efficiency of the analysis. In this case, the gene set for "assembly of the primary cilium" contained components that would be expected to be upregulated, such as the BBSome and IFT machinery. Simultaneously, genes involved cell cycle and centrosome maturation, which play an important role in ciliary assembly, would be expected to be downregulated (Pedersen et al. 2008). This shows the importance of referring back to the gene level after pathway analysis.

Within the network of "assembly of the primary cilium" there were a larger number of DE genes from the BBS starved vs fed experiment than the control group. However, this may be due to statistical bias created by analysing a larger patient than control group. This was a known limitation of the study design, and was not possible to overcome due to limited access to control material at the time of conducting experiments. However, one may be able to infer differences if seen in the control group and not patients. 10 genes that interacted within the complex of transition zone/centrosome were downregulated in control and not in the BBS experiment. These included centrosomal genes involved in microtubule nucleation and centrosome integrity, or were negative regulators of ciliogenesis.

The gene products of 6 genes identified in the complex are involved in mitotic spindle dynamics and centrosome integrity. HAUS3, HAUS6, and HAUS8 form a complex responsible for mitotic spindle generation and cytokinesis (Uehara et al. 2009; Lawo et al. 2009). The complex localises to interphase centrosomes and regulates the formation of mitotic spindles. Deletion of HAUS subunits was shown to result in delayed attachment of microtubules to kinetochores and diminished centrosome integrity (Lawo et al. 2009; Uehara et al. 2009). CKAP5 also plays an essential role in dynamics, organisation, and nucleation of the mitotic spindle (Gergely et al. 2003). CEP72 localises to the pericentriolar material (PCM) of the centrosome and acts to recruit other centrosomal proteins for centrosome duplication (Kodani et al. 2015). Furthermore, *CEP72* overexpression was shown to induce chromosomal instability and abnormal mitotic microtubule assembly (Lüddecke et al. 2016). CEP152 directly interacts with CEP72 and functions together to promote centrosome duplication (Kodani et al. 2015; Brown et al. 2013). CEP152 also acts as a molecular scaffold to orchestrate the interaction between PLK4 and CENPJ (Cizmecioglu et al. 2010), both of which are also involved in centrosome duplication and were downregulated in response to starvation (**figure 5.14**). Analysis of transcriptomic data for these genes identified that gene expression was lower in fed BBS cultures than fed controls, where expression of *HAUS3* and *CEP152* was significantly reduced (**figure 5.15**). The finding that mitotic spindle organisation

is disrupted in cycling BBS cells was supported by GSEA, which established a significantly negative enrichment of the mitotic spindle gene set for fed BBS vs control.

CEP97 and *KIF24,* which negatively regulate ciliogenesis, were also significantly downregulated in control, but not BBS, serum starved cultures. CEP97 forms an inhibitory complex with CCP110, which prevents premature ciliogenesis (Spektor et al. 2007). KIF24 mediates this process by ensuring destabilisation of microtubules at the mother centriole, to prevent nucleation of the basal body (Kobayashi et al. 2011). Furthermore, upon re-entry into the cell cycle, *KIF24* is upregulated to stimulate depolymerisation of microtubules at the basal body, a process driven by the kinase, NEK2 (S. Kim et al. 2015). Thus, cilia assembly is induced by inactivation of the CCP110-CEP97 complex and KIF24, via phosphorylation by kinases, such as TTBK2 (Čajánek and Nigg 2014). Downregulation of *CEP97* and *KIF24* is therefore a logical finding for cells stimulated for ciliogenesis by serum starvation. Interestingly, *CCP110* is significantly upregulated in BBS starved vs fed. The fact that *CCP110, CEP97*, and *KIF24* expression is reduced or opposite to what is expected for the assembly of the cilium, may suggest that the process of inhibition of ciliogenesis is dysregulated.

5.3.3 Many upregulated proteins linked to ciliary pathways in starved BBS vs control

RNA-seq analysis detected the global mRNA transcripts to provide a comprehensive snapshot of the cellular physiological state. Identification of DE genes revealed that there were large differences between starved and fed cells, resulting in up to 6000 DE genes. On the other hand, there were smaller differences between BBS and control fibroblasts. There were only 29 DE genes for starved BBS vs control comparison, which was anticipated to be the most biologically relevant comparison to this study. The reason for the smaller number of DE genes may be due to one of two explanations. Firstly, starvation may have reduced the basal cellular activity to a state where the BBS and control samples were more homogeneous, and thus differential expression could not be inferred. The second explanation may be for the opposite reason; starvation may have resulted in a greater level of heterogeneity across all cells, that transcript levels within biological groups were too variable to identify statistical differences, particularly with a small number of samples. The issue seen for transcripts of starved BBS vs control was not observed with the proteomics dataset. This is possibly due to the greater stability of proteins compared to mRNA, where inside a cell, mRNA has a half-life of several hours, whereas proteins can last up to 2 days (Schwanhäusser et al. 2011).

PANTHER enrichment testing was used to simplify the data and categorise DE proteins into biological groups. Out of 18 pathways from the Reactome database shown to be significantly enriched, 5 were recognised as being involved in the regulation or maintenance of the primary cilium, namely "assembly of the primary cilium", "hedgehog 'off' state", "RHO GTPase effectors", "centrosome maturation", and "cytoskeletal regulation by RHO GTPases". The DE proteins enriched for these pathways were predicted by pathway topology mapping to be highly interconnected. It was then considered how BBS1 and its interacting partners were incorporated into the network, with the postulation that the *BBS1* p.M390R variant was driving the pathways to be upregulated in contrast to controls. Due to the complexity of the network presented in figure 5.20, it was challenging to make reliable inferences without further validation. However, the implication of the interaction of BBS1 within the pathway network has been considered and shall be discussed here.

BBS1 was incorporated into the network via LZTFL1, and the GPCR that controls the induction of HH signalling, SMO. LZTFL1 is a cytoplasmic protein that has been hypothesised to interact with BBSome as a negative regulator of its trafficking to the ciliary membrane. Together LZTFL1 and the BBSome were found to regulate trafficking of SMO in the primary cilium (Seo et al. 2011). Furthermore, *LZTFL1* was recognised as the 17th gene to manifest BBS in a consanguineous family, with unusual presentation featuring situs inversus and insertional polydactyly, in addition to visual, renal, endocrine, and cognitive phenotypes (Marion, Stutzmann, et al. 2012). These BBS-like features were shown to be caused by defective regulation of SMO, as loss of function of *LZTFL1* resulted in increased *SMO, PTCH1*, and *GLI2* expression (Marion, Stutzmann, et al. 2012). Within the present study, SMO was classified to function within the "HH 'off' state", due to the presence of negative regulators of HH, such as KIF7, NUMB, and the proteasome (Liem et al. 2009; Di Marcotullio et al. 2006; Gerhardt et al. 2016). For example, KIF7 is thought to localise to the basal body and block entry of GLIs into the primary cilium, which promotes degradation of GLIs by transporting them away from the base of the cilium (Liem et al. 2009; Goetz and Anderson 2010). To determine whether "HH 'off' state" is a true finding, it is crucial to understand the localisation of proteins, such as SMO and the proteasome; information that mass spectrometry cannot provide. Localisation of some proteins determines their function. For example, if SMO has accumulated in the cytoplasm of the cell, then this induces processing of repressor GLIs. However, SMO localises to the ciliary membrane upon binding of HH ligand to PTCH1, which activates GLIs and triggers downstream HH pathways (Rohatgi et al. 2007; Goetz and Anderson 2010) (see figure 1.2). As it requires additional purification steps to isolate membrane bound

proteins (Mirza et al. 2007), within this experiment it is more likely that SMO has been detected in the cytoplasm. This would need to be verified by protein localisation experiments.

BBS1 was also interconnected to the network via direct interaction with RABIN8 (encoded by *RAB3IP*), which is a guanine nucleotide exchange factor (GEF) required for activation of RAB8 (Hattula et al. 2002). The interaction of the BBSome, RABIN8, and RAB8 is crucial for the mediation of vesicle docking and fusion to the ciliary membrane (Nachury et al. 2007; Jin et al. 2010). Furthermore, relocalisation of RABIN8 to the centrosome by interaction with another small GTPase, RAB11, is required for initiation of ciliogenesis (Knödler et al. 2010; Westlake et al. 2011; Feng et al. 2012). This occurs upstream of the interaction with BBS1 (Westlake et al. 2011). Although the majority of BBSome components, such as *BBS2, BBS4*, and *BBS7,* are not essential for cilia assembly, knockdown of *BBS1* resulted in reduction of cilia on RPE cells (Loktev et al. 2008; Jin et al. 2010). This evidence suggests that the interaction between RABIN8 and BBS1, but not other components of the BBSome, is necessary for cilia assembly (Nachury et al. 2007; Jin et al. 2010). Within the network, the connection with RABIN8 links BBS1 to components of "centrosome maturation" and "assembly of the primary cilium", such as CEP290, OFD1, EXOCs, and ARF4, but also proteins involved in "RHO GTPase Effectors", including MAPRE1, CLASP1, YWHAG, and YWHAE.

RHO GTPases act as a molecular switch to control cytoskeletal dynamics and cellular remodelling (Pan et al. 2007). RABIN8 has been shown to modulate actin organisation and reduce stress fibre formation (Hattula et al. 2002). As a result, it has been suggested that RABIN8 modulates delivery of vesicles to the ciliary membrane through reorganisation of actin and microtubules (Hattula et al. 2002). This is supported by RABIN8 interaction with the exocyst complex, featuring EXOC1-8 (Feng et al. 2012), which in turn interacts with CDC42 to regulate the cytoskeleton (Zhang et al. 2001). Components of the "RHO GTPase effectors" that are involved in actin remodelling were significantly upregulated in starved BBS vs control, such as CDC42, ROCK1/2, RHOC, and members of the ARP2/3 complex, including ARPC2-5. The loss of BBS1 has been demonstrated to increase expression of RHO GTPases, which increased actin fibres and prompted defective ciliogenesis. This led to the conclusion that BBS proteins may also be involved in actin regulation (Hernandez-Hernandez et al. 2013), and that actin can negatively modulate ciliogenesis (Kim et al. 2010; Pan et al. 2007). The mechanism for how BBS1 regulates the cytoskeleton is not understood, however it is possible that it occurs through RABIN8 and its many effectors related to cytoskeletal organisation.

The results presented by pathway enrichment and topological mapping analysis has successfully categorised a large complex dataset and presented the data to aid hypothesis generation. The main finding is that BBS1 interacts indirectly with members of 5 pathways related to ciliary function, that were statistically positively enriched in starved BBS vs control experiments. This occurs through direct interaction with SMO or RABIN8. As a result, a new hypothesis was developed stating that the observed upregulation of the components in the presented pathways are due to reduced function of BBS1, caused by harbouring the variant p.M390R. This may be driven by diminished affinity for BBS1 with its interactors. It has previously been shown that p.M390R induces a conformational change in BBS1 that reduces its affinity with ARL6 (Mourão et al. 2014). Therefore, it is possible that p.M390R also affects the direct interaction with RABIN8, which results in the accumulation of proteins in BBS cells. The RNA-seq data may also support this hypothesis, as a greater number of upregulated transcripts for "assembly of the primary cilium" were found in BBS starved vs fed compared to control starved vs fed. As previously discussed, this may be a statistical artefact, thus must be validated in a larger sample size with a greater number of control samples.

5.3.4 Experimental limitations

5.3.4.1 Pathway enrichment analysis

Statistical enrichment tools, such as GSEA and PANTHER used within this chapter, addressed limitations of other pathway analyses, such as overrepresentation testing, by not restricting the data by p-value thresholds (Khatri et al. 2012). In the enrichment tests, genes are not treated equally but ranked by a metric, such as log2(FC) used in this study. Although a more reliable analysis than overrepresentation testing, there are limitations. It is commonly understood that a gene or a complex of genes can function in multiple different pathways; however, enrichment testing considers each pathway independently. For example, the proteasome is a negative regulator of multiple cellular processes, including Wnt signalling, HH signalling, cell cycle, apoptosis, and cancer associated pathways (Adams 2003; Malicki and Johnson 2017). Therefore, differential regulation of a component such as the proteasome may return many significant pathways, where it is challenging to ascertain whether they are true findings without functional determination.

5.3.4.2 Pathway topology mapping analysis

Gene enrichment analysis is well complemented by pathway topology mapping, as it demonstrates how pathways are interconnected and cross-talking. As demonstrated in this thesis, it also allows identification of complexes that are up or downregulated, or that are contributing the most to significant enrichment. The utility of topological analysis relies on well characterised pathways, where some interactions will be more comprehensively studied than others. For example, within the STRING database, interactions involved in cilium assembly were extensively documented by experimental data or in curated databases. On the other hand, interactions for genes involved in adipogenesis were less confident, thus to strengthen the network generated for **figure 5.21**, connections obtained from text mining were also included.

5.3.4.3 Database annotation

Pathway analysis can be an invaluable tool for simplifying overwhelmingly large datasets into biologically meaningful and statistically significant groups, as has been demonstrated within this study. However, their utility is limited by the robustness of databases and gene sets, which are currently struggling to keep up with the high-resolution datasets generated by NGS technologies. Data used to create gene sets and pathways are typically generated from gene expression experiments implemented by different labs under different conditions, often with different cell types (Khatri et al. 2012). Although most databases have been manually curated to ensure gene sets are reliable and non-overlapping (Liberzon et al. 2015; Croft et al. 2011), databases are currently lacking information about the exact conditions experiments were undertaken. This is particularly relevant to tissue specific findings, as pathways may be distinctive in different cells due to cell-specific gene expression signatures. Furthermore, different transcript isoforms cannot be distinguished by databases, thus are mapped to the same gene, removing tissue or condition specific information (Khatri et al. 2012).

Another obvious limitation is that pathway analysis can only be completed with genes that are annotated. This immediately removes all non-protein coding transcripts, the function of which are predominantly undefined. Therefore, within this study this immediately excluded ~2000 transcripts from RNA-seq experiments from consideration. Moreover, PANTHER is comprised of 13,387 human protein-coding genes (Mi et al. 2017) meaning that ~7000 genes are not annotated in PANTHER. These limitations are well comprehended by curators and are being addressed by new database builds, which are frequently being released.

5.3.4.4 Limits of the study design and future work

This study was designed to observe the differences between cell cycling and ciliated fibroblast cultures between BBS and controls. This identified thousands of differentially expressed genes and proteins when serum starved and serum fed cells were compared. There was less significant variation between BBS and control groups, which was reflected by fewer differentially expressed genes. One reason for fewer significant differences between biological groups is likely to be heterogeneity between different samples. Regulation of gene expression may be affected by genetic background or environmental factors, thereby generating variability between different cultures within the same biological group. In order to reduce this variance, greater sample sizes would be required to increase statistical power.

In total, 14,573 protein-coding transcripts and 6,683 proteins were detected from fibroblasts, generating a 46% rate of protein detection for expressed genes. This is a high detection rate, which can be attributed to the reduced sample complexity achieved by separating samples into 20 peptide fractions each. On the other hand, many BBS proteins, including BBS1 and its interactors, were not detected by iTRAQ proteomics. This may be due to the low expression levels of these genes in fibroblasts (for example, *RAB3IP* mean log2CPM= 1.77; *BBS1* mean log2CPM= 1.88; whereas *GAPDH* mean log2CPM= 10.45, supplemental data, **table S.2**). It is possible to increase the depth of coverage for a single cell type, as 10,255 proteins were identified from HeLa cells (Nagaraj et al. 2011). However, this required separation into 72 fractions, which accounted for 288 hours of mass spectrometry instrument time (Nagaraj et al. 2011). This depth of coverage was not realistic to obtain for this project, due to the expense of fractionation and extensive instrument running time required for the number of samples used. Furthermore, sample pooling may have also reduced the sensitivity of the mass spectrometer, as sample complexity would have increased. This was an unavoidable step within this experiment due to the high cost of iTRAQ proteomics technology. The main limitation of pooling was that protein expression for samples could not be considered individually, and thus it was not possible to identify the true variance of the experiment.

Fibroblasts may not be considered as a disease relevant cell type, as BBS patients do not have a known dermal phenotype. However, they were used in this study as they can be isolated from a dermal skin biopsy that can be taken from patients in the clinic via a simple procedure. This results in a cellular disease model that, in theory, harbours the same genetic background of each patient. Furthermore, fibroblasts were also suited to this study design as they are well-known to ciliate

efficiently under serum starved conditions (Schneider et al. 2005; Christensen et al. 2013). In this study, these cells presented a valuable cellular resource that were manageable to culture to vast numbers when large quantities of RNA and protein were required. They provided an appropriate cellular model for a large-scale pilot project, such as this one, where implementation of novel techniques may have squandered precious patient tissues if experimental strategies had not gone to plan. This study has subsequently allowed development of multi-omic methodology for future work, where more disease relevant tissue, or models derived from iPSC differentiations, may be utilised.

Even though the skin does not exhibit a known molecular phenotype in BBS patients, this study has uncovered some interesting findings that would warrant further research. Pathway analysis of iTRAQ proteomics data for starved BBS vs control comparisons identified the upregulation of key modulators of ciliogenesis, including those involved in cilium assembly, actin cytoskeleton remodelling, and HH signalling. I have hypothesised that this upregulation may be an effect of reduced affinity of BBS1 with its interactors, such as SMO or RABIN8, caused by the p.M390R variant. Firstly, this finding would need to be validated in a larger sample size to determine whether this collective upregulation is reproducible. Next, it would be interesting to establish whether this difference is enhanced in BBS fibroblasts when stimulated to induce ciliary trafficking. Smoothened agonist (SAG) may present an efficient method to accomplish this, as it induces an upregulation of SMO trafficking to the primary cilium (Sinha and Chen 2006). Ciliary localisation of SMO was shown to be depleted in cells lacking *BBS1*, both with and without SAG treatment (Seo et al. 2011). Therefore, this would be a suitable agonist to test this hypothesis. Preliminary testing would investigate the localisation of SMO in response to SAG in *BBS1* p.M390R fibroblasts compared to control by immunochemical staining. If this is successful, a targeted assay could be developed, featuring the upregulated proteins in key pathways identified in this study, to establish whether treatment induces a significant increase in gene and protein expression in BBS compared to control cells.

Chapter 6: Generation of an *in vitro* retinal model from BBS patient-derived induced pluripotent stem cells for the study of retinal dystrophy

6.1 Background

6.1.1 Development of the eye

There are three primary germ layers that can be produced from a human blastocyst after fertilisation, during the phase of gastrulation. These layers are the endoderm, being the inner most layer that gives rise to epithelial linings including most organs; the mesoderm, which forms mesenchymal cells, muscles, and aspects of the cardiovascular system; and the ectoderm, the outer most layer that manifests as the central and peripheral nervous systems, sensory epithelia, and the epidermis. These germ layers are formed from embryonic stem cells that, driven by chemical factors such as morphogens, eventually differentiate into their terminal cell fate. Two germ layers are required to interact to form the complex vertebrate eye, which are formed from the neural ectoderm, the surface ectoderm, and the periocular mesenchyme (Heavner and Pevny 2012).

Genes involved with eye development are highly conserved across all bilaterally sighted species. The importance of these key developmental genes is emphasised by the fact that aberrant variation can lead to severe congenital defects (Graw 2003; Danno et al. 2008; Azuma et al. 1999). A series of intricately balanced transcription factors coordinate to form a regulatory network for accurate development of different eye components. The first set of these transcription factors activate to molecularly define the eye field, a single region within the anterior neural plate of the developing forebrain, which are triggered shortly after gastrulation. These are termed the eye field transcription factors (EFTFs), which feature *PAX6, RX, SIX3,* and *LHX2,* and are coordinated by *SOX2* and *OTX2* (Zuber et al. 2003). Following the establishment of the developmental midline, the single eye field is divided into bilateral hemispheres, coordinated by *SHH* and *SIX3*, to make two distinct sites for eye development (Jeong et al. 2008). Evagination of these individual sites then leads to the formation of the optic vesicle.

The formation of the eye structure requires a double invagination caused by close mutual interactions between the neural and surface ectoderms (Heavner and Pevny 2012; Graw 2010)

(figure 6.1). The surface ectoderm invaginates and separates to form the lens, leaving behind what then becomes the cornea and the epithelial layers at the front of the eye. Simultaneously, the neural ectoderm forms from the optic vesicle, an outgrowth of the neural tube, which invaginates to form the optic cups, comprised of the neural retina (NR), retinal pigmented epithelial (RPE) cells, and other epithelial layers continuous with the retina. The periocular mesenchyme is a population of neural crest cells that differentiate into cells responsible for formal ocular development, such as the stroma, ciliary body and extraocular muscles, blood vessels, and endothelium of the eye.

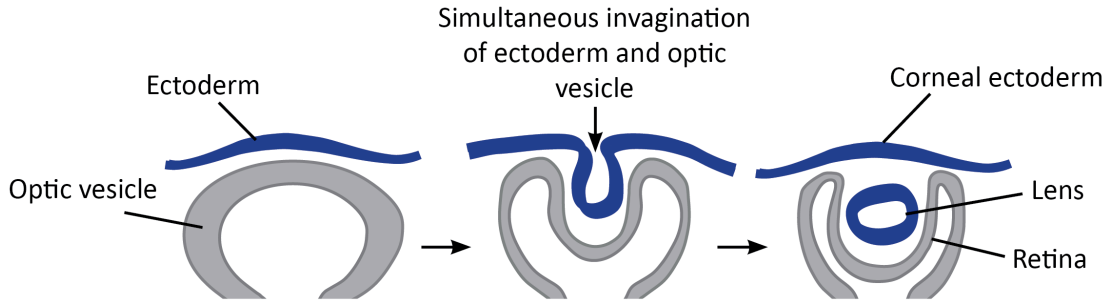

Figure 6.1 – A schematic to depict the formation of the retina and lens from the optic vesicle and surface ectoderm during development.

6.1.2 The retina

The retina develops from the two layers of the optic cup, where the inner layer represents the NR and the outer layer the RPE. The prospective NR is first characterised by the expression of *VSX2*, whereas the future RPE expresses *MITF* (Horsford et al. 2005; Capowski et al. 2014). These two genes have an antagonistic relationship, which essentially creates a partition between the NR and the RPE to allow for separate differentiation. *MITF* is essential for RPE specification and expression occurs prior to *VSX2*, where upon *VSX2* upregulation, *MITF* is repressed in the neural retina (figure 6.2). *In vivo*, this phenomenon is orchestrated by the surface ectoderm, which secretes fibroblast growth factors, FGF1 and FGF2 (Horsford et al. 2005). The RPE has a fundamental role in maintaining and supporting the cells within the retina (Marmorstein et al. 1998), and is recognisable by its distinctive dark appearance. An intimate mutually beneficial relationship exists between photoreceptors and the RPE; deleterious mutations in one cell type results in degeneration of the other (Gu et al. 1997).

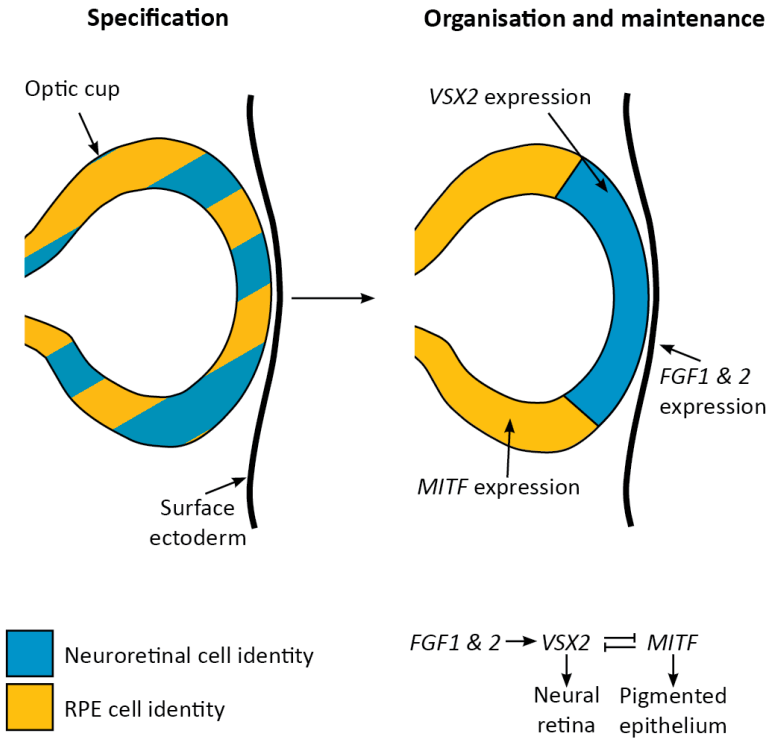

Figure 6.2 – Specification and organisation of the optic cup, where *MITF* and *VSX2* expression dictate the formation of the RPE (yellow) and NR (blue), respectively. This process is orchestrated by FGF1 and FGF2, expressed by the surface ectoderm. Adapted from Horsford et al. (2005).

The mature retina consists of three layers of laminar organisation; the outer nuclear layer, where the sensory photoreceptors reside; the inner nuclear layer, represented by the processor interneurons called horizontal, bipolar, and amacrine cells; and the retinal ganglion cell layer, which is responsible for transmitting the chemical signal to the visual centres in the brain (Heavner and Pevny 2012). Additionally, the retinal extracellular environment is maintained by specialised Müller glial cells, which span the entirety of the neural retina and regulates various homeostatic and metabolic pathways (Luo et al. 2008). Neural retina progenitor cells differentiate in a temporal manner that is conserved across vertebrates, where each cell type is characterised by certain cell markers (summarised in **figure 6.3**).

The photoreceptor exists as two cell types, namely rods and cones. The rod photoreceptors are responsible for scotopic vision and are highly sensitive to low light levels (100 fold more sensitive than its cone counterpart (Luo et al. 2008)), which is detected by the only visual pigment expressed in these cells, called rhodopsin. Conversely, cone photoreceptors are receptive to conditions of bright light, in addition to providing visual acuity and colour perception. Cones are accountable for trichromatic vision, as each cone cell expresses one of three visual pigments; S-opsin absorbs short-wavelength 'blue' light; M-opsin, sensitive to medium-long wavelength

'green' light; and L-opsin, which allows detection of long wavelength 'red' light (Nathans et al. 1986).

During human development, differentiation from progenitor to mature rod photoreceptor occurs over several weeks. At fetal week (Fwk) 10.5, rods are identified by expression of the rod-specific transcription factor *NRL*. By Fwk 11-12, markers indicative of functionality appear, such as recoverin (encoded by *RCVRN*), *CRX*, and *IRBP*. It is not until Fwk 15 that proteins essential for photosensitivity, including rhodopsin, emerge (Hendrickson et al. 2008). Conversely, cone photoreceptors start to materialise at fwk 8, with S-opsin expressing from fwk 11, which is earlier than rods despite being outnumbered 20:1 in the mature human retina. In maturing photoreceptors, outer segment elongation initiates at ~fwk 32 (Vajzovic et al. 2012) (**figure 6.3**).

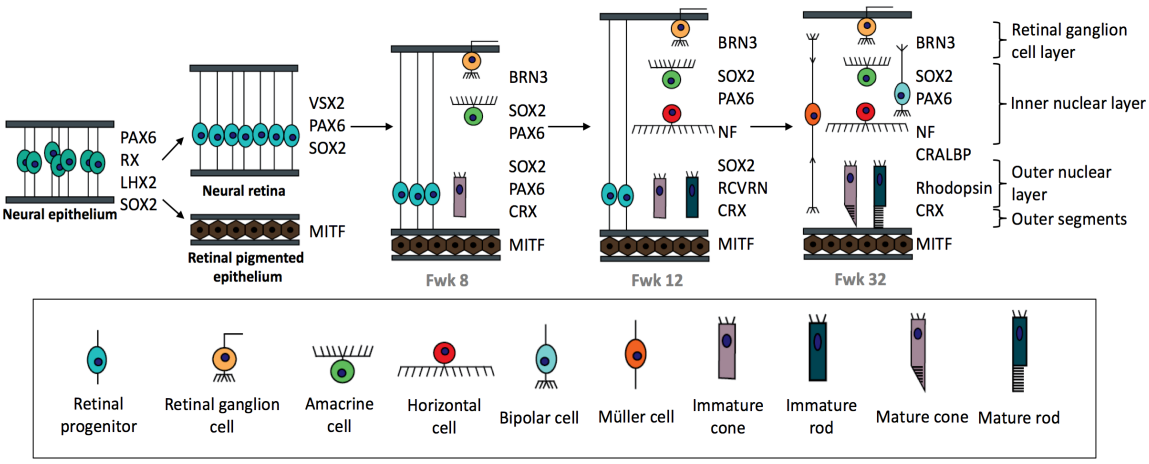

Figure 6.3 – The laminated structure of the retina. Differentiation of the retinal cell layers from the neural epithelium in human development, where key markers are expressed throughout development, indicative of different cell types.

6.1.3 The photoreceptor and the cilium

Blindness caused by photoreceptor degeneration is one of the most genetically heterogeneous disorders in humans with over 250 genes identified, with a prevalence of around 1 in 3000 (Hartong et al. 2006; Wright et al. 2010). The largest category of gene mutations that cause photoreceptor degeneration are involved in ciliary trafficking, accounting for 20% of these 250 genes (Daiger et al. 1998). This demonstrates that there is obviously a link between the photoreceptor and the cilia. The answer to understanding this link can be found by studying the photoreceptor structure in more detail.

The photoreceptor cell stretches across the outer nuclear layer of the retina, creating synaptic connections with neurons found in the inner nuclear layer, such as bipolar and horizontal cells. The cell projects deeper into the retina, forming two distinct compartments; the mitochondria rich inner segment and the photosensitive outer segment. The outer segment is a dynamic structure consisting of layers of membranous discs, maintenance of which requires rapid morphogenesis and turnover of proteins. In fact, it can turn over up to 10% of its volume per day to prevent accumulation of proteins that may be damaged by photo-oxidation (Besharse et al. 1977). One role of the RPE is to phagocytose the distal outer segment to prevent accumulation of damaged proteins through photooxidative stress (Finnemann et al. 1997). The outer segment, however, does not possess its own machinery to synthesise essential proteins and so they must be delivered from the inner segment via the connecting cilium (**figure 6.4**).

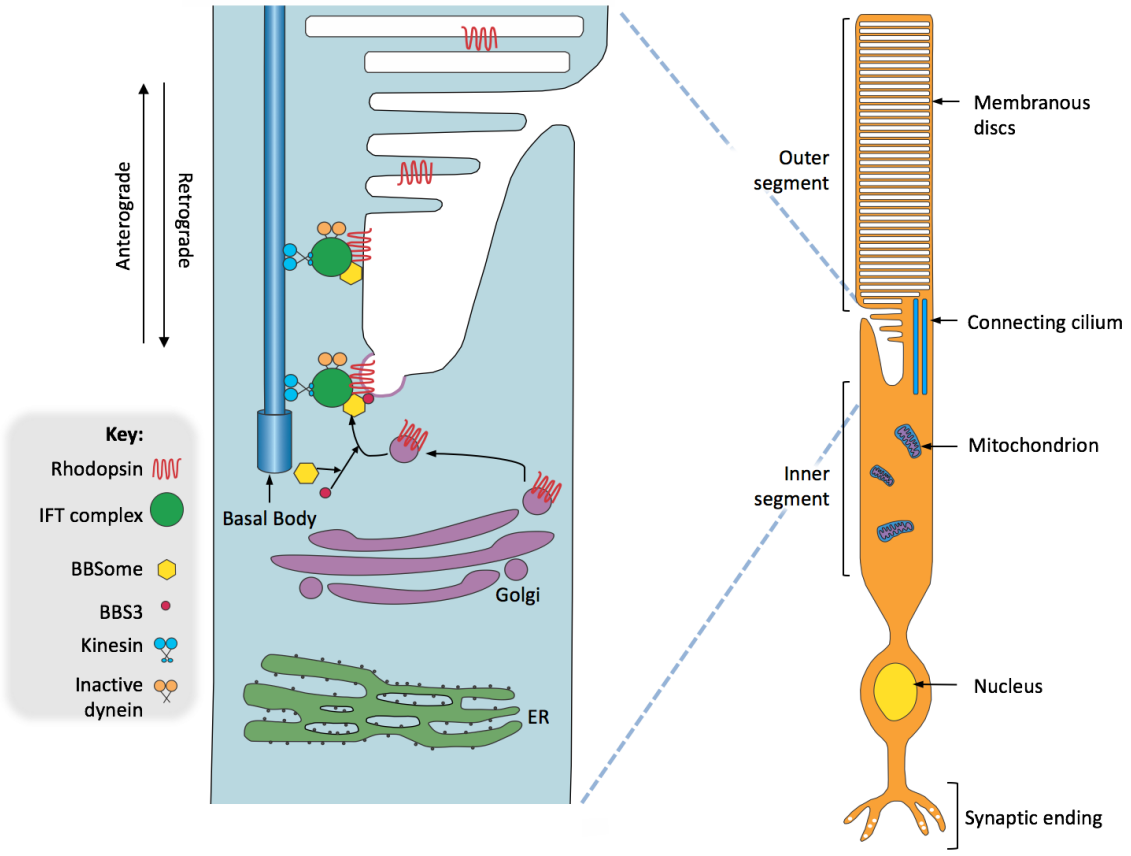

Figure 6.4 – Schematic of a rod photoreceptor and the connecting cilia machinery. Phototransduction proteins, such as rhodopsin, are synthesised and folded in the inner segment of the rod photoreceptor. The BBSome recruits the folded protein for entry into the connecting cilium and is trafficked to the outer segments by anterograde intraflagellar transport (Wright et al. 2010).

The presence of the cilium in the photoreceptor means that the outer segment behaves as a complex specialised cilium, the axoneme of which stretches more than half the length of the

outer segment in rods and the entire length in cone cells (Luby-Phelps et al. 2008). Under a light stimulus, the connecting cilium is vital for rapid shuttling of membrane-bound and membrane-associated proteins that are essential for phototransduction, such as rhodopsin, to the optic discs that run perpendicular to the axoneme (Wolfrum and Schmitt 2000; Smith et al. 2013). The connecting cilium is also required for cell maintenance, as well as retrograde transport of intraflagellar transport (IFT) machinery (Krock et al. 2009). Although very similar in structure to the standard primary cilium, the connecting cilium is essentially the structural equivalent to the transition zone of primary cilia on other cell types (Mykytyn et al. 2004; Pazour et al. 2002). Through this organelle, the turnover of proteins must be efficient and well organised as each minute an estimated 2000 opsin molecules are transported to the outer segment through a narrow space of only 0.25µm in diameter (Besharse et al. 1985). As one would surmise, aberrations in trafficking within this complex organelle can lead to severe complications.

6.1.4 Impacts of impaired connecting cilium

Evaluation of ciliopathy *in vivo* models can provide an insight into the pathobiology of retinal degeneration when IFT is disrupted. Aberrant anterograde IFT in mice, induced through knock out of *Kif3a*, led to complete mislocalisation and accumulation of key phototransduction proteins in the inner segment, ultimately resulting in progressive photoreceptor death (Marszalek et al. 2000). Many groups have recapitulated this study with disruption of other principle IFT components, such as IFT88 (Pazour et al. 2002), IFT122 (Boubakri et al. 2016; Keady et al. 2011; Ross et al. 2005; Kulaga et al. 2004), and IFT20 (Keady et al. 2011). Although this shows a clear link with IFT function, the BBS mouse model shows that regulation of vesicle trafficking and entry/exit to IFT has an impact on photoreceptor structure and degeneration. For example, the *Bbs6*$^{-/-}$ mouse shows a progressive loss of photoreceptors where onset starts at 3 weeks of age and continues until total loss of the outer segment by 26 weeks of age (Ross et al. 2005). Additionally, transmission electron microscopy (TEM) shows structural disorganisation and fragmentation of outer segments at 13 weeks of age (Ross et al. 2005). This pattern is mimicked in the *Bbs1*$^{-/-}$, *Bbs2*$^{-/-}$, *Bbs4*$^{-/-}$, and *Bbs1*$^{M390R/M390R}$ knock-in mice (Nishimura et al. 2004; Mykytyn et al. 2004; Davis et al. 2007; Kulaga et al. 2004). Although these studies closely imitate the progressive loss of eyesight seen in BBS patients, the exact mechanism for photoreceptor death still remains uncertain.

Accumulation of phototransduction proteins in an erroneous compartment in the cell, such as rhodopsin and arrestin localised across the inner segment and outer nuclear layer, is commonly

seen in BBS mutant models (Louie et al. 2010; Mockel et al. 2012). It is thought that this is due to impaired docking and fusion caused by the faulty BBS proteins, which leads to high toxicity, and subsequent cell death (Moritz et al. 2001). However, it cannot be ascertained in these studies whether the BBS mutation was the primary cause of the mislocalisation, or these proteins have become misplaced due to the cell undergoing apoptosis. Conversely, there is evidence that the BBSome interactor Rab8 docks and fuses with rhodopsin-bearing vesicles, thus providing a functional link between rhodopsin and the BBSome (Moritz et al. 2001). Furthermore, a study featuring a *BBS4*$^{-/-}$ mouse found that even as early as 2 weeks of age mislocalised opsin proteins were seen and, to a lesser extent, some apoptotic nuclei (Abd-El-Barr et al. 2007). They hypothesised that since the quantity of mislocalised proteins was greater than the latter, apoptosis must succeed this aberrant localisation, and thus photoreceptor death occurs as a consequence of defectively trafficked proteins.

Datta et al. (2015) have suggested that the cause of photoreceptor death is not due to mislocalisation of phototransduction proteins, but of aberrant protein localisation to the outer segment (Datta et al. 2015). In contrast to other studies, they found that rhodopsin did not noticeably accumulate in the inner segment of a *Lztfl1/(Bbs17)*$^{-/-}$ mouse retina. In the outer segment, however, there was an enrichment of membrane fusion Syntaxin proteins, Stx3 and Stxbp1, which are normally confined to the connecting cilium and are components of the SNARE complex required for vesicle trafficking. Suitably, Stx3 has previously been shown to be involved in the trafficking of rhodopsin (Chuang et al. 2007; Mazelova et al. 2009). Datta and colleagues hypothesised that a lack of functional proteins in the inner segment may contribute to apoptosis of the photoreceptors, and the BBSome functions to prevent entry of non-outer segment proteins to the outer segment (Datta et al. 2015).

Although a great deal of focus has been dedicated to researching the role of anterograde transport in photoreceptor degeneration, the process of retrograde transport has received less interest. It is generally accepted that, due to the large turnover of the photoreceptor through shedding vesicles at its distal tip, damaged proteins are removed and do not need shuttling back to the cell body (Insinna and Besharse 2008). Additionally, some proteins, such as transducin, require ultra-rapid movement from the outer segment to the inner segment, and can do so passively, without the need for IFT (Sokolov et al. 2002). Nevertheless, knock down of the retrograde IFT molecule dynein-2 in zebrafish led to truncated outer segments and swollen axonemes, induced by the accumulation of anterograde machinery at the ciliary tip (Krock et al.

2009). This study showed that retrograde transport is required for recycling of IFT machinery, but not for phototransduction proteins. However, the build-up of IFT proteins alone could be responsible for cell stress and ultimately apoptosis.

One aspect that is particularly reproducible in ciliopathy retinopathy models is the highly unstructured and disorganised optic discs. Perhaps this alone is the reason that these cells undergo apoptosis, as they have reduced functionality (Datta et al. 2015).

To further support the model that suggests apoptosis is triggered due to accumulation of proteins in the inner segment, Mockel et al. shows that the unfolded protein response (UPR) is upregulated in *Bbs12* depleted retina (Mockel et al. 2012). This is a mechanism that is initiated due to stress on the endoplasmic reticulum caused by protein misfolding, and aims to restore homeostasis to the cell. If cellular balance is not reinstated, apoptosis is activated. Indeed, inhibition of certain players in the UPR, such as a caspase12 inhibitor (INH) or BiP agonist valproic acid (VPA), significantly decreased the number of apoptotic nuclei in the outer nuclear layer. Interestingly, they also saw a swelling of the ER cisternae in the photoreceptors of the *Bbs12*$^{-/-}$ mouse, which they accredited to the accumulation of proteins in the inner segment (Mockel et al. 2012) . This strongly suggests a mechanism that accounts for apoptosis in BBS photoreceptors, and an interesting opportunity for therapeutic intervention, especially since VPA is already being used for retinal diseases such as ischemia-reperfusion injury (Zhang et al. 2012).

From a review of the literature, it is clear that understanding the mechanism of photoreceptor degeneration in BBS is not trivial, and there are many contradictory studies. To add an additional layer of complexity, although it is believed that the connecting cilium behaves similarly to a typical primary cilium, mislocalisation of proteins has not been recorded in other tissue types. This could be explained by a degree of tissue specificity, where BBS proteins behave differently in photoreceptors compared to other primary cilia.

6.1.5 Protocols for iPSC differentiation into optic cups

As previously described, the development of the mammalian eye is a complex process, which requires the coordination of multiple events to occur at precise time points. The idea of generating an eye *in vitro* therefore, may seem far-fetched and futuristic. In recent years however, there has been vast progress in being able to develop features of the eye from embryonic stem cells (ESCs), such as lens-like structures (Yang et al. 2010), retinal neurons, or

RPE cells (Meyer et al. 2009; Lamba et al. 2006; Osakada et al. 2008). The greatest breakthrough came from the Sasai lab, who used murine ESCs to recreate optic cup development, including invagination of the optic vesicle, expression of RPE and NR markers, and generation of all major neural retinal layers (Eiraku et al. 2011). Astoundingly, this morphogenesis occurred without the presence of a lens or surface ectodermal tissues, suggesting that the process is self-directed and influenced by internal transcription factors only (Eiraku et al. 2011). The following year, the same group published a similar study, showing that a self-directed optic cup can also form using human ESCs (Nakano et al. 2012).

Building on the foundation made by Sasai's lab, the next advancement was recapitulating the success of photoreceptor development from ESCs in iPSCs (Zhong et al. 2014; Wright et al. 2014; Meyer et al. 2011), which could later be utilised to model retinal diseases with patient material, such as Usher syndrome (USH) and Leber Congenital Amaurosis (LCA) (Tucker et al. 2013; Parfitt et al. 2016). In recent years, a wealth of publications have been released describing protocols with different modifications, such as the use of a variety of exogenous factors, two- or three-dimensional systems, or different stem cell culturing systems (Tucker et al. 2013; Zhong et al. 2014; Meyer et al. 2009; Reichman et al. 2014; Sridhar et al. 2013; Mellough et al. 2012; Lamba et al. 2006; Osakada et al. 2008; Gonzalez-Cordero et al. 2013; Boucherie et al. 2013). Despite the contrasting methods seen in these studies, they were all able to reach the same conclusion, where precursor photoreceptors were expressed.

An extra level of achievement was reached by the Canto-Soler lab, who demonstrated that their differentiation protocol was able to promote maturation of photoreceptors to produce intracellular membrane discs similar to outer segment discs (Zhong et al. 2014). Furthermore, they conducted perforated patch clamp experiments on putative rod cells to establish if there was evidence of functionality when stimulated by light. In total, two out of thirteen cells tested exhibited a response to the light stimulus. What is surprising about this particular accomplishment is that unlike other labs, this protocol did not add exogeneous agonists or antagonists to their culturing system. The phenomenon of retinal cup formation occurred spontaneously and only through secreted endogeneous factors in this native-like micro-environment (Zhong et al. 2014).

6.1.6 Aims of investigation

Murine systems are popular models for the study of rod-cone dystrophy in BBS, since they mimic the progressive disease presentation seen in patients. However, with the advent of iPSC technology, we are presented with a novel opportunity to create an *in vitro* human model system from iPSCs derived from BBS patient fibroblasts.

There are clear advantages to using a model developed from patient-derived iPSCs. For one, retinal tissue from BBS patients is inaccessible, unless obtained post-mortem, and thus the patient's degenerative progression cannot be investigated *in vivo*. Therefore, iPSC derived optic cups can be used to interrogate the pathobiology of rod-cone dystrophy, specifically investigating whether there is an underlying developmental defect or a light stimulus is required to induce degeneration. Furthermore, successful development of the retina would allow validation of therapeutic interventions, such as gene therapy correction or drug screening, in a patient specific model system.

The work in this chapter utilises patient iPSC cells derived by the HipSci project funded by the Wellcome Trust and the Medical Research Council. In total, skin biopsies from 50 BBS patients were accepted for reprogramming by the project. From the available reprogrammed lines, 3 BBS patients with retinal dystrophy and 2 clinically healthy control lines were chosen. The three patients in question were BBS001, BBS007 and BBS008, all of whom carry a homozygous p.M390R variant in *BBS1*.

The aim of this investigation was to develop a protocol for differentiation of these 5 iPSC lines into 3D optic cups. The procedure was based on the published methods by Zhong and colleagues (2014), which was chosen due to the demonstration of photoreceptor functionality. Additionally, Zhong's method was less complex than others available, such as Nakano et al. (2012), which was reported as technically challenging to achieve for large-scale cultures. The main feature of this study that differed from other published methods was the use of a iPSC culturing system comprised of vitronectin and E8 media, in contrast to Matrigel and mTESR.

The differentiation protocol was followed up to day 198 (28 weeks) until mature optic cups had developed. To determine whether differentiation followed the same course as *in vivo* development, key gene markers were assayed for expression at 8 time points throughout the protocol. The results of this chapter demonstrate a large amount of differentiation efficiency

between different cell lines. Two out of five cell lines failed to differentiate completely. As this affected both patient and control lines, this was not a problem specific to BBS. On the other hand, iPSCs derived from patient BBS001 differentiated well into mature optic cups. Immuno-histochemical staining exhibited a well organised structure of retinal cells, expression of phototransduction proteins, and correct localisation of connecting cilium proteins. The ultrastructure was investigated by transmission electron microscopy, which revealed presence of mitochondria rich inner segments, connecting cilia and nascent outer segments. Finally, within this chapter it is discussed whether this *in vitro* model could realistically be scalable, reproducible and utilised as a means of therapeutic testing.

6.2 Results

6.2.1 Maintenance and experimental design

6.2.1.1 Culturing of iPSC lines

iPSC lines were reprogrammed from patient fibroblasts by the HipSci laboratories at the Wellcome Trust Sanger Institute. Following the popularisation of stem cell culturing, there have been a myriad of culturing systems released for research use, such as different matrices and media. The lines obtained from HipSci were feeder-free and utilised the E8 media with vitronectin matrices. This culturing system was adopted and maintained throughout iPSC culturing. The key difference in differentiation protocol in this study compared to Zhong et al. was the use of vitronectin/E8 culturing system over a Matrigel matrix and mTESR medium.

The lines received were contamination free and checked for mycoplasma prior to differentiation, as well as passing HipSci quality control tests, including karyotyping. To minimise variability, it was ensured that each line had been passaged 3 times to expand the culture prior to commencing differentiation. The level of spontaneous differentiation was monitored to enhance differentiation efficiency. Each line was subjected to rigorous testing for pluripotency, such as assessment by PluriTest (Müller et al. 2011), at the HipSci laboratories.

6.2.1.2 Validation of disease causing mutation in iPSC lines

Before commencing differentiation of the BBS lines into 3D optic cups it was important to validate the presence of the primary driving mutation, *BBS1* p.M390R, in each line. The reason for this was twofold: first to ensure that the patient identity was correct, and second, to confirm that the mutation had not been lost in the reprogramming process, for example, through a deletion or chromosomal aberration. Sanger sequencing of genomic DNA extracted from iPSCs confirmed the presence of the reference genotype in control lines, CTRL037 and CTRL038. Similarly, sequencing affirmed the mutation in all three BBS patient lines, derived from patients BBS008, BBS001 and BBS007, where substitution of a T to a G resulted in an amino acid change from methionine to arginine (**figure 6.5**). Further details for each individual in the study are presented in **table 6.1**.

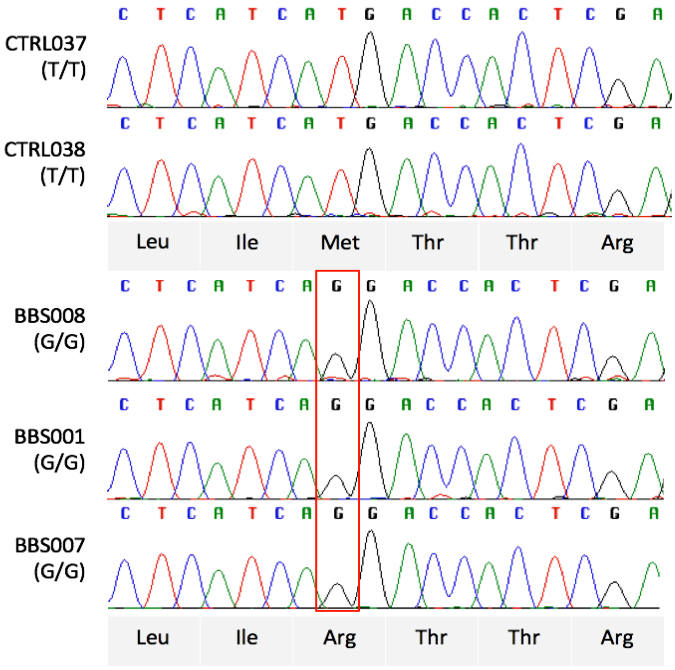

Figure 6.5 – Confirmation by Sanger sequencing of homozygous BBS1 p.M390R in iPSC lines derived from patients BBS001, BBS007, and BBS008. This variant was absent from control lines, CTRL037 and CTRL038.

	iPSC line ID	Donor age	Donor sex	Genotype allele 1	Genotype allele 2
Patient	BBS001	30-34	Female	*BBS1* p.M390R	*BBS1* p.M390R
Patient	BBS007	40-44	Male	*BBS1* p.M390R	*BBS1* p.M390R
Patient	BBS008	35-39	Male	*BBS1* p.M390R	*BBS1* p.M390R
Control	CTRL037	65-69	Male	Reference	Reference
Control	CTRL038	40-44	Female	Reference	Reference

Table 6.1 – Table presenting patient and control iPSC line information, where genotype, age, and sex of each individual included in the study is shown.

6.2.1.3 Identification of a suitable reference gene

Correct quantitative real-time PCR (qPCR) practice requires the use of at least one reference gene to normalize the target gene data to. This normalisation reduces the variability found through differences in mRNA extraction and reverse transcription performance, prior to qPCR setup. Housekeeping genes are genes that are required for typical cell survival, such as metabolic genes, and are commonly used as a reference since they are constitutively expressed. One of the most commonly used reference genes is *GAPDH*. However, there have been reports that *GAPDH* and other popular housekeeping genes can vary across different organisms, conditions, and importantly, cell types (Kozera and Rapacz 2013; Barber et al. 2005).

In this study, the aim was to assess the variability of gene expression over a period of *in vitro* development. During this period, a variety of ever-changing cell types will develop and mature. It was therefore imperative that a reliable housekeeping gene was established that remained consistently expressed over the differentiation time frame. In order to determine a suitable housekeeping gene, cDNA was obtained from mRNA extracted from the patient BBS001 line at day 0, 6, 16, 37, 60, 98, and 124 after initial differentiation. Six commonly used housekeeping genes were selected to test, namely *GAPDH, HPRT1, SNRPD3, LDHA, TBP,* and *PGK1*. Samples were normalised to day 0 within each gene set and the fold change plotted across the time course (figure 6.6).

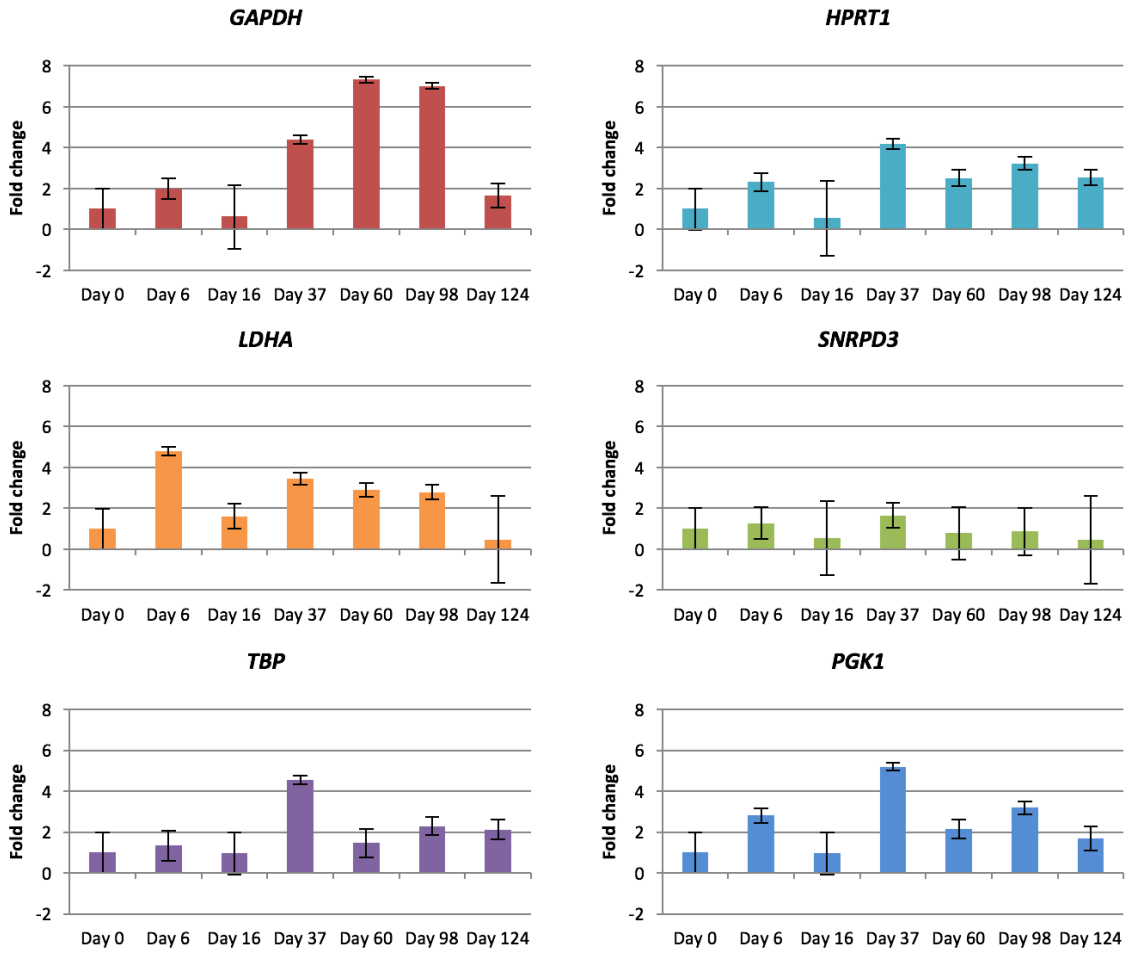

Figure 6.6 – Relative fold change of expression for 6 putative reference genes from optic cup samples taken over a 124-day differentiation process. iPSCs were derived from patient BBS001. Error bars show transformed 95% confidence intervals (CI) for each experiment, which was performed in triplicate.

The qPCR results showed that *GAPDH* exhibited the greatest variability, which demonstrated the greatest fold change difference from day 0 at day 60 (FC = 7.32; day 0 relative expression = 1, 95% CI= 0.01, 2.01; day 60 relative expression= 7.32, 95% CI= 7.18, 7.46). This supports studies

that show *GAPDH* is not consistently expressed across different tissue types (Barber et al. 2005). Other putative reference genes also showed variability over the time course, for example, the largest differences in fold change existed between day 0 and day 37 for *HPRT1* (FC = 4.18; day 0 relative expression = 1, 95% CI= 0.02, 2.02; day 37 relative expression= 4.18, 95% CI= 3.94, 4.42), day 6 for *LDHA* (FC = 4.79; day 0 relative expression= 1, 95% CI= 0.02, 2.02; day 6 relative expression = 4.79, 95% CI= 4.58, 5.00), day 37 for *TBP* (FC = 4.55; day 0 relative expression= 1, 95% CI= 0.01, 2.01; day 37 relative expression = 4.55, 95% CI= 4.34, 4.78), and day 37 for *PGK1* (FC = 5.19; day 0 relative expression = 1, 95% CI= 0.01, 2.01; day 37 relative expression= 5.19, 95% CI= 5.00, 5.39). Additionally, *TBP* was an unsuitable reference gene since the C_T values were relatively low, ranging between 27 and 30. The most consistently expressed gene was *SNRPD3*, which showed a maximum increase at day 37 (FC = 1.90; day 0 relative expression= 1, 95% CI= 0.00, 2.00; day 37 expression = 1.90, 95% CI= 1.03, 2.26). As *SNRPD3* showed the most consistent expression over the time course, this was utilised as the reference gene for all following qPCR experiments.

6.2.2 Differentiation of iPSCs into 3D optic cups

As described above, differentiations were undertaken using 5 independent iPSC lines. However, due to variables outlined here and in detail in the discussion, there was limited success in reaching the terminal differentiation point for many of these cell lines (day 198/ week 28). The best quality differentiations followed the correct gene expression patterns and morphology at a suitable frequency so that samples could be taken for analysis throughout. One particular line, from patient BBS001, differentiated accordingly, mimicking gene expression of both *in vivo* development and other published protocols (Meyer et al. 2009; Zhong et al. 2014; Parfitt et al. 2016; Hirami et al. 2009). The results described first will follow the differentiation this line, BBS001. Differentiations of other lines were then compared to BBS001, the findings of which are outlined later.

6.2.2.1 Embryoid body formation

The first stage of the differentiation protocol required the removal of anti-differentiation factors from the culturing media. The media was changed from E8 media to embryoid body media (EBM), which was supplemented with ROCK inhibitor on the first day of differentiation (day 0) to prevent dissociation induced apoptosis. This change in conditions resulted in the formation of embryoid bodies (EBs), three-dimensional non-adherent cell aggregates. These aggregates epitomise the molecular and cellular morphogenic interactions implemented during development, and can also

differentiate to give rise to the three germ layers (Rungarunlert et al. 2009). Thus, this was a common method for inducing differentiation *in vitro* for both iPSC and ESC systems. Images of iPSC colonies and resulting EBs from patient BBS001 can be found in the appendix, **figure A1 A-B**.

Over the first days of EB development it was expected that the expression of pluripotency markers would decrease and developmental transcription factors would be upregulated (Meyer et al. 2009). qPCR analysis of *NANOG* expression was conducted from material extracted from day 0, day 6, and day 16 of differentiation. Over this time period, *NANOG* expression decreased, suggesting that the differentiation procedure had successfully commenced. RT-PCR analysis for *NANOG* complemented qPCR data, showing that expression had diminished by day 6. On the other hand, *OCT4* expression decreased more gradually and residual expression was still visible by day 37 (**figure 6.7**).

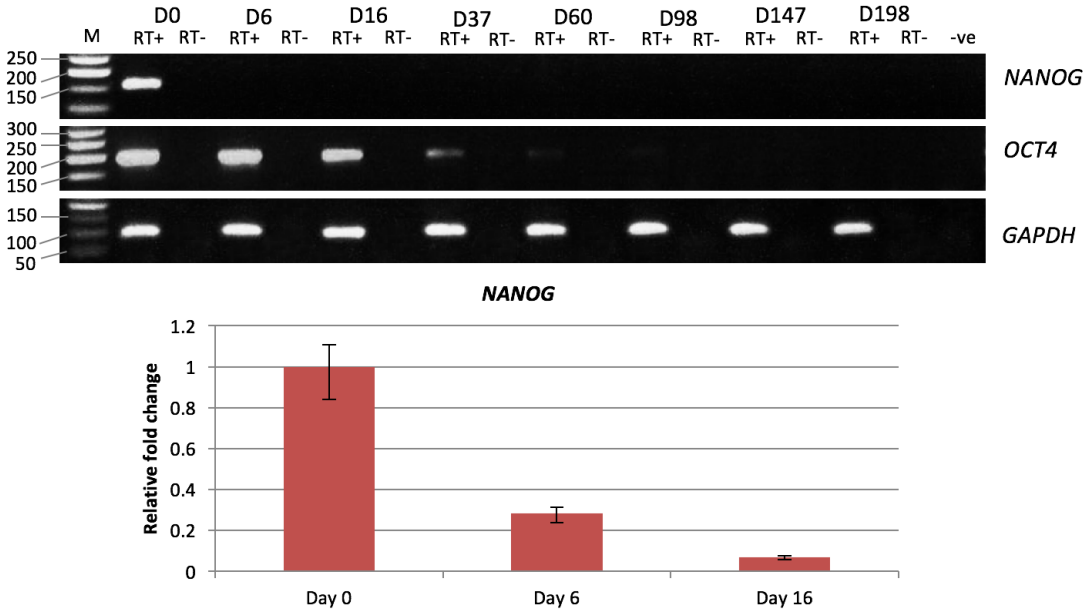

Figure 6.7 – Gene expression of pluripotency genes from patient BBS001 samples. RT-PCR expression (top) of *NANOG* and *OCT4* over the differentiation process from day 0 to day 198. *GAPDH* was used as a positive control. Relative fold change was determined by qPCR (bottom) for *NANOG* over day 0, 6 and 16 of differentiation. Samples were normalised to the reference gene *SNRPD3* and day 0 for each experiment. Error bars represent transformed 95% CI calculated from triplicate experiments. Relative fold change was calculated through calculation of $2^{(-\Delta\Delta CT)}$.

On day 4 of the differentiation protocol, the media was switched from EBM to neural induction media (NIM), which contained N2 supplementation. This media was designed to encourage differentiation towards a neuroepithelial specification, which was reflected by expression of

EFTFs. On day 6, the matured EBs were transferred to laminin-coated culture plates, which allowed the colonies to adhere and promoted the development of eye field domains.

Through day 6 to day 37, as expected, *PAX6* and *RX* showed a close relationship by imitating each other's expression patterns. Together these genes are responsible for the formation of the eye field during gastrulation (Zuber et al. 2003). RT-PCR analysis of both *PAX6* and *RX* showed that expression had commenced by day 6 (**figure 6.8**). *PAX6* is crucial for the development of the eye, which is demonstrated by an eyeless phenotype caused by deletion of the *Pax6* homologue in *D. melanogaster* (Quiring et al. 1994). RT-PCR experiments verify the presence of both *PAX6(-5a)* and *PAX6(+5a)* isoforms, which is evident by the two bands on the agarose gel. *RX* is involved in upregulation of other EFTFs including *LHX2*, which was coregulated with *RX* and *PAX6* during EB maturation, and was strongly expressed by day 6. Both *RX* and *LHX2* are involved in regionalisation of the optic vesicle, and lack of their expression results in failure to define between the neural retina and the RPE in the native retina (Yun et al. 2009; Medina-Martinez et al. 2009).

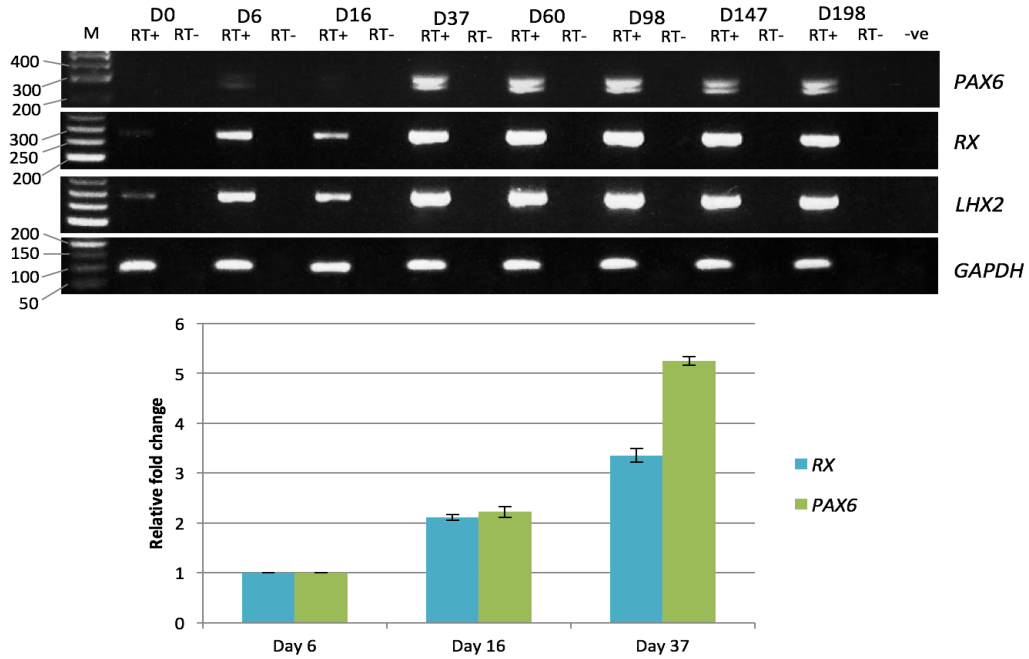

Figure 6.8 – Gene expression analysis of EFTF genes from patient BBS001 samples. RT-PCR expression (top) of *PAX6, RX* and *LHX2* over the differentiation process from day 0 to day 198. *GAPDH* was used as a positive control. Relative fold change was determined by qPCR (bottom) for *RX* and *PAX6* over day 6, 16, and 37 of differentiation. Samples were normalised to the reference gene *SNRPD3* and day 6 for each experiment. Error bars represent transformed 95% CI calculated from triplicate experiments. Relative fold change was calculated through calculation of $2^{(-\Delta\Delta CT)}$.

6.2.2.2 Formation of neural retina and retinal pigmented epithelium

On day 16 of the differentiation protocol, the media was switched to retinal differentiation media (RDM) containing B27 supplementation. B27 has been shown to promote neural progenitor differentiation and facilitate development of phototransduction machinery (Mellough et al. 2012). Switching to RDM encouraged differentiation towards a retinal progenitor fate and initiated expression of markers that distinguish the optic vesicle.

Neural retinal (NR) domains in BBS001 patient cultures were identifiable by their distinctive horseshoe shaped lamination that is evocative of the optic cup *in vivo* (figure 6.9 A-C). On day 28 of culture, these structured NR domains were manually detached and transferred to non-coated cell culture plates (figure 6.9D), where they matured into non-adherent three-dimensional optic cups. Once in suspension, the NR thickened over time and RPE gradually pigmented (figure 6.9 E-J). Contrary to *in vivo* development, the RPE was isolated from the neural retina, either packaged at the tip of the cup (figure 6.9 F-G) or in the middle (figure 6.9 I-J). By day 70 of differentiation, these cups had formed a well-defined self-laminated NR structure and approximately 80% were expressing RPE.

To facilitate maintenance and maturation of the optic cup, the media was supplemented with FBS and taurine at day 42. FBS had been shown to enhance retinal differentiation (Nakano et al. 2012), but also prevented necrosis at the centre of the cup (Zhong et al. 2014). Retinoic acid at 1µM was added to the media at day 70, and then reduced to 0.5µM at day 98. Zhong et al. had previously shown that intervention with retinoic acid significantly enhanced rhodopsin expression by week 21 (Zhong et al. 2014). These conditions were able to support the thickening of the NR up until the experiment end point, day 198.

For specification and subsequent cell identity of the NR, two antagonistically expressed regulatory genes are pivotal, *MITF* and *VSX2*. During development of the mammalian eye, the surface ectoderm controls expression of *MITF* and *VSX2* through *FGF1* and *FGF2* signalling (Horsford et al. 2005). In the native NR, *VSX2* represses *MITF* expression and restricts it to the RPE. Therefore, these signalling pathways lead to the compartmentalisation of the NR where *VSX2* is prominent, and the RPE, which is characterised by *MITF* expression (Horsford et al. 2005).

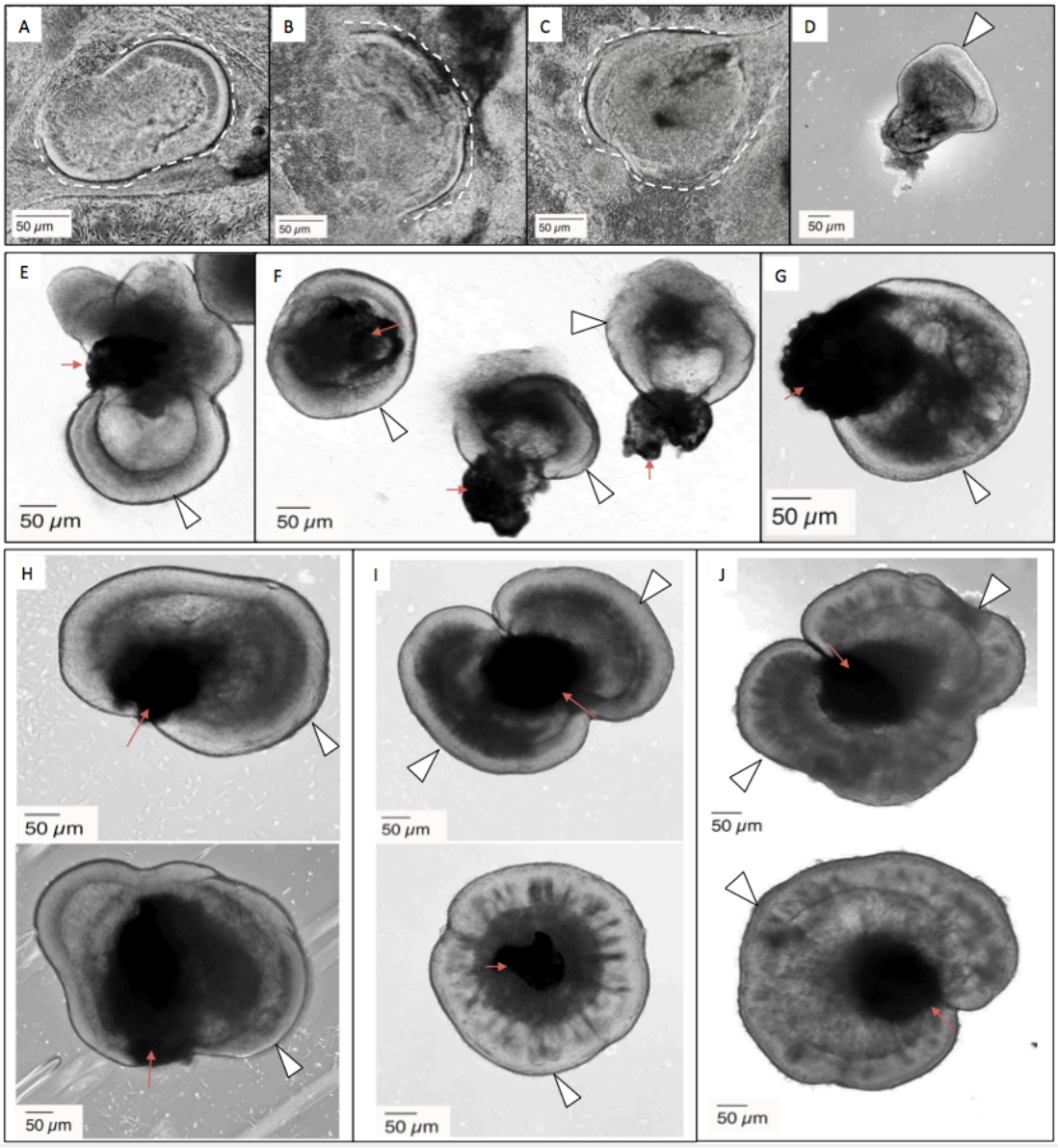

Figure 6.9 –The progression of differentiating optic cups from patient BBS001 samples at day 28 through to day 198. A-D) Cultures at day 28 show distinctive horseshoe shaped NR (dashed line, arrow heads), which was manually lifted for culture in suspension (D). E-G) At day 70 the NR (arrow heads) had thickened and RPE (red arrow) had formed. H-J) Optic cups in suspension show thickening of the NR over time where H) day 98, I) day 147 and J) day 198.

Concordantly, it was important to assess the presence of these two genes during the development of the neural retina. RT-PCR analysis showed *VSX2* and *MITF* expression was activated as early as day 6 of the differentiation process (**figure 6.10**). qPCR experiments demonstrated a 4.42-fold increase in *MITF* between day 6 to day 16 (day 6 mean ΔC_T= 4.39, 95% CI= 4.22, 4.56; day 16 mean ΔC_T= 2.25, 95% CI= 2.09, 2.39; p= 0.0012), which was maintained at day 37 (day 37 mean ΔC_T= 2.34, 95% CI= 2.19, 2.48). By day 60, *MITF* expression began to rise

again, which coincided with RPE development, suggesting an RPE specific expression of *MITF* (day 60 mean ΔC_T= 1.69, 95% CI= 1.58, 1.81). *VSX2* expression increased 288.27-fold from day 6 to day 60 (day 6 ΔC_T= 4.47, 95% CI= 4.23, 4.72; day 60 mean ΔC_T= -3.70, 95% CI= -4.05, -3.35; p<0.0001), which perhaps quenched *MITF* expression in the neural retina. Both genes were maintained through to the end point of differentiation, as shown by RT-PCR analysis (figure 6.10).

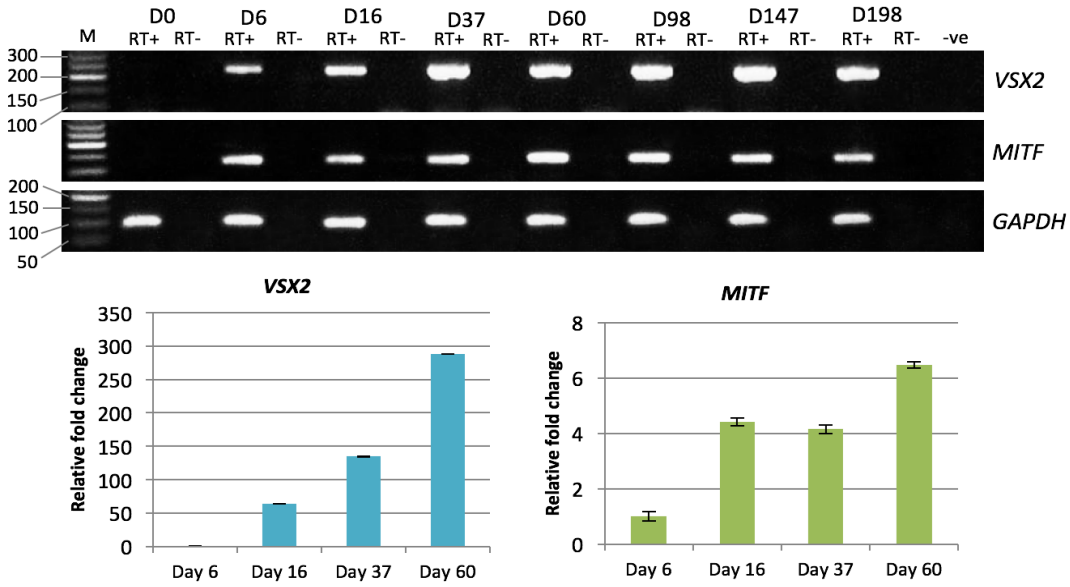

Figure 6.10 – Gene expression of neural retina specification markers from patient BBS001 samples. RT-PCR expression (top) of *VSX2* and *MITF* over the differentiation process from day 0 to day 198. *GAPDH* was used as a positive control. Relative fold change was determined by qPCR (bottom) for *VSX2* and *MITF* over day 6, 16, 37 and 60 of differentiation. Samples were normalised to the reference gene *SNRPD3* and day 6 for each experiment. Error bars represent transformed 95% CI calculated from triplicate experiments. Relative fold change was calculated through calculation of $2^{(-\Delta\Delta CT)}$.

6.2.2.3 Expression of photoreceptor specific cells

Photoreceptor precursors were expected to develop in the apical layer of the optic cups (Zhong et al. 2014). The putative expression and localisation of photoreceptor precursor cells was investigated through analysis of two photoreceptor specific genes, *CRX* and *RCVRN*. *CRX* is a transcription factor specific to rod and cone photoreceptor precursors and is key for the prospective expression of phototransduction proteins. Failure to express *CRX* in mice *in vivo* leads to lack of outer segments and subsequent photoreceptor degeneration (Furukawa et al. 1999; Swaroop et al. 2010). *RCVRN* encodes a phototransduction protein, Recoverin, that precedes expression of opsins in development of the fetal retina (Hendrickson et al. 2008). This pattern was mimicked in the optic vesicles studied here, as *RCVRN* expression was observed in the

cytoplasm and the developing outer segment of both rod and cone photoreceptors (**Figure 6.12 A-B**).

RT-PCR experiments indicated that *RCVRN* expression was lowly expressed at day 6-16, which was earlier than has been reported in similar studies (**Figure 6.11**). Previously, the earliest account of *RCVRN* expression was day 60 (Tucker et al. 2013). qPCR analysis demonstrated that *RCVRN* expression increased from day 37 through day 98 by 19.72-fold (day 37 mean ΔC_T= 2.73, 95% CI= 2.61, 2.86; day 98 ΔC_T= -1.57, 95% CI= -1.58, -1.55; p<0.0001), then significantly heightened during maturation to day 147 to a fold change increase of 139.96 compared to day 37 (day 147 ΔC_T= -4.39, 95% CI = -4.40, -4.39; p<0.0001). *CRX* expression showed a similar upregulation during the same time frame (relative fold change (RFC)= 7.95; day 37 mean ΔC_T= 2.13, 95% CI= 2.04, 2.23; day 147 mean ΔC_T= -0.86, 95% CI= -0.93, -0.79; p<0.0001), though qPCR analysis indicated that the difference between day 37 and day 147 was less pronounced than for *RCVRN*. RT-PCR analysis suggested a later onset of expression at day 98 and a down-regulation prior to the end of differentiation at day 198.

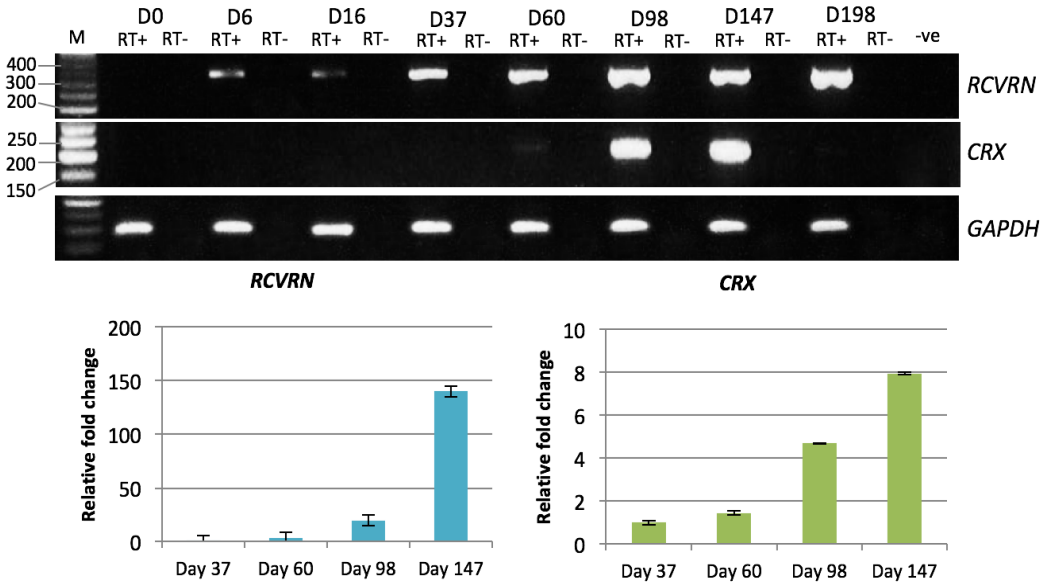

Figure 6.11 – Gene expression analysis of phototransduction markers from patient BBS001 samples. RT-PCR expression (top) of *RCVRN* and *CRX* over the differentiation process from day 0 to day 198. *GAPDH* was used as a positive control. Relative fold change was determined by qPCR (bottom) for *RCVRN* and *CRX* over day 37, 60, 98, and 147 of differentiation. Samples were normalised to the reference gene *SNRPD3* and day 37 for each experiment. Error bars represent transformed 95% CI calculated from triplicate experiments. Relative fold change was calculated through calculation of $2^{(-\Delta\Delta CT)}$.

The expression of phototransduction proteins in patient BBS001 can be compared against a control line that was previously published by Parfitt et al. (2016) (figure 6.12 A-B). At week 21 (day 147), the control photoreceptors differentiated by Parfitt and colleagues was abundantly expressing recoverin, cone arrestin and rhodopsin in the apical cell layer of the optic cup.

Within patient BBS001, immunohistochemistry specific for recoverin at day 98 mimicked expression seen for Parfitt et al. Furthermore, CRX was also expressed in cells expressing recoverin, demonstrating that the majority of CRX- and RCVRN-positive cells were concentrated in the outer nuclear layer (Figure 6.12 C-D). By day 198, limited expression of phototransduction proteins, rhodopsin and cone arrestin was detected (figure 6.12 E-F). In comparison to the number of recoverin expressing cells, it was surprising that so few cells matured to exhibit phototransduction proteins. This frequency of cells was also less than seen at the same time point in other protocols (Parfitt et al. 2016; Zhong et al. 2014) (figure 6.12 A-B). However, the majority of rhodopsin- and cone arrestin-positive cells resided in the inner nuclear layer. Therefore, it may be possible that these cells had not yet reached maturity and were yet to migrate to the outer nuclear layer in these cups.

Taken together, this demonstration showed that it was possible to differentiate a BBS patient-derived iPSC line into a model of a 3D optic cup. This was achievable in a temporal manner similar to that of *in vivo* retinal development and closely recapitulated the protocol that was adapted from Zhong and colleagues (2014).

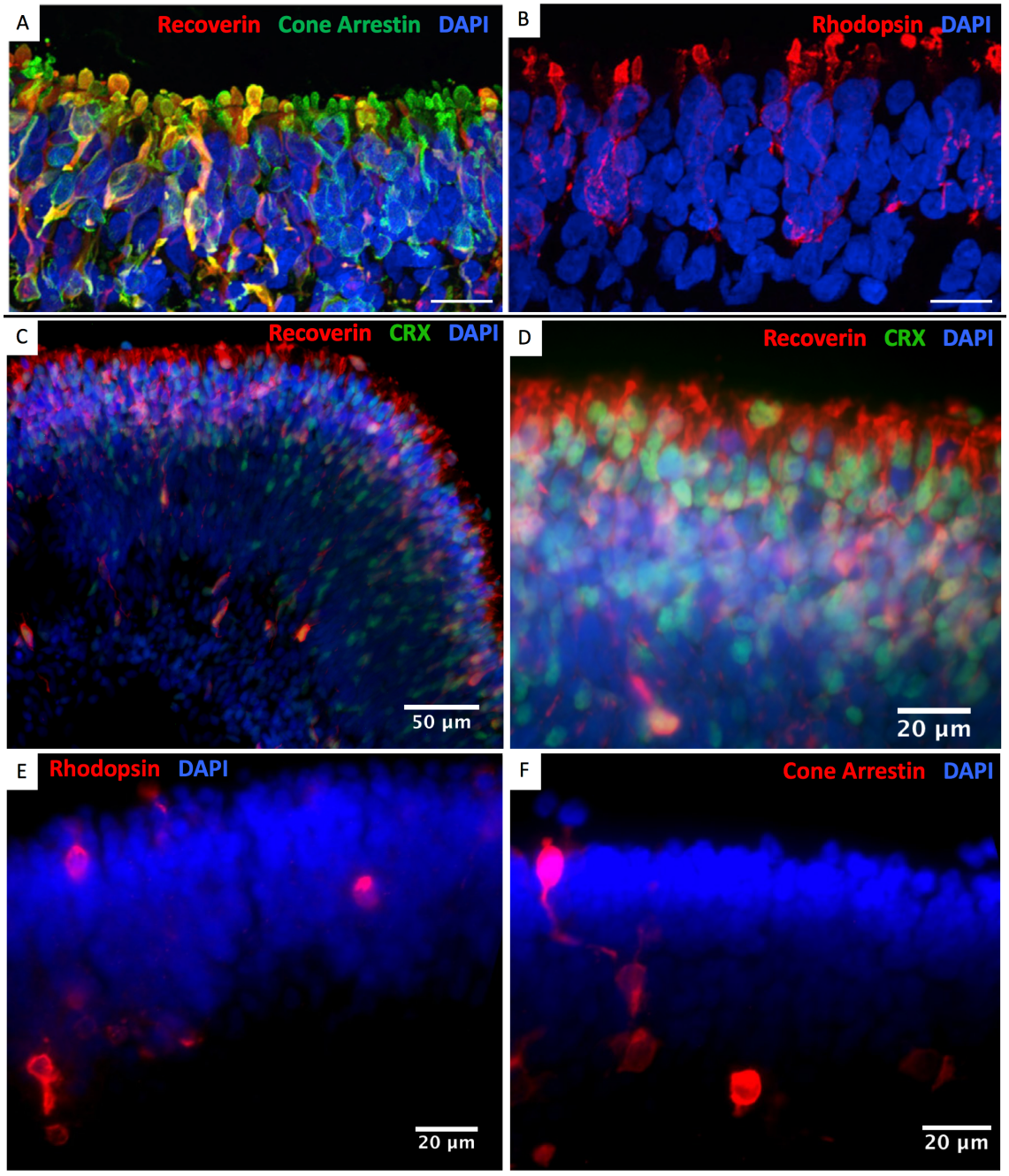

Figure 6.12 – Immunohistochemical analysis of phototransduction proteins in optic cups from patient BBS001, with comparative control images taken from Parfitt et al. (2016). A-B) Optic cups differentiated from control human iPSCs generated by Parfitt et al. at week 21 (~day 147) demonstrating expression of recoverin and cone arrestin (A) and rhodopsin (B) in the apical layer. Scale bar = 20µm. C-D) Expression of recoverin and CRX at day 98. E) Some rhodopsin positive cells at day 198. F) Expression of cone arrestin positive cells at day 198. Nuclei were stained with DAPI.

6.2.3 Differentiation of iPSCs from BBS008 and CTRL037

The aim of this study was to determine whether it was possible to differentiate BBS lines into 3D optic cups and assess the practicality of developing this protocol as an *in vitro* model to study

disease pathobiology. Although it has been demonstrated that the former was an achievable goal, there were some issues experienced during this project that will contribute towards the latter goal.

Consistent with previous reports (Meyer et al. 2009; Yu et al. 2007; Zhong et al. 2014), there was considerable variability in efficiency for differentiation across the five different lines. Indeed, even differentiations from the same individual showed variability, where some cultures had defined morphology, yet others failed to generate viable embryoid bodies or neural retinas. Through RT-PCR analysis to validate the presence of key genes throughout attempted differentiations, only 35% of efforts showed inclusive expression of *PAX6, RX, VSX2, MITF, CRX,* and *RCVRN* (n=26). Unfortunately, the control line CTRL038 and patient line from BBS007 failed to differentiate into any operable cultures.

The level of *PAX6* expression has been correlated with the differentiation efficiency of hiPSCs into the neural retina (Meyer et al. 2009). Therefore, it was hypothesised that there was a link with expression levels of *PAX6*, *NANOG,* and *RX,* and differentiation efficiency. Samples were chosen from CTRL037 and BBS008 that differentiated with the most successful morphology and investigated further by qPCR. Since BBS001 followed the correct morphology and temporal gene expression, these samples were used to compare against other differentiations (**figure 6.13**). Images of iPSC colonies and resulting EBs from patient BBS008 and control line CTRL037 can be found in the appendix, **figure A1 C-F**.

During the first 16 days of differentiation, both BBS008 and CTRL037 exhibited downregulation of *NANOG* (**figure 6.13A**). Compared to BBS001 and BBS008, CTRL037 had significantly higher levels of expression of *NANOG* at day 0 (CTRL037 ΔC_T mean= 2.02, 95% CI= 1.67, 2.36; BBS001 ΔC_T mean= 2.92, 95% CI= 2.64, 3.21; CTRL037 vs BBS001 p= 0.0162; BBS007 ΔC_T mean= 2.98, 95% CI= 2.70, 3.25; CTRL037 vs BBS008 p= 0.0129). In CTRL037, NANOG expression decreased as expected over the first 16 days, however, it remained significantly higher than BBS001 and BBS008 over this period (day 16: CTRL037 ΔC_T mean= 4.21, 95% CI= 3.87, 4.57; BBS001 ΔC_T mean= 6.75, 95% CI= 6.28, 7.24; CTRL037 vs BBS001 p= 0.0011; BBS008 ΔC_T mean= 5.41, 95% CI= 5.22, 5.60; CTRL037 vs BBS008 p= 0.0043). It is possible that this may be linked to the poor differentiation efficiency seen in CTRL037 cultures.

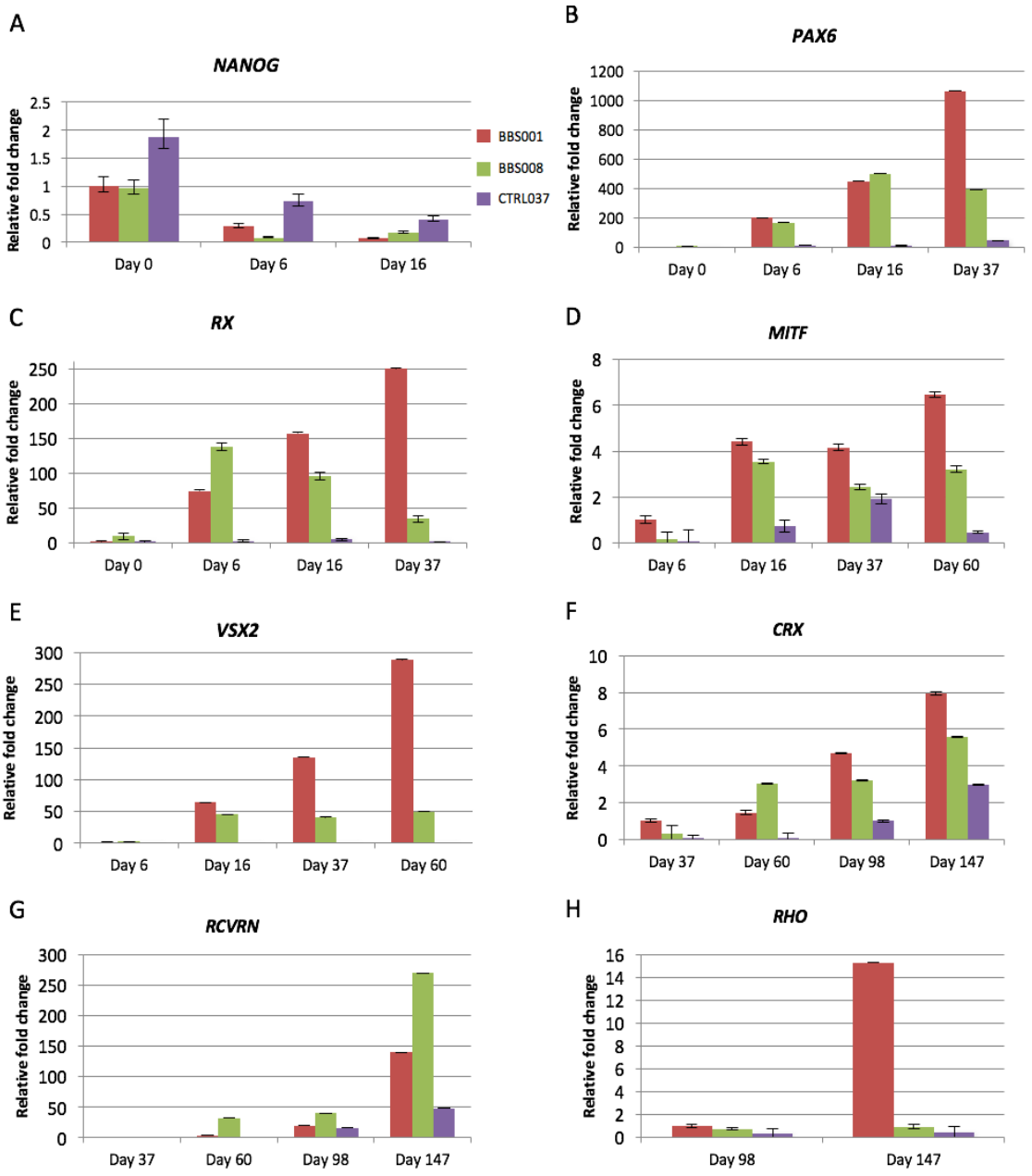

Figure 6.13 A-H – Relative fold change of gene expression profiles by qPCR for key genes involved in the development of the retina. Expression for BBS008 (green) and CTRL037 (purple) were normalised to BBS001 (red), allowing comparison across the differentiation process. All experiments were primarily normalised to the reference gene *SNRPD3*. Error bars represent transformed 95% CI calculated from triplicate experiments. Relative fold change was calculated through calculation of $2^{(-\Delta\Delta CT)}$.

RX and *PAX6* expression for BBS008 followed the same pattern of increased expression as BBS001 from day 0 to day 6 (*RX*: RFC= 137.6; *PAX6*: RFC= 166.96) (**figure 6.13 B-C**), but decreased as differentiation continued through to day 37, effecting *RX* in particular (*RX*: RFC= 34.1; *PAX6*: RFC=393.10). Though there was reduced expression of *RX* and *PAX6* compared to BBS001, BBS008 still exhibited expression of other key markers by qPCR analysis, such as *MITF* (**figure**

6.13D), *VSX2* (figure 6.13E), *CRX* (figure 6.13F), and *RCVRN* (figure 6.13G). Within CTRL037 cultures, *PAX6* did not reach the same high expression levels as BBS001 or BBS008 (day 37: CTRL037 ΔC_T mean= 1.13, 95% CI= 1.00, 1.26; BBS001 ΔC_T mean= -3.35, 95% CI= -3.45, -3.25; BBS008 ΔC_T mean= -1.91, 95% CI= -2.02, -1.80), and *RX* was very minimally expressed with RFC of 0.94, compared to 249.70 and 34.17 for BBS001 and BBS008, respectively (day 37: CTRL037 ΔC_T mean= 10.03, 95% CI= 9.08, 10.98; BBS001 ΔC_T mean= 1.99, 95% CI= 0.75, 3.22; BBS008 ΔC_T mean= 4.85, 95% CI= 3.30, 6.41) (figure 6.13 B-C). These results would fit with personal observations that the least successful line for differentiation was CTRL037, which concordantly had the lowest *PAX6* and *RX* expression levels over the first 37 days of differentiation. Furthermore, CTRL037 also exhibited low *VSX2* expression compared to BBS001 and BBS008 (day 60: CTRL037 RFC= 0.11, ΔC_T mean= 7.62, 95% CI= 7.29, 7.96; BBS001 RFC= 288.27, ΔC_T mean=-3.70, 95% CI= -4.04, -3.35; BBS008 RFC= 49.71, ΔC_T mean= -1.16, 95% CI= -1.63, -0.69). Despite this, there was expression of *CRX* and *RCVRN* from day 98 onwards (figure 6.13 F-G; figure 6.14). This may be due to collection of optic cups with more successful morphology at later time points, and is a reflection on how variable this line was throughout the differentiation process.

The viable CTRL037 and BBS008 differentiations were kept in culture until day 147 and 160, respectively. The optic cups with a morphology featuring defined NR and RPE were retained for immunohistochemical staining. Both CTRL037 and BBS008 expressed *CRX* and *RCVRN* at day 98 and day 147, respectively (figure 6.13 F-G). Overall, CTRL037 optic cups had a less organised structure (figure 6.14 D-G) and showed little lamination compared with BBS001. Although *CRX* and *RCVRN* were detected in the apical layer of the cup, expression was sparse and disjointed in places (Figure 6.14 D-F). Conversely, BBS008 showed a well-ordered outer nuclear layer with consistent *CRX* and *RCVRN* expression throughout the optic cup (Figure 6.14 A-B). In the outer nuclear layer, cone arrestin was also distinguishable by day 98 in CTRL037 (Figure 6.14G) and abundant by day 160 in BBS008 (Figure 6.14C). On the other hand, rhodopsin was absent in both cultures. qPCR analysis validated minimal expression of *RHO* in CTRL037 and BBS008 compared to BBS001 from day 98 to day 147 (BBS001 RFC= 15.32, ΔC_T mean= 4.20, 95% CI= 4.16, 4.23; BBS008 RFC= 0.94, ΔC_T mean= 8.23, 95% CI= 8.06, 8.39; CTRL037 RFC= 0.40, ΔC_T mean= 9.74, 95% CI= 8.89, 10.05) (figure 6.13H). Overall, there was a greater level of variability between cups derived from CTRL037 and BBS008 compared with BBS001, which resulted in fewer optic cups for experiments.

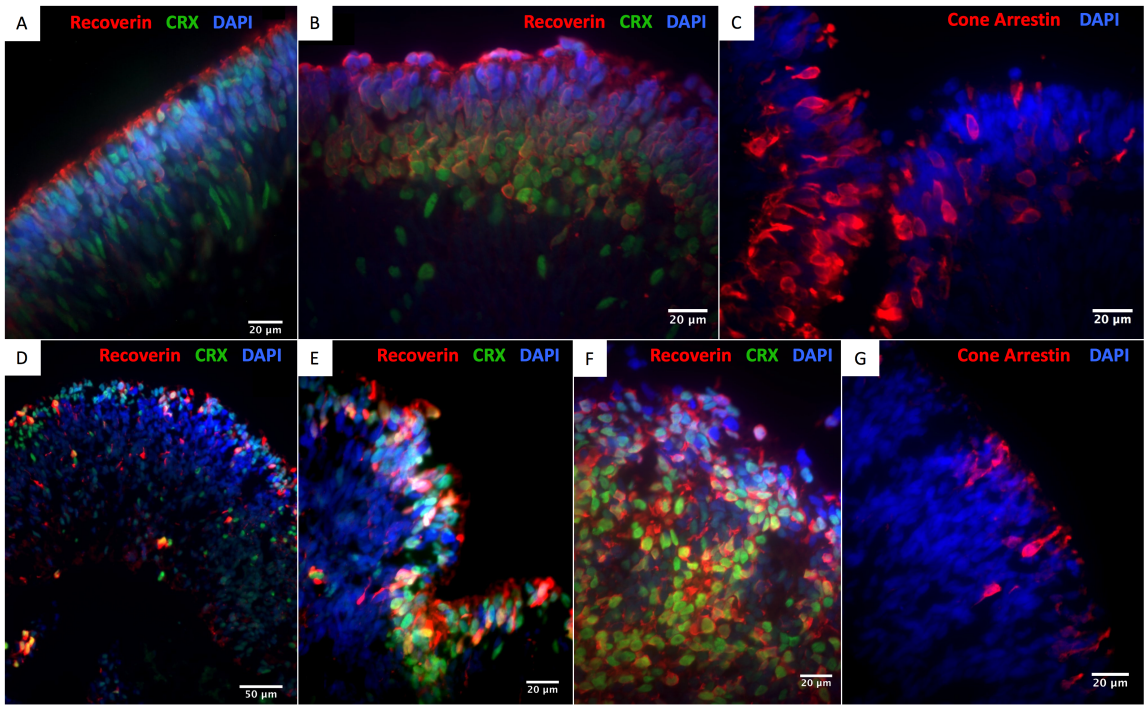

Figure 6.14 – Immunohistochemical analysis of phototransduction proteins in optic cups from CTRL037 and BBS008. A-B) Expression of recoverin and CRX in the apical layer of line derived from patient BBS008 at day 147 (A) and at day 160 (B). C) Expression of cone arrestin from BBS008 at day 160. D-F) Recoverin and CRX positive cells in the control line CTRL037 at day 98. G) Some cone arrestin positive cells in CTRL037 at day 98. Nuclei were stained with DAPI.

6.2.4 BBS-specific staining

To my knowledge, this is the first account of using BBS patient-derived iPSCs to obtain optic cups *in vitro*. Therefore, the next aim was to ascertain whether there was a BBS-specific phenotype observed in these photoreceptors, which may provide an insight into the disease pathobiology. These experiments were conducted in BBS001 derived optic cups collected at day 198 of differentiation.

Immunohistochemical staining revealed that BBS1 was exclusively expressed on the apical surface of optic cups and appeared to localise to the basal body of the developing photoreceptors (**figure 6.15 A-B**). The location of the basal body was verified by immunohistochemical staining for BBS5, another constituent member of the BBSome complex (**figure 6.15C**). Using an antibody for acetylated tubulin, some connecting cilia extending from the basal body were identifiable (**figure 6.15 B-D**). Taken together, this suggests that the p.M390R missense variant in *BBS1* does

not prevent expression or localisation of the BBSome, nor does it prevent connecting cilia from forming.

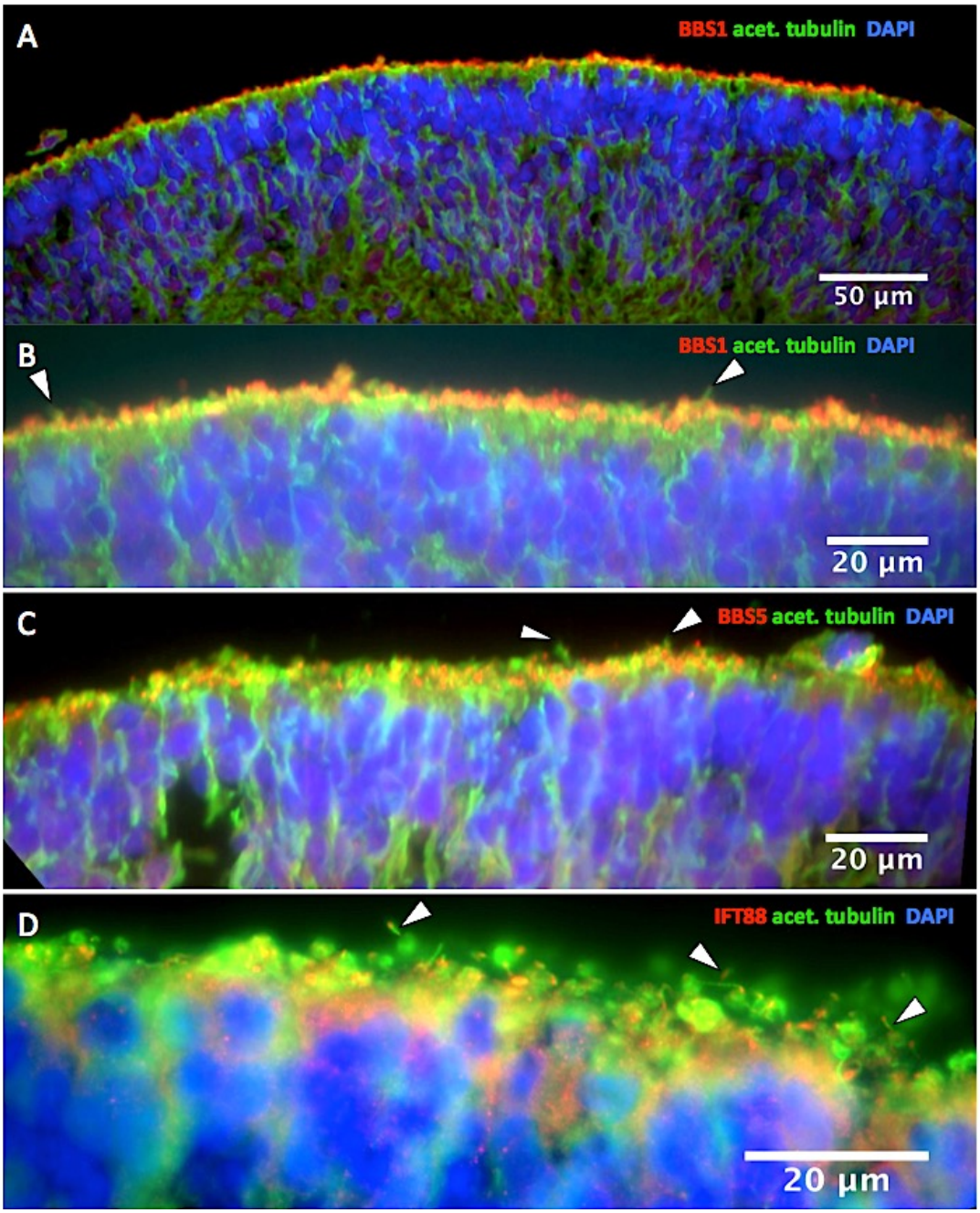

Figure 6.15 – Immunohistochemical staining to observe localisation of primary cilia proteins BBS1 (A-B), BBS5 (C), and IFT88 (D). Assays were completed in day 198 optic vesicles from patient BBS001. Acetylated tubulin was used as a marker for connecting cilia, which project from the apical surface of the optic cups (white triangles). Nuclei were stained with DAPI.

Finally, it was investigated whether there was functional intraflagellar transport and ciliogenesis occurring within the patient optic cup. IFT88 is a member of the IFT-B particle, which is vital for anterograde transport to the ciliary tip. Immunohistochemical staining showed that IFT88

localised to the connecting cilium, which was identifiable by staining of acetylated tubulin (figure 6.15D). There were some IFT-B particles present at the ciliary tip, suggesting that there was active ciliogenesis taking place. These results indicate that the defect within the BBSome does not affect ciliogenesis during photoreceptor development. This suggests that the p.M390R variant in *BBS1* induces photoreceptor degeneration after development of the photoreceptor rather than causing an early-developmental defect.

6.2.5 Transmission Electron microscopy

Transmission electron microscopy (TEM) is a powerful tool to investigate the ultrastructural features of a cell in exact detail. This method was utilised to determine an accurate picture of photoreceptor development and observe outer segment formation within BBS001 optic cups at day 198. It was also an opportunity for identification of any abnormalities that could be a result of the disease genotype. TEM processing and imaging was conducted by Dr Peter Munro from the UCL Institute of Ophthalmology.

TEM images of the apical-most surface of these vesicles showed inner segments of photoreceptors, identifiable due to their ellipsoid shape and abundant mitochondria (figure 6.16 A-H). The presence of tight junctions and an outer limiting membrane demonstrated polarisation and organisation of this cell layer (figure 6.16A, F, H). Inner segments that were sectioned in the corresponding plane of view showed an intact emerging connecting cilium (figure 6.16 B-E). The lengths of these cilia varied, possibly due to cells maturing at different rates. Some images displayed horizontal cross-sections of the basal body or centrosome, revealing their distinctive microtubule arrangement (9x3 + 0 or 9x2 + 0, respectively), at the apex of the cell (figure 6.16A, F-H). Vertical cross-sections revealed the complex structure of the developing connecting cilium, including the axoneme that extended from the basal body into a membranous vesicle at the distal tip. This club-shaped vesicle has become a characteristic of the developing outer segment in iPSC derived optic cups (Parfitt et al. 2016; Nakano et al. 2012). In fact, the BBS optic cups showed presence of disorganised material in this outer segment, which could be attributed to disordered, developing outer segment discs, similar to that seen by Parfitt et al (2016). This is particularly prominent in figure 6.16D.

As described earlier, there were difficulties in reaching the end point of well-developed optic cups for all lines. As a consequence of this, there were no control optic cups available for comparison TEM experiments. However, the results I have presented here are substantiated by control optic

cups in the literature, which present indistinguishable structural features (Zhong et al. 2014; Parfitt et al. 2016). Therefore, at this stage in development there are no obvious developmental defects seen in this BBS patient's optic cups.

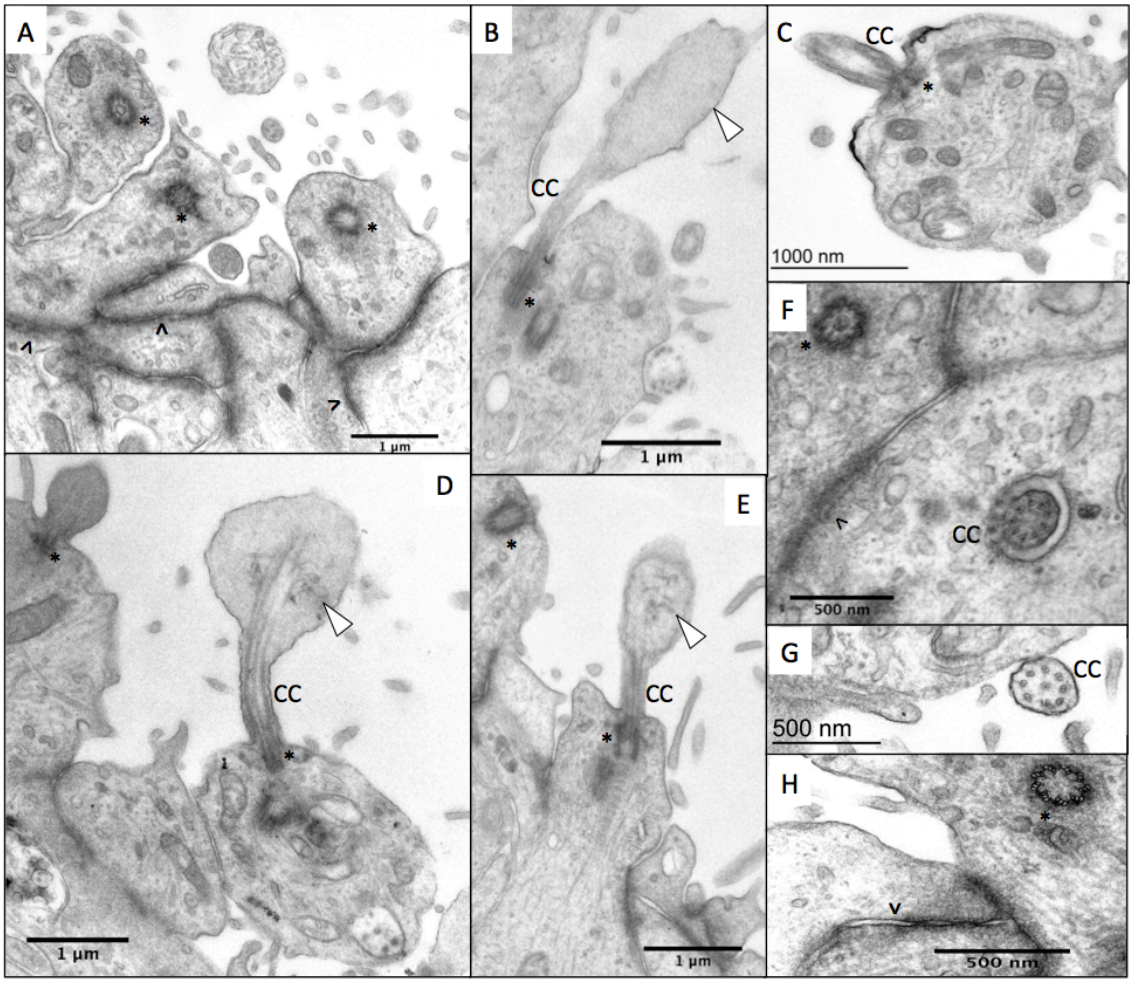

Figure 6.16 – Transmission electron microscopic analysis of optic cups at day 198 from patient BBS001. Images were taken from the apical layer and display mitochondria rich inner segments (A, C, D, F), basal bodies (*), tight junctions (^), connecting cilia (CC) and club-like nascent outer segments (arrow heads). Basal bodies (*) are distinguishable from connecting cilia (CC) due to their microtubule arrangement of 9x3+0 (H), compared to 9x2+0 (G). TEM and imaging was completed by Dr Peter Munro at the UCL Institute of Ophthalmology.

6.3 Discussion

6.3.1 Formation of optic cups from BBS iPSCs

6.3.1.1 Recapitulation of the neural retina

It has been shown for the first time that BBS patient-derived iPSCs can be differentiated into 3D optic cups. These cups recapitulated *in vivo* development, including topological organisation of retinal neural cells and expression of distinct genes required for mammalian eye development. Furthermore, in late stage differentiations spanning day 147 to day 198, mature phototransduction proteins were detectable, including rhodopsin and cone arrestin. Disappointingly, there were fewer mature rhodopsin- and cone arrestin-expressing cells in the cultures of this study compared to published protocols (Nakano et al. 2012; Zhong et al. 2014), despite abundant expression of CRX and RCVRN. However, rhodopsin expression has been shown to be variable, perhaps depending on retinoic acid exposure or particular differences between cell lines (Zhong et al. 2014; Mellough et al. 2012).

Moreover, the neural retina was able to form without the presence of the surface ectoderm, which is known to have a crucial role in retinal specification. Zhong et al. (2014) hypothesise that this was able to occur due to the formation of a micro-environment, established by cell-cell interactions and endogenous secretions. These signalling cues were able to induce self-driven lamination that led to a well-organised NR displaying the innate inner nuclear and outer nuclear layers seen in the native retina. The neuronal cells were not as abundant and more scattered than *in vivo*, which this is likely due to lack of rigid borders in the optic cup *in vitro*.

6.3.1.2 Formation of connecting cilia and outer segments

The ultrastructure of BBS patient-derived optic cups was analysed by transmission electron microscopy. TEM imaging identified an organised and well-structured photoreceptor layer, where connecting cilia had started to emerge from mitochondria-rich inner segments. Connecting cilia were identifiable by their 9 + 0 arrangements and long microtubular axonemes that extended from a distinctive basal body. At the distal tip of some of these connecting cilia, club-like outer segments were starting to form. These immature structures were reminiscent of the developing outer segments of photoreceptors *in vivo* at fetal week 18 (Hendrickson et al. 2008). In some cases, disordered tubulovesicular structures were distinguishable in the nascent outer segment. This material could be rudimentary undeveloped discs, such as those published from other *in vitro* developing photoreceptor cultures (Saga et al. 1996; Parfitt et al. 2016).

During formation of the optic cup *in vivo*, the NR and RPE form adjacent layers, where each is mutually beneficial to the other. The RPE is known to be essential for the development and maintenance of photoreceptors, but its direct role in maturation of the NR is not entirely understood. In the current literature, the remarkable characteristic of *in vitro* photoreceptor differentiation is that they can develop without direct contact between the NR and the RPE; the cultures presented here were no exception. Although a great proportion of optic cups developed RPE cells, they were always seen bundled at the tip or in the middle of the cup, and never lining the surface of the cup. RPE cells secrete a wealth of growth factors, including FGF1, FGF2, insulin-like growth factor 1, ciliary neurotrophic factor, and pigment epithelium derived factor (German et al. 2008). It may be that these secretions were enough to initiate photoreceptor development, and thus direct RPE contact is not essential for the very early stages of outer segment formation. However, it is known that during development *in vivo,* photoreceptors extend and elongate towards the RPE (German et al. 2008). As a result, it is likely that outer segment maturation cannot occur without support and direct contact with RPE cells. With regards to functionality, efficient phototransduction and maintenance of the visual cycle cannot occur without direct contact between photoreceptors and RPE. Therefore, adjustments should be made to current protocols to develop methods for co-culturing RPE and photoreceptors in their native configuration.

6.3.2 Expression of the BBSome and IFT machinery

Having established the presence of photoreceptors in BBS patient-derived optic cups, it was investigated whether cilia proteins were expressed and correctly localised in the BBS patient line BBS001. My findings show that BBS1 localised to the apical surface of the optic cup, where inner segments have formed. It was then determined that BBS5 is expressed in the same outer layer as BBS1. These two proteins are components of the BBSome. In some cases, acetylated tubulin staining identified connecting cilia, which emerged from inner segments positive for BBS1 and BBS5. I therefore surmise that the BBSome is localised to the basal body of nascent connecting cilium of these rudimentary photoreceptors. This is consistent with localisation of the BBSome within photoreceptors *in vivo* (Jin et al. 2010; Smith et al. 2013).

Although it is generally thought that BBS patients are born with normal vision, it is not 100% certain as most patients are diagnosed after eyesight has started to deteriorate. Through executing these experiments, the target was to gain some understanding of the pathobiology of

retinal degeneration in BBS patients. However, from the example of the line obtained from patient BBS001, there were no recognisable developmental abnormalities related to BBS protein localisation. Consistent with the findings presented here, the p.M390R point mutation has not been shown to affect BBS1 localisation *in vivo*, but is predicted to influence entry and exit into the cilium to control IFT (Liew et al. 2014; Jin et al. 2010; Mourão et al. 2014). Therefore, IFT88, a member of the IFT-B machinery critical for photoreceptor assembly and maintenance (Pazour et al. 2002), was explored. Immunohistochemical staining for IFT88 demonstrated expression of some IFT particles at the distal tips of the connecting cilia. These results indicate that active ciliogenesis was able to occur, which suggests that during development the BBSome allows normal entry into the connecting cilium.

Despite the underlying BBS1 mutation seen in this patient, the differentiation procedure was consistent with *in vivo* development, and BBS and IFT proteins localised correctly. Therefore, from the experiments conducted in this study, there were no recognisable BBS-specific abnormalities discovered in these developing photoreceptors. This observation is supported by a functional study by Mockel et al. (2012), where apoptosis of *Bbs12*$^{-/-}$ photoreceptors occurred only once the outer segments had differentiated and matured. Accordingly, I would surmise that in general BBS proteins are not required for the development of photoreceptors, but are essential for photoreceptor function. This late deterioration of photoreceptors could provide an opportunity for early intervention that would prevent visual loss before it occurs.

If BBS proteins are involved only in functionality of photoreceptors, this points to the conclusion that visual degeneration is triggered by a light stimulus. In support of this, the BBSome has demonstrated some light sensitivity. BBS5 has been shown to undergo phosphorylation in response to light to facilitate arrestin transport (Smith et al. 2013) and BBS4 is required for light-dependent protein transport to the outer segment (Abd-El-Barr et al. 2007). Encompassing evidence from my own and published studies, I present a model as the foundation of photoreceptor degeneration in BBS patients: The transportation of phototransduction proteins to the outer segment is activated by a light stimulus. These phototransduction proteins are shuttled from the inner segment to the outer segment via the connecting cilium, entry into which is controlled by the BBSome. Under a light stimulus, IFT must occur efficiently as an estimated 2000 opsin molecules are delivered each minute through the connecting cilium (Wolfrum and Schmitt 2000; Pazour et al. 2002). As a result of a faulty BBSome, the proteins cannot be shuttled quickly enough, thus proteins accumulate in the inner segment and outer nuclear layer (Abd-El-

Barr et al. 2007; Mockel et al. 2012). This accumulation leads to stress on the endoplasmic reticulum and triggers the unfolded protein response pathway and imminent apoptosis (Mockel et al. 2012). In order to investigate this hypothesis further, I would need to conduct light response experiments on photoreceptors with mature outer segments, and importantly have control photoreceptors available for these studies.

6.3.3 Variability within iPSC cultures

Upon starting this project, the aim was to differentiate 5 different iPSC lines. Unfortunately, the success rate was low and two lines failed to express key retinal genes entirely. Due to the irregularities between cultures being so great, comparison between lines for functional experiments would be unreliable. It would be impossible to ascertain whether significant differences were real or due to experimental variability. Variation between lines is not uncommon in iPSC cultures (Hirami et al. 2009; Zhong et al. 2014; Meyer et al. 2009; Kilpinen et al. 2017) and shall be addressed in this section.

6.3.3.1 Embryoid body formation

Embryoid body formation can create one of the greatest sources of variation in stem cell differentiation cultures. Although they present an efficient method of initiating differentiation, EB size, number, and quality are difficult to control. These variables influence the cell-cell spatial coordination within the EB, which affects the course of differentiation and makes reproducibility difficult to ensure (Rungarunlert et al. 2009).

There are different methods for maintaining EBs. A static suspension method was used in this project. This cultures all EBs for each experiment in a single environment, such as in a T25 cell culture flask. This method was advantageous because it was the least labour intensive and is popular for differentiations into neural lineages (Rungarunlert et al. 2009). On the other hand, difficulties were faced when generating aggregates due to agglomeration of multiple EBs and in some cases premature attachment to the surface of the culture plate. These problems led to high cell death rates and heterogeneity within a single experiment, which could have had a big impact on differentiation downstream. These issues could be overcome in future experiments through use of a different method for EB development, such as a hanging drop or using a 96-well U-bottomed plate. Both techniques prevent agglomeration and ensure high uniformity as EBs are kept separate from one another (Rungarunlert et al. 2009).

6.3.3.2 Epigenetic variability

During one differentiation, a bundle of cardiomyocytes was discovered, which formed around day 20 in the control line CTRL038 (supplemental data, movie S.1). The presence of this demonstrates the unpredictability associated with iPSC culture and shows the importance of checking cultures regularly for any abnormalities. This phenomenon may also have occurred due to reprogramming inefficiencies associated with incomplete reprogramming of the epigenome. Reprogramming of somatic cells to iPSCs is induced through epigenetic transformation and not a genetic alteration. Although studies have shown that at a transcriptomic level iPSCs are almost identical to ESCs, multiple reports have identified that iPSCs possess an epigenetic memory retained from their host cell (Lister et al. 2011; Kim et al. 2010; Polo et al. 2010; Hu et al. 2010). This epigenetic memory can be transmitted at a high frequency through differentiation and affect differentiation propensity. For example, cardiomyocytes and fibroblasts both belong to the mesodermal germ layer. Therefore, the line CTRL038 may have been predisposed to differentiate into cardiomyocytes due to its epigenetic memory of mesenchymal cells. This could also explain why no viable differentiations were obtainable from this line.

Using keratinocytes as a host cell for reprogramming could enhance the chance of successful optic cup differentiation. Compared to dermal fibroblasts, keratinocytes have proved to be twice as fast and 100-fold more efficient at reprogramming (Maherali et al. 2008; Barrero et al. 2012). Additionally, keratinocytes arise from the ectoderm and thus are the same germ lineage as the neural retina. iPSCs derived from Usher patient keratinocytes have been used to generate defined neural retina and RPE layers, closely mimicking the configuration seen *in vivo* (Tucker et al. 2013). As yet, this has not been reported in iPSCs derived from fibroblasts.

6.3.3.3 Genetic variability

The recent study from the HipSci project characterised iPSC lines derived from 189 individuals and showed that the greatest variation factor in reprogrammed iPSCs was in fact genetic background from the donor (Kilpinen et al. 2017). These differences were shown to affect pluripotency, cell morphology and importantly, differentiation capacity. Concordantly, in my experiments, large differences were noticeable in differentiation capacity between lines, irrespective of whether they were BBS patient-derived or from healthy control individuals. For example, the iPSCs derived from BBS001 differentiated very well compared to the line from BBS007, and the control line CTRL038, which both failed to produce operable differentiations.

One of the contributors to genetic variability may arise from somatic mutations found in dermal fibroblasts. The skin is typically prone to ultraviolet radiation exposure, which causes a high frequency of point mutations that would otherwise be absent in germ cells. The selection and accumulation of somatic mutations occurs through reprogramming and passaging selection during fibroblast and iPSC culture (Gore et al. 2011). This also poses problems downstream for therapeutic uses of iPSCs and introduces the need for high-throughput functional genomics to identify detrimental or carcinogenic mutations before clinical use. The HipSci project have addressed this issue by implementation of genotyping arrays on both somatic cells and iPSCs prior to banking. This aimed to validate that genetic integrity was maintained throughout the reprogramming process and allowed removal of cell lines that accumulated aberrant variants, such as chromosomal aberrations or structural variants (Kilpinen et al. 2017).

6.3.3.4 Media and culturing techniques

The results described here were consistent with the protocol published by Zhong et al. (2014) despite the difference in iPSC culturing system. However, the success rate of differentiation was generally lower than Zhong and colleagues, who were able to successfully differentiate three iPSC lines with 50-70% efficiency. One reason for the difference in efficiency may be due to using an E8 and vitronectin system, which was not possible to optimise due to time constraints within this project.

E8 was developed by Chen et al. (2011) as a replacement to mTESR for iPSC feeder-free culture (Chen et al. 2011). mTESR tends to be more commonly used, but contains over 20 highly variable ingredients, typically obtained from animal sera, such as albumin and Matrigel (Ludwig and A Thomson 2007). As a result, variability across cultures grown with mTESR can be substantial. On the other hand, E8 is nonxenogeneic and contains 8 essential ingredients that ensure pluripotency and cell health, whilst controlling for variability. Despite the benefits of iPSC culture using E8, there have yet to be any published reports that have used E8 maintained iPSCs for differentiation. Similarly, no studies are available demonstrating that iPSCs grown on vitronectin can differentiate successfully to a neuroepithelial lineage. In all cases in the literature, a Matrigel substrate is utilised for differentiation into neurospheres. Matrigel's efficiency is owing to its heterogeneity, which favourably recapitulates the extracellular matrix *in vivo* (Hughes et al. 2010). In contrast, vitronectin is comprised of a single component, which may not provide the same level of support.

6.3.4 Future role for this model

6.3.4.1 Optimisation of culture system

Optic cup development is a complex process orchestrated by elaborate time sensitive signalling pathways. This process was possible without the addition of exogenous factors to the culturing system. Explicitly, this was one of the reasons why the Zhong et al. protocol was chosen over other publications. A low maintenance, unelaborate protocol was more compatible for a single person to undertake several simultaneous differentiations. In hindsight, the lack of exogenous factors created a long-winded experiment with low success rates.

For future experiments, the addition of retinogenesis specific agonists and antagonists could be included to enhance photoreceptor differentiation and efficiency. As an example, inhibition of BMP and Wnt signalling can increase *PAX6* and *RX* expression (Meyer et al. 2009), which could enhance the prosperity of eye field development in the early stages of differentiation. Similarly, Notch suppression by DAPT treatment has shown to exhibit strong acceleration of photoreceptor differentiation (Nakano et al. 2012) and doubled early *CRX* and *RCVRN* expression between week 3 and 4 of differentiation (Reichman et al. 2014). Additionally, introduction of FGF2 may be sufficient to represent an environment similar to the surface ectoderm. These exogeneous factors will require precise optimisation with time sensitive delivery. If used appropriately, they may enhance reproducibility, as differentiation will no longer depend on micro-environmental cues or endogenous secretions.

There are other modifications that would benefit future experiments. A small adjustment would be to switch a 6-well culture plate with a 96-well U-bottomed plate. This would allow more precision and control over the development of each individual optic cup. Additionally, this technique would reduce some of the variability caused by EB agglomeration. On the other hand, this would increase maintenance time each day, which would make multiple lines difficult to manage for one person.

Another option would be to switch from a 3D culture system to 2D culture. The benefit of this technique is a higher cell viability, as growth factors do not have to penetrate several cell layers of the optic cup. Additionally, multiple homogenous biopsies can be easily extracted for analysis or further subculture, without affecting future development. Tucker et al. (2013) were successful in their recapitulation of layered RPE and NR, similar to *in vivo* configuration, and expression of opsins by day 120 of culture (Tucker et al. 2013).

6.3.4.2 Model for therapeutic use

This project has shown differentiation of BBS patient-derived iPSCs into optic cups is possible, but there is much to accomplish before utilisation as a disease model. The greatest barrier to overcome is the lack of reproducibility and unpredictability of the differentiation process. The best way to ensure reproducibility would be to increase the number of differentiations per line and also add clones to assess differences in the reprogramming procedure. To reliably evaluate biological differences between disease and control groups, multiple lines for each group would need to be successfully differentiated. As a whole, this process would be extremely time consuming and expensive. On the other hand, if optimised properly, this technique could provide an exciting patient specific model for therapeutic uses.

In this research field, the challenge is to recreate functional outer segments that are rich with membranous discs. Zhong et al. (2014) revealed some excitation in response to a light stimulus with immature nascent outer segments similar to those generated in this project. Due to time restraints, it was not possible to test BBS optic cups for functionality. For future work, demonstration of a light response may provide answers to the role of the BBSome in photoreceptors. If functionality were observed, the utility of the optic cups could be invaluable to test therapeutics in a patient specific model. For example, gene therapy models are being developed to rescue eyesight in *BBS1*$^{M390R/M390R}$ knock-in mice. This involves injecting a functional gene copy packaged in Adeno-Associated Virus (AAV) vectors into the retina. Before use in patients, this therapy would need to be optimised and validated in a patient-specific model. Furthermore, this model could be used for a drug screen to decelerate the loss of photoreceptors. Mockel et al. demonstrated that drug intervention to inhibit the unfolded protein response pathway prolonged photoreceptor viability in *Bbs12*$^{-/-}$ mice (Mockel et al. 2012).

Finally, there is potential for the transplantation of photoreceptors grown *in vitro* into the patient retina (Gonzalez-Cordero et al. 2013; MacLaren et al. 2006; Lakowski et al. 2011; Tucker et al. 2013; Homma et al. 2013; Lamba et al. 2010; Tucker et al. 2011). iPSCs derived from patient somatic cells allows autologous transplantation, which would avoid immune rejection and therefore treatment with immunosuppressors. There has been success with transplantation of retinal cells derived from both ESC and iPSC models back into mouse disease models (MacLaren et al. 2006; Gonzalez-Cordero et al. 2013). Following transplantation, photoreceptor precursors expressing *CRX* have been shown to integrate into the host retina, extend neuronal projections,

develop outer segments that extend towards the RPE, and induce partial restorations of electrophysiological function (Tucker et al. 2011; Lakowski et al. 2011; Tucker et al. 2013).

Chapter 7: Final Conclusions

7.1 Summary of findings

This project was the first of its kind to investigate multiple biological materials from BBS patients with the overall objectives to:

- Investigate whether interfamilial heterogeneity, which presents as differences in disease onset or severity, is a consequence of genetic modifying alleles that interact with BBS1 in the primary cilium.

- Utilise different proteomic profiling technologies to discover molecular biomarkers that may be used to measure disease progression or detect secondary features of BBS from urine and plasma biofluids.

- Using a multi-omic strategy, integrate transcriptomic and proteomic profiling data to uncover differentially regulated pathways between BBS and control fibroblasts that may elucidate BBS pathology.

- Develop a tissue specific model from BBS patient-derived iPSCs for the investigation of retinal dystrophy *in vitro*.

- Trial novel technologies, as well as develop protocols for use in future multi-omic studies for rare diseases.

The main findings of this thesis are discussed here and summarised in **figure 7.1**:

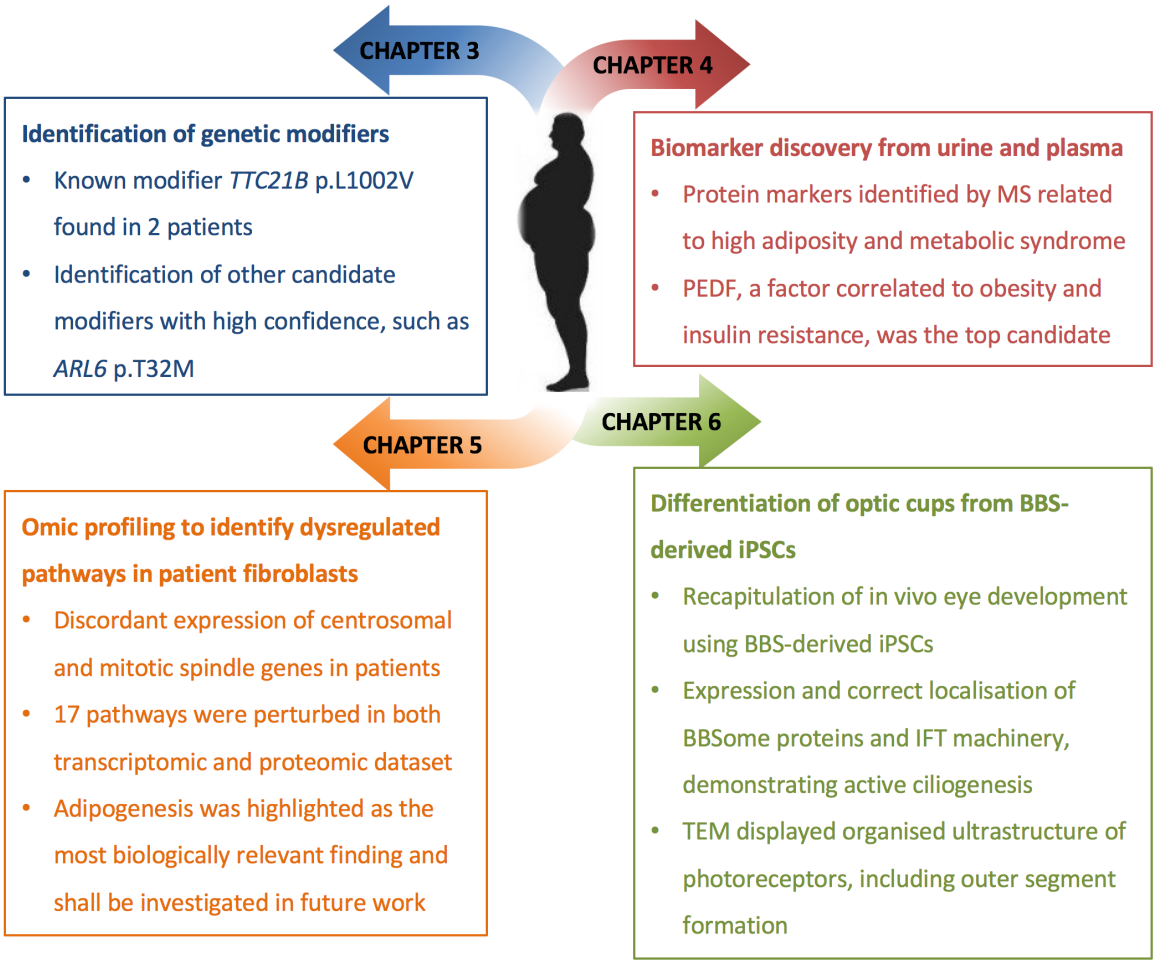

Figure 7.1 – Summary of findings for each chapter.

7.1.1 There is evidence of secondary modifying variants in BBS patients, which may point towards the underlying causes of phenotypic variability. However, these variants do not contribute to the overall mutational burden of BBS

Phenotypic heterogeneity is a common feature of BBS, which has been proposed to be controlled by genetic modifiers (Beales et al. 2003; Katsanis, Lupski, et al. 2001; Khanna et al. 2009; Zaghloul and Katsanis 2009; Katsanis, Ansley, et al. 2001; Badano et al. 2006; Zaki et al. 2011). This variability was reflected within the cohort of this project; therefore, it was investigated whether genetic modifiers could be identified that may attribute to interfamilial heterogeneity (described in chapter 3). Within this project, I generated a 4-tier categorisation system for identification of modifiers from whole genome sequencing data. This demonstrated its effectiveness through the discovery of 3 known modifiers in 4 patients, namely *RPGRIP1L* p.A229T (Khanna et al. 2009), *TTC21B* p.L1002V (Davis et al. 2011), and *BBS12* p.S429T (Zaghloul et al. 2010). Furthermore, an additional 4 variants, *ARL6* p.T32M, *BBS4* p.V266G, *CEP164* pQ1410*, and *AHI1* p.R548H, were

uncovered and assigned high confidence due to evidence collected by protein interactions with BBS1, pathogenic prediction algorithms, or for the case of *AHI1*, a previously established modifying gene in ciliopathies (Tory et al. 2007; Louie et al. 2010; Coppieters et al. 2010). Further work will be required to determine whether putative modifiers have the capability of altering the expression *BBS1* p.M390R in these patients. This would best be achieved through a series of gene correction studies discussed further in chapter 3.

Mutational burden analysis was then completed to establish whether the genetic modifying variants found in BBS patients attributed to enhanced genetic burden. This project determined that, within this cohort and another set of 16 patients sequenced by whole exome sequencing, there was no evidence that modifying variants contributed to the total mutational load of the disease. This was determined with 2 primary cilia gene sets (ciliopathy and SCGSv1 panels), and a non-related CMT disease gene panel for use as a control. The finding of no mutational burden supported a previous study, which determined a similar finding with 31 ciliopathy patients (Shaheen et al. 2016). Further investigation of patients with *BBS1* p.M390R uncovered that they had significantly fewer variants than control individuals when the *BBS1* p.M390R allele was removed from the total variant count. This was an unexpected finding as it conflicted with the result of the burden analysis with 16 WES BBS patients. This outcome may have complicated our understanding of the underlying genetics of *BBS1* p.M390R patients. Further work is required to determine whether this was a true finding, or a consequence of small sample sizes or experimental design.

7.1.2 Proteomic profiling uncovered molecular biomarkers that may be indicative of secondary features of obesity

This was the first study to utilise biofluids obtained from ciliopathy patients for identification of protein biomarkers (described in chapter 4). The current marker tested in BBS clinics is CRP, a generic indicator of inflammation. Thus, there was a need for molecular biomarkers that are indicative of disease progression, that may be specific to BBS, or specific to phenotypic features of BBS. This project implemented label-free mass spectrometry and a novel targeted technology, SOMAscan, to detect differentially expressed proteins in urine and plasma between BBS patients and control subjects. Both technologies and biofluids revealed markers that are associated with obesity or secondary features of high adiposity, such as type II diabetes or cardiovascular disease.

The biomarker identification strategy used in this project uncovered candidate markers, such as PEDF and ORM1 from urine, and ApoM from plasma. Importantly, further investigation must be completed to determine the reliability of these findings in a larger BBS cohort. Biomarker discovery hierarchy described by Rifai et al. (2006) indicates that verification should be implemented with 100s of patients (**figure 4.1**). This could be completed in a large profiling project, implementing urine samples from all BBS patients in the clinic, for which there are 520 patients in the UK (Prof P. Beales, personal communication). The benefit of urine is that it is non-invasive and already routinely obtained for clinical testing.

7.1.3 A multi-omic strategy identified dysregulated pathways in BBS cell cycling and ciliated fibroblasts compared to controls

Fibroblasts from patients and controls were investigated under serum fed and serum starved conditions, to compare differences in cell cycling and ciliated cells (described in chapter 5). To our knowledge, the resulting 3-way comparison strategy was the first example of an investigation of differences between ciliated and non-ciliated cells, as well as the global analysis of patient derived cultures. This analysis uncovered a highly complex response to starvation, as well as some biologically relevant changes between BBS and controls at the pathway level. These processes shall be targets for further validation as a consequence of this study.

Pathway analysis of RNA-seq genes uncovered discordant expression of centrosomal genes between BBS and control cells. It was revealed that genes involved in mitotic spindle nucleation, such as *HAUS3* and *CEP152*, were significantly downregulated in serum fed BBS cultures compared to controls. This may suggest a disruption of mitotic spindle organization or cytokinesis in BBS cell cycling cultures. This study also investigated proteomics data from iTRAQ mass spectrometry, which revealed upregulated processes, including regulators of actin cytoskeleton, cilia assembly, and hedgehog signaling. Incorporation of BBS1 and its interacting proteins into a protein network with DE proteins allowed assessment of possible mechanisms for upregulation of these pathways. Although highly complex, this technique enabled speculation that the *BBS1* p.M390R variant in patient cells has reduced affinity of BBS1 with its interactions, and thus has resulted in accumulation of proteins.

The gene set adipogenesis was positively enriched by both transcriptomic and proteomic profiling technologies in the serum fed BBS vs control comparison, which elucidate mechanistic insight

into obesity manifested by BBS patients. Although, individually genes were rarely statistically significant, GSEA determined that the collective trend was significant. Evaluation of the individual genes and pathway topology mapping identified that components predominantly involved in mitochondrial function were upregulated, which may suggest that the enrichment for adipogenesis was driven by unbalanced mitochondrial activity in BBS fed cells compared to controls. This work may support previous findings that suggest increased adiposity in BBS patients, thought to be driven by aberrantly upregulated adipogenic drivers (Marion et al. 2009; Marion et al. 2012). Further work to validate this finding was discussed in chapter 5.

Within this thesis, RNA-seq was analysed by looking at changes in gene expression and dysregulation of pathways between biological groups. However, integration of RNA-seq data with genetic variation information can be used to aid analysis of the genome. For example, differences in gene expression may be linked to regulatory variants, such as eQTLs, which influence the expression level of genes (Morley et al. 2004). Although, large sample sizes are required for eQTL association studies, this project identified some interesting variants for further study, such as *RPGRIP1L*, which was flagged in both the genetic modifier and RNA-seq studies (discussed in chapter 3 and 5, respectively). There were 2 qualifying variants in *RPGRIP1L* from the 4-tier categorisation system found in 2 patients; p.A229T in BBS003 and p.A1183G in BBS013 (**table 3.3iv**). In addition, *RPGRIP1L* was highlighted as discordantly expressed in response to starvation in patients compared to controls (**figure 5.14 and 5.15**). Therefore, future work would also entail analysis of the RNA-seq data to determine whether there are genetic explanations for significant changes in gene expression, such as with *RPGRIP1L*.

7.1.4 BBS-patient derived iPSCs were used to develop a model for investigation of retinal dystrophy *in vitro*, which successfully recapitulated gene expression of the native developing retina

Within this thesis, it was demonstrated for the first time that iPSCs derived from BBS patient fibroblasts could differentiate into 3D optic cups, which exhibited the key stages of gene expression required for *in vivo* retinal development (described in chapter 6). Investigation of the ultrastructure by TEM of optic cups at day 198 of differentiation revealed nascent photoreceptor outer segments, morphologically similar to differentiating outer segments from wild type cells in other studies (Saga et al. 1996; Zhong et al. 2014; Parfitt et al. 2016). Furthermore, this was the first study to investigate the expression BBSome proteins within iPSC models, which

demonstrated that BBS1 and BBS5 were localised to the basal body of the connecting cilia, as expected. Although at this stage of development there were no control cultures for comparison, using knowledge taken from the literature (Smith et al. 2013; Jin et al. 2010; Parfitt et al. 2016), it appeared that the BBSome localisation and ciliogenesis was not affected in the BBS patient-derived optic cups.

Several types of biological material obtained from BBS patients were utilised in this thesis, namely urine, plasma, and dermal fibroblast cultures. Although each element has its advantage, they are not tissue types that are known to exhibit BBS-specific phenotypes. Disease relevant tissues obtained from patients may be a more valuable source of biological information, such as renal progenitors to study polycystic kidney disease, hypothalamic neurons for investigation of satiety response, or retina to model rod-cone dystrophy. These tissues are difficult to obtain from patients for research purposes, which is why iPSCs are a highly valuable resource. A future project would be to implement models, such as BBS patient-derived optic cups, in a multi-omic study with the framework developed in this thesis. This would provide valuable insight into the pathological mechanisms within a disease relevant tissue, as well as being able to study the development of the disease *in vitro*.

7.2 Considerations for further work

This thesis doubled as a pilot project to develop and implement a multi-omic strategy for investigation of rare diseases. Typical challenges were faced that relate to studying rare disease cohorts, such as disease heterogeneity and small sample sizes. However, having confronted these challenges, future studies can be devised with improved study designs, encompassing optimised standard operating protocols (SOPs). Some considerations to consider for future work are described here.

Patients with the variant *BBS1* p.M390R were chosen for this study for two key reasons. Firstly, this is the most common variant in European populations (Mykytyn et al. 2002), therefore the largest population of patients in the BBS clinic carry at least one disease causing *BBS1* p.M390R variant. This not only means that this variant carries the greatest burden on the population, but also that larger cohorts can be assembled of patients with the same genotype, attributing to more statistical power. Secondly, despite it being very common in the population, the pathobiology of *BBS1* p.M390R is not well understood, partly accredited to the complex nature of a hypomorphic variant. This is reflected by the spectrum of phenotypes associated with *BBS1* p.M390R, where

presentation varies from all cardinal features of BBS to asymptomatic individuals (Beales et al. 2003; Estrada-Cuzcano et al. 2012; Cox et al. 2012). In many ways, this disease heterogeneity complicated the outcome of the study and due to statistical problems created by small sample size, it made interpretation of results quite challenging. For example, there were few differentially expressed genes found at the transcriptome level between starved BBS and control fibroblasts (described in chapter 5). It is possible that the study of a loss of function mutation may have exhibited a greater difference.

As well as advancing our understanding of *BBS1* p.M390R pathobiology, this study has demonstrated proof of principle, and ensured optimisation of methodology and expertise for future experiments. Increased sample size has already been discussed as a limitation throughout this thesis, therefore, this would need to be addressed first. There are 520 patients that are registered to attend the BBS clinics headed by Professor Phil Beales, of which ~166 individuals (32%) manifest at least one *BBS1* p.M390R variant (Prof P. Beales, personal communication). The next stage of this analysis would therefore be to recruit patients for validation of findings. Further to this, patients with different disease-causing variants could also be recruited to the study. The variant p.C91Lfs*5 in *BBS10* is responsible for ~20% of BBS cases, and is the next most common mutation found in BBS (Forsythe and Beales 2013), present in ~145 registered patients at Professor Beale's clinics (Prof P. Beales, personal communication). It would be interesting to determine whether the findings of this study were specific to *BBS1* p.M390R, or a general finding to BBS by comparison between these two genotypes. As a loss of function variant, I would hypothesise that patients with *BBS10* variants would exhibit greater differences compared to controls. Finally, it would be imperative that a reliable control group be recruited for this study, where it would be preferable that there are double the number of controls to patients. This would benefit statistical power and allow reliable inferences from the data.

Appendix

AHI1	*BBS9*	*EVC2*	*NEK1*	*SPATA7*	*USH1C*
ARL13B	*BBIP1*	*GLIS2*	*NEK8*	*TCTN1*	*USH1G*
ARL6	*C2orf71*	*GPR98*	*NPHP1*	*TCTN2*	*USH2A*
ATXN10	*C8orf37*	*HYLS1*	*NPHP3*	*TMEM138*	*VHL*
B9D1	*CC2D2A*	*IFT43*	*NPHP4*	*TMEM216*	*WDPCP*
B9D2	*CCDC28B*	*IFT80*	*OFD1*	*TMEM231*	*WDR19*
BBS1	*CDH23*	*IFT172*	*PCDH15*	*TMEM237*	*WDR35*
BBS10	*CEP164*	*INVS*	*PKD2*	*TMEM67*	*XPNPEP3*
BBS12	*CEP290*	*IQCB1*	*PKHD1*	*TOPORS*	*ZNF423*
BBS2	*CEP41*	*KIF7*	*RPGR*	*TRIM32*	
BBS4	*DFNB31*	*LCA5*	*RPGRIP1*	*TTC21B*	
BBS5	*DYNC2H1*	*MKKS*	*RPGRIP1L*	*TTC8*	
BBS7	*EVC*	*MKS1*	*SDCCAG8*	*TULP1*	

Table A.1 – Genes known to cause nonmotile ciliopathies, referred to as the ciliopathy panel.

ADCY3	*CEP97*	*HSPB11*	*NEK4*	*RPGR*	*TRAPPC9*
AHI1	*CLDN2*	*HTR6*	*NEK8*	*RPGRIP1*	*TRIM32*
ALMS1	*CLUAP1*	*HTT*	*NGFR*	*RPGRIP1L*	*TRIP11*
ARF4	*CP110*	*HYLS1*	*NME5*	*RTTN*	*TTC12*
ARL13B	*CRB3*	*IFT122*	*NME7*	*SASS6*	*TTC21B*
ARL3	*CROCC*	*IFT140*	*NOTO*	*SCLT1*	*TTC26*
ARL6	*CTNNB1*	*IFT172*	*NPHP1*	*SDCCAG8*	*TTC30A*
ASAP1	*DCDC2*	*IFT20*	*NPHP3*	*SEPT2*	*TTC30B*
ATXN10	*DCDC2*	*IFT27*	*NPHP4*	*SEPT7*	*TTC8*
AZI1	*DFNB31*	*IFT43*	*OCRL*	*SHH*	*TTK*
B9D1	*DPYSL2*	*IFT46*	*ODF2*	*SLC47A2*	*TTLL3*
B9D2	*DRD1*	*IFT52*	*OFD1*	*SMO*	*TTLL6*
BBS1	*DRD2*	*IFT57*	*PARD3*	*SNAP25*	*TTLL9*
BBS10	*DRD5*	*IFT74*	*PARD6A*	*SNX10*	*TUBA1A*
BBS12	*DVL1*	*IFT80*	*PCDH15*	*SPATA7*	*TUBA1C*
BBS2	*DYNC2H1*	*IFT81*	*PCM1*	*SPEF2*	*TUBA4A*
BBS4	*DYNLT1*	*IFT88*	*PDE6D*	*SSNA1*	*TUBB2A*
BBS5	*DYX1C1*	*INPP5E*	*PDZD7*	*SSTR3*	*TUBB2B*
BBS7	*EVC*	*INVS*	*PHF17*	*STIL*	*TUBB3*
BBS9	*EVC2*	*IQCB1*	*PIBF1*	*STK38L*	*TUBE1*
BBIP1	*EXOC4*	*KIF17*	*PKD1*	*STOML3*	*TUBGCP2*
C21orf2	*EXOC5*	*KIF19*	*PKD1L1*	*STX3*	*TUBGCP3*
C2CD3	*EXOC6*	*KIF24*	*PKD2*	*SUFU*	*TUBGCP4*
C2orf71	*EXOC6B*	*KIF27*	*PKHD1*	*SYNE2*	*TUBGCP5*

C8orf37	*FAM161A*	*KIF3A*	*PLK1*	*TBC1D30*	*TUBGCP6*
CBY1	*FBF1*	*KIF3B*	*POC1A*	*TBC1D7*	*TULP1*
CC2D2A	*FLNA*	*KIF3C*	*PTCH1*	*TCTN1*	*TULP3*
CCDC28B	*FOPNL*	*KIF7*	*PTPDC1*	*TCTN2*	*USH1C*
CCDC37	*FOXJ1*	*LCA5*	*RAB11A*	*TCTN3*	*USH1G*
CCDC41	*FUZ*	*LZTFL1*	*RAB11FIP3*	*TMEM138*	*USH2A*
CDH23	*GLI1*	*MAK*	*RAB17*	*TMEM216*	*VHL*
CENPJ	*GLI2*	*MAL*	*RAB23*	*TMEM231*	*WDPCP*
CEP104	*GLI3*	*MAPRE1*	*RAB3IP*	*TMEM237*	*WDR19*
CEP135	*GLIS2*	*MCHR1*	*RAB8A*	*TMEM67*	*WDR35*
CEP164	*GPR161*	*MDM1*	*RABL5*	*TNPO1*	*WDR60*
CEP250	*GPR98*	*MKKS*	*RFX3*	*TOPORS*	*XPNPEP3*
CEP290	*GSK3B*	*MKS1*	*RILPL1*	*TPPP2*	*ZNF423*
CEP41	*HAP1*	*MLF1*	*RILPL2*	*TRAF3IP1*	
CEP72	*HNF1B*	*NEK1*	*RP1*	*TRAPPC10*	
CEP89	*HSPA8*	*NEK2*	*RP2*	*TRAPPC3*	

Table A.2 – Nonmotile cilia genes from the SYSCILIA gold standard version 1 gene panel.

PLEKHG5	*FGD4*	*PIKFYVE*	*IKBKAP*	*DNMT1*	*CARS*
KIF1B	*HSPB8*	*SOD1*	*SETX*	*DYNC1H1*	*NARS*
MFN2	*TDP1*	*SOX10*	*GJB1*	*INF2*	*IARS*
YARS	*SLC12A6*	*RAB7A*	*PRPS1*	*TFG*	*VARS*
NGF	*LITAF*	*CCT5*	*MED25*	*HARS*	*SARS*
LMNA	*KARS*	*SH3TC2*	*PMP22*	*KIF1A*	*EPRS*
NTRK1	*GAN*	*FIG4*	*TRPV4*	*FAM134B*	*SPTLC3*
MPZ	*SEPT9*	*GARS*	*AARS*	*HK1*	*GNB4*
EGR2	*CTDP1*	*HSPB1*	*LRSAM1*	*RARS*	*PDK3*
SBF2	*DNM2*	*ARHGEF10*	*MARS*	*TARS*	
BSCL2	*PRX*	*GDAP1*	*SPTLC2*	*FARSA*	
IGHMBP2	*DCTN1*	*NDRG1*	*ATP7A*	*FARSB*	
WNK1	*HOXD10*	*SPTLC1*	*ATL1*	*DARS*	

Table A.3 – Genes associated with causing Charcot-Marie-Tooth disease, from (Gonzaga-Jauregui et al. 2015).

Sample ID	Status	Batch number	Disease
BBS001	Patient genome	WGS batch 2	Bardet-Biedl Syndrome
BBS002	Patient genome	WGS batch 2	Bardet-Biedl Syndrome
BBS003	Patient genome	WGS batch 2	Bardet-Biedl Syndrome
BBS004	Patient genome	WGS batch 2	Bardet-Biedl Syndrome
BBS005	Patient genome	WGS batch 2	Bardet-Biedl Syndrome
BBS006	Patient genome	WGS batch 2	Bardet-Biedl Syndrome
BBS007	Patient genome	WGS batch 2	Bardet-Biedl Syndrome

BBS008	Patient genome	WGS batch 2	Bardet-Biedl Syndrome
BBS009	Patient genome	WGS batch 2	Bardet-Biedl Syndrome
BBS010	Patient genome	WGS batch 2	Bardet-Biedl Syndrome
BBS011	Patient genome	WGS batch 2	Bardet-Biedl Syndrome
BBS012	Patient genome	WGS batch 2	Bardet-Biedl Syndrome
BBS013	Patient genome	WGS batch 2	Bardet-Biedl Syndrome
BBS014	Patient genome	WGS batch 2	Bardet-Biedl Syndrome
BBS015	Patient genome	WGS batch 2	Bardet-Biedl Syndrome
SRS002	Control genome	WGS batch 3	Clinically healthy
SRS003	Control genome	WGS batch 3	Clinically healthy
SRS005	Control genome	WGS batch 3	Clinically healthy
SRS006	Control genome	WGS batch 3	Clinically healthy
SRS008	Control genome	WGS batch 3	Clinically healthy
SRS009	Control genome	WGS batch 3	Clinically healthy
SRS011	Control genome	WGS batch 3	Clinically healthy
SRS012	Control genome	WGS batch 3	Clinically healthy
SRS014	Control genome	WGS batch 3	Clinically healthy
SRS015	Control genome	WGS batch 3	Clinically healthy
SRS017	Control genome	WGS batch 3	Clinically healthy
SRS018	Control genome	WGS batch 3	Clinically healthy
SRS020	Control genome	WGS batch 3	Clinically healthy
SRS021	Control genome	WGS batch 3	Clinically healthy
SRS023	Control genome	WGS batch 3	Clinically healthy
SRS024	Control genome	WGS batch 3	Clinically healthy
SRS026	Control genome	WGS batch 3	Clinically healthy
SRS027	Control genome	WGS batch 3	Clinically healthy
SRS029	Control genome	WGS batch 3	Clinically healthy
SRS030	Control genome	WGS batch 3	Clinically healthy
IBD001	Control genome	WGS batch 1	Very Early Onset Inflammatory Bowel Disease
IBD002	Control genome	WGS batch 1	Very Early Onset Inflammatory Bowel Disease
IBD003	Control genome	WGS batch 1	Very Early Onset Inflammatory Bowel Disease
IBD006	Control genome	WGS batch 1	Very Early Onset Inflammatory Bowel Disease
IBD007	Control genome	WGS batch 1	Very Early Onset Inflammatory Bowel Disease
IBD010	Control genome	WGS batch 1	Very Early Onset Inflammatory Bowel Disease
IBD011	Control genome	WGS batch 1	Very Early Onset Inflammatory Bowel Disease
IBD012	Control genome	WGS batch 3	Very Early Onset Inflammatory Bowel Disease
IBD016	Control genome	WGS batch 3	Very Early Onset Inflammatory Bowel Disease
IBD018	Control genome	WGS batch 3	Very Early Onset Inflammatory Bowel Disease
UCLG_502	Patient exome	WES batch 3	Bardet-Biedl Syndrome
UCLG_518	Patient exome	WES batch 3	Bardet-Biedl Syndrome
UCLG_520	Patient exome	WES batch 3	Bardet-Biedl Syndrome
UCLG_523	Patient exome	WES batch 3	Bardet-Biedl Syndrome

UCLG_508	Patient exome	WES batch 3	Bardet-Biedl Syndrome
UCLG_511	Patient exome	WES batch 3	Bardet-Biedl Syndrome
UCLG_512	Patient exome	WES batch 3	Bardet-Biedl Syndrome
UCLG_513	Patient exome	WES batch 3	Bardet-Biedl Syndrome
UCLG_514	Patient exome	WES batch 3	Bardet-Biedl Syndrome
UCLG_519	Patient exome	WES batch 3	Bardet-Biedl Syndrome
UCLG_521	Patient exome	WES batch 3	Bardet-Biedl Syndrome
UCLG_522	Patient exome	WES batch 3	Bardet-Biedl Syndrome
UCLG_524	Patient exome	WES batch 3	Bardet-Biedl Syndrome
UCLG_525	Patient exome	WES batch 3	Bardet-Biedl Syndrome
UCLG_506	Patient exome	WES batch 3	Bardet-Biedl Syndrome
UCLG_507	Patient exome	WES batch 3	Bardet-Biedl Syndrome
PE_1	Control exome	WES batch 1	Cleft lip and palate
PE_2	Control exome	WES batch 1	Cleft lip and palate
PE_3	Control exome	WES batch 1	Cleft lip and palate
PE_27	Control exome	WES batch 4	Severe Combined Immunodeficiency
PE_7	Control exome	WES batch 4	Spina Bifida Occulta
PE_8	Control exome	WES batch 4	Spina Bifida Occulta
PE_5	Control exome	WES batch 4	Spina Bifida Occulta
PE_4	Control exome	WES batch 4	Spina Bifida Occulta
PE_10	Control exome	WES batch 4	Spina Bifida Occulta
PE_11	Control exome	WES batch 4	Spina Bifida Occulta
PE_51	Control exome	WES batch 5	Congenital Melanocytic Naevus
PE_55	Control exome	WES batch 5	Severe Combined Immunodeficiency
PE_43	Control exome	WES batch 5	Congenital Melanocytic Naevus
PE_45	Control exome	WES batch 5	Multiple Congenital Anomalies
PE_40	Control exome	WES batch 5	Congenital Melanocytic Naevus
PE_42	Control exome	WES batch 5	Congenital Melanocytic Naevus
PE_36	Control exome	WES batch 5	Congenital Melanocytic Naevus
PE_38	Control exome	WES batch 5	Congenital Melanocytic Naevus
PE_37	Control exome	WES batch 5	Congenital Melanocytic Naevus
PE_37	Control exome	WES batch 5	Congenital Melanocytic Naevus
PE_49	Control exome	WES batch 5	Congenital Melanocytic Naevus
PE_53	Control exome	WES batch 5	Congenital Melanocytic Naevus
PE_52	Control exome	WES batch 5	Congenital Melanocytic Naevus
PE_54	Control exome	WES batch 5	Congenital Melanocytic Naevus
PE_41	Control exome	WES batch 5	Congenital Melanocytic Naevus

Table A.4 – Whole genome and whole exome sequencing samples and batches used in burden analysis, with disease information.

FGG	VTN	IGHG1	PROS1	C3	TF
FGA	F2	C1S	IGHG2	FN1	C1QA
FGB	CLU	APOE	CFH	SERPINA1	HBA1
ALB	PON1	ITIH4	IGHA1	HBB	PLG
APOA1	C4B	C1QC	C1QB	SERPINC1	C1R
C4A	F9	CRP	HP		

Table A.5 – Proteins detected by both SOMAscan and LC-MS/MS from plasma.

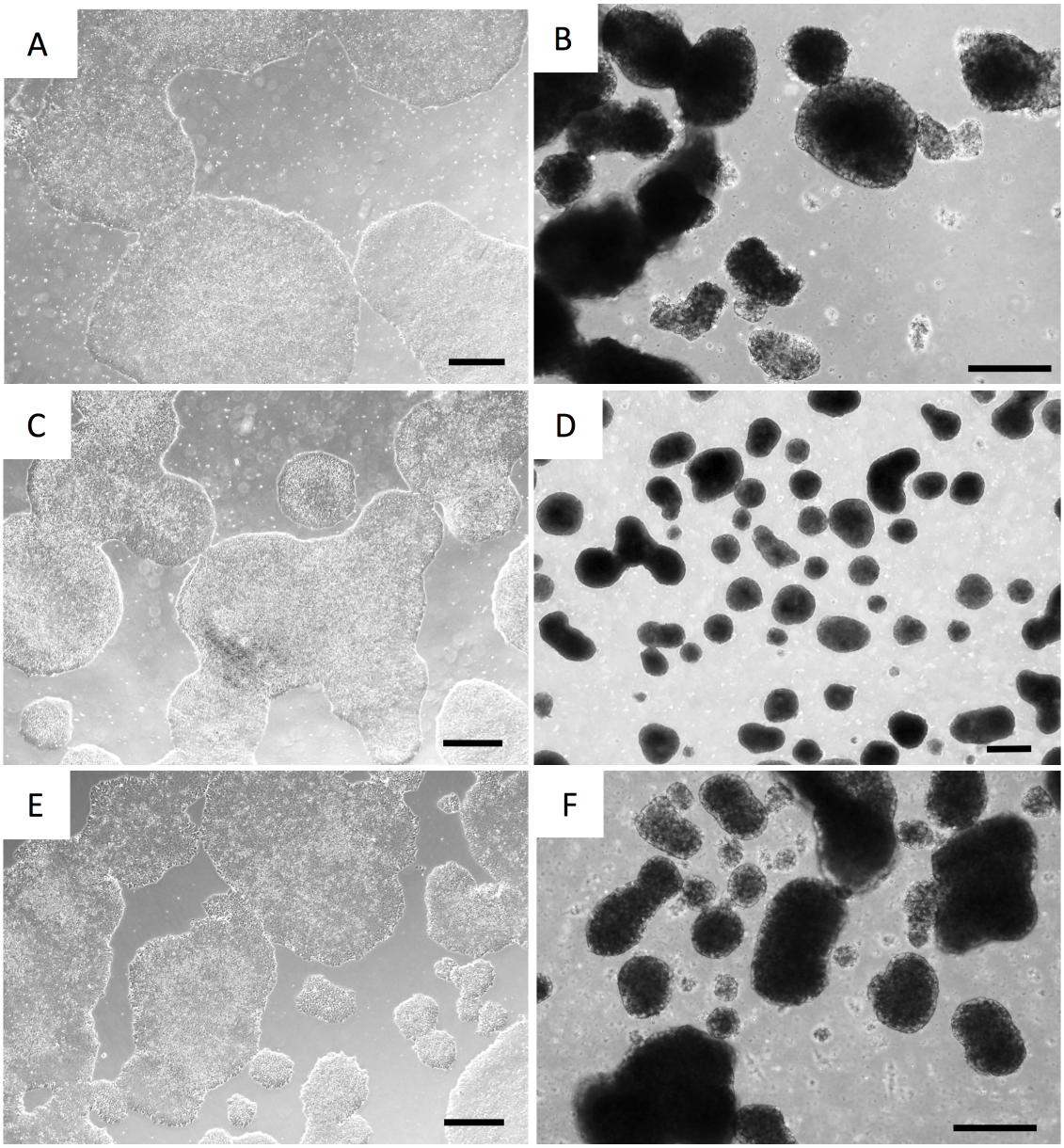

Figure A.1 – iPSC colonies and embryoid bodies from lines used in chapter 6. iPSC colonies (left images) and EBs at day 4 of differentiation (right images) are shown from BBS001 (A-B), BBS008 (C-D), and CTRL037 (E-F). Scale bars for A, C and E = 100µm, scale bars for B, D and F = 50µm.

Supplemental data

All supplemental data is available on the USB stick attached to the back page of this thesis, including raw files from transcriptomic and proteomic experiments. All files are outlined in **table S.1** below.

File ID	File name	Description	Location
Table S.1	Supplemental data contents	Contents of the supplemental data folders	Supplemental_data
Table S.2	RNA-seq_expression-matix-cpm.txt	Normalised gene expression data from RNA-seq, expressed as log2CPM	Supplemental_data>RNA-seq
Table S.3	RNA-seq_expression-matrix-rpkm.txt	Normalised gene expression data from RNA-seq, expressed as reads per kilobase million (RPKM)	Supplemental_data>RNA-seq
Table S.4	RNA-seq_edger-disease.txt	Master file of DE transcripts produced by the edge R programme	Supplemental_data>RNA-seq
Table S.5	RNA-seq_Fed.BBSvscontrol	Processed RNA-seq data with DE transcripts for fed BBS vs control	Supplemental_data>RNA-seq
Table S.6	RNA-seq_Starved.BBSvscontrol	Processed RNA-seq data with DE transcripts for starved BBS vs control	Supplemental_data>RNA-seq
Table S.7	RNA-seq_BBS.starvedvsfed	Processed RNA-seq data with DE transcripts for BBS starved vs fed	Supplemental_data>RNA-seq
Table S.8	RNA-seq_control.starvedvsfed	Processed RNA-seq data with DE transcripts for control starved vs fed	Supplemental_data>RNA-seq
Table S.9	iTRAQ_Hallmark_bbs.starvedvsfed.gsea_report_for_na_neg_enriched.xlsx	Hallmark gene sets negatively enriched in BBS starved vs fed from iTRAQ data	Supplemental_data>RNA-seq
Table S.10	iTRAQ_Hallmark_bbs.starvedvsfed.gsea_report_for_na_pos_enriched.xlsx	Hallmark gene sets positively enriched in BBS starved vs fed from iTRAQ data	Supplemental_data>RNA-seq
Table S.11	iTRAQ_Hallmark_control.starvedvsfed.gsea_report_for_na_neg_enriched.xlsx	Hallmark gene sets negatively enriched in control starved vs fed from iTRAQ data	Supplemental_data>RNA-seq
Table S.12	iTRAQ_Hallmark_control.starvedvsfed.gsea_report_for_na_pos_enriched.xlsx	Hallmark gene sets positively enriched in control starved vs fed from iTRAQ data	Supplemental_data>RNA-seq
Table S.13	iTRAQ_Hallmark_fed.bbsvscontrol.gsea_report_for_na_neg_enriched.xlsx	Hallmark gene sets negatively enriched in starved BBS vs control from iTRAQ data	Supplemental_data>RNA-seq

Table S.14	iTRAQ_Hallmark_fed.bbsvscontrol.gsea_report_for_na_pos_enriched.xlsx	Hallmark gene sets positively enriched in starved BBS vs control from iTRAQ data	Supplemental_data>RNA-seq
Table S.15	iTRAQ_Hallmark_starved.bbsvscontrol.gsea_report_for_na_neg_enriched.xlsx	Hallmark gene sets negatively enriched in fed BBS vs control from iTRAQ data	Supplemental_data>RNA-seq
Table S.16	iTRAQ_Hallmark_starved.bbsvscontrol.gsea_report_for_na_pos_enriched.xlsx	Hallmark gene sets positively enriched in fed BBS vs control from iTRAQ data	Supplemental_data>RNA-seq
Table S.17	RNA_Hallmark_BBS.starvedvsfed_gsea_report_for_neg_enriched.xls	Hallmark gene sets negatively enriched in BBS starved vs fed from RNA-seq data	Supplemental_data>RNA_GSEA
Table S.18	RNA_Hallmark_BBS.starvedvsfed_gsea_report_for_pos_enriched.xls	Hallmark gene sets positively enriched in BBS starved vs fed from RNA-seq data	Supplemental_data>RNA_GSEA
Table S.19	RNA_Hallmark_control.starvedvsfed_gsea_report_for_neg_enriched.xls	Hallmark gene sets negatively enriched in control starved vs fed from RNA-seq data	Supplemental_data>RNA_GSEA
Table S.20	RNA_Hallmark_control.starvedvsfed_gsea_report_for_pos_enriched.xls	Hallmark gene sets positively enriched in control starved vs fed from RNA-seq data	Supplemental_data>RNA_GSEA
Table S.21	RNA_Hallmark_starved.bbsvscontrol_gsea_report_for_neg_enriched.xls	Hallmark gene sets negatively enriched in starved BBS vs control from RNA-seq data	Supplemental_data>RNA_GSEA
Table S.22	RNA_Hallmark_starved.bbsvscontrol_gsea_report_for_pos_enriched.xls	Hallmark gene sets positively enriched in starved BBS vs control from RNA-seq data	Supplemental_data>RNA_GSEA
Table S.23	RNA_Hallmark_fed.bbsvscontrol_gsea_report_for_neg_enriched.xls	Hallmark gene sets negatively enriched in fed BBS vs control from RNA-seq data	Supplemental_data>RNA_GSEA
Table S.24	RNA_Hallmark_fed.bbsvscontrol_gsea_report_for_pos_enriched.xls	Hallmark gene sets positively enriched in fed BBS vs control from RNA-seq data	Supplemental_data>RNA_GSEA
Table S.25	BBS-limma-plasma_SOMA.txt	Limma file for SOMAscan plasma	Supplemental_data>Proteomics> SOMAscan
Table S.26	plasma_SOMA_raw.txt	Raw SOMAscan data from SOMAlogic	Supplemental_data>Proteomics> SOMAscan
Table S.27	SOMA_plasma_analysis	SOMAscan processed data, with RFU for each individual	Supplemental_data>Proteomics> SOMAscan
Table S.28	Plasma_label_free_MS	Data processed by Progenesis QIP from plasma SDS-PAGE fractionated label-free LC-MS/MS	Supplemental_data>Proteomics>Label-free_mass_spectrometry
Table S.29	Urine_label_free_MS	Data processed by Progenesis QIP from urine label-free LC-MS/MS	Supplemental_data>Proteomics>Label-free_mass_spectrometry
Table S.30	iTRAQ_BBS.starvedvsfed	Processed iTRAQ data with DE proteins for BBS starved vs fed	Supplemental_data>Proteomics> iTRAQ_mass_spectrometry
Table S.31	iTRAQ_control.starvedvsfed	Processed iTRAQ data with DE proteins for control starved vs fed	Supplemental_data>Proteomics> iTRAQ_mass_spectrometry

Table S.32	iTRAQ_fed.BBSvscontrol	Processed iTRAQ data with DE proteins for fed BBS vs control	Supplemental_data>Proteomics> iTRAQ_mass_spectrometry
Table S.33	iTRAQ_starved.BBSvscontrol	Processed iTRAQ data with DE proteins for starved BBS vs control	Supplemental_data>Proteomics> iTRAQ_mass_spectrometry
Table S.34	iTRAQ_Treated-raw_quant.txt	iTRAQ raw normalised quantitation data from BGI, for serum starved fibroblasts	Supplemental_data>Proteomics> iTRAQ_mass_spectrometry
Table S.35	iTRAQ_Untreated-raw_quant.txt	iTRAQ raw normalised quantitation data from BGI, for serum fed fibroblasts	Supplemental_data>Proteomics> iTRAQ_mass_spectrometry
Table S.36	iTRAQ_peptideSummary_treated.txt	iTRAQ peptide summary from serum starved fibroblasts	Supplemental_data>Proteomics> iTRAQ_mass_spectrometry
Table S.37	iTRAQ_proteinSummary_treated.txt	iTRAQ protein summary from serum starved fibroblasts	Supplemental_data>Proteomics> iTRAQ_mass_spectrometry
Table S.38	iTRAQ_peptideSummary_untreated.txt	iTRAQ peptide summary from serum fed fibroblasts	Supplemental_data>Proteomics> iTRAQ_mass_spectrometry
Table S.39	iTRAQ_proteinSummary_untreated.txt	iTRAQ protein summary from serum fed fibroblasts	Supplemental_data>Proteomics> iTRAQ_mass_spectrometry
Table S.40	Ingenuity_variant_count_vs_controls	Number of variants at each stage of the Ingenuity filtering cascade for WGS patient and control samples	Supplemental_data>Other
Table S.41	SOMAsan-Assay-protein-panel.pdf	List of 1310 proteins in the SOMAscan assay	Supplemental_data>Other
Table S.42	Urine_PANTHER_pathway_output	The output from PANTHER overrepresentation test from 726 urinary proteins discovered by HPLC-MS/MS	Supplemental_data>Proteomics>Label-free_mass_spectrometry
Movie S.1	IMG_1210.MOV	Recording of cardiomyocytes that differentiated in culture from CTRL038 iPSCs	Supplemental_data>Other

Table S.1 – Supplemental data and material available on the USB stick attached to this thesis.

Bibliography

1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., et al. (2015). A global reference for human genetic variation. Nature 526, 68–74.

Abd-El-Barr, M.M., Sykoudis, K., Andrabi, S., Eichers, E.R., Pennesi, M.E., Tan, P.L., Wilson, J.H., Katsanis, N., Lupski, J.R., and Wu, S.M. (2007). Impaired photoreceptor protein transport and synaptic transmission in a mouse model of Bardet-Biedl syndrome. Vision Res 47, 3394–3407.

Abdi, H., and Williams, L.J. (2010). Principal component analysis. WIREs Comp Stat 2, 433–459.

Aberle, H., Bauer, A., Stappert, J., Kispert, A., and Kemler, R. (1997). beta-catenin is a target for the ubiquitin-proteasome pathway. EMBO J 16, 3797–3804.

Abyzov, A., Urban, A.E., Synder, M., Gerstein, M. (2011). CNVnator: an approach to discover, genotype, and characterise typical and atypical CNVs from family and population genome sequencing. Genome Res 21, 974-984.

Adams, J. (2003). The proteasome: structure, function, and role in the cell. Cancer Treat Rev 29 Suppl 1, 3–9.

Adzhubei, I., Jordan, D.M., and Sunyaev, S.R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet Chapter 7, Unit7.20.

Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. Nat Methods 7, 248–249.

Aggarwal, S., and Yadav, A.K. (2016). Dissecting the iTRAQ Data Analysis. Methods Mol Biol 1362, 277–291.

Aguet, F., Brown, A.A., Castel, S., Davis, J.R., Mohammadi, P., Segre, A.V., Zappala, Z., Abell, N.S., Fresard, L., Gamazon, E.R., et al. (2016). Local genetic effects on gene expression across 44 human tissues. BioRxiv.

Aldahmesh, M.A., Li, Y., Alhashem, A., Anazi, S., Alkuraya, H., Hashem, M., Awaji, A.A., Sogaty, S., Alkharashi, A., Alzahrani, S., et al. (2014). IFT27, encoding a small GTPase component of IFT particles, is mutated in a consanguineous family with Bardet-Biedl syndrome. Hum Mol Genet 23, 3307–3315.

Ali, A.T., Hochfeld, W.E., Myburgh, R., and Pepper, M.S. (2013). Adipocyte and adipogenesis. Eur J Cell Biol 92, 229–236.

Alshammari, T.M., Al-Hassan, A.A., Hadda, T.B., and Aljofan, M. (2015). Comparison of different serum sample extraction methods and their suitability for mass spectrometry analysis. Saudi Pharm J 23, 689–697.

Amador-Arjona, A., Elliott, J., Miller, A., Ginbey, A., Pazour, G.J., Enikolopov, G., Roberts, A.J., and Terskikh, A.V. (2011). Primary cilia regulate proliferation of amplifying progenitors in adult hippocampus: implications for learning and memory. J Neurosci 31, 9933–9944.

Anderson, N.L., and Anderson, N.G. (2002). The human plasma proteome: history, character, and diagnostic prospects. Mol Cell Proteomics 1, 845–867.

Ansley, S.J., Badano, J.L., Blacque, O.E., Hill, J., Hoskins, B.E., Leitch, C.C., Kim, J.C., Ross, A.J., Eichers, E.R., Teslovich, T.M., et al. (2003). Basal body dysfunction is a likely cause of pleiotropic Bardet-Biedl syndrome. Nature 425, 628–633.

Ardrey, R.E. (2003). Liquid chromatography – mass spectrometry: an introduction (Chichester, UK: John Wiley & Sons, Ltd).

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25, 25–29.

Ashcroft, A. (1997). Ionization methods in organic mass spectrometry (Cambridge: Royal Society of Chemistry).

Ashley, E.A. (2016). Towards precision medicine. Nat Rev Genet 17, 507–522.

Ayyagari, R., Demirci, F.Y., Liu, J., Bingham, E.L., Stringham, H., Kakuk, L.E., Boehnke, M., Gorin, M.B., Richards, J.E., and Sieving, P.A. (2002). X-linked recessive atrophic macular degeneration from RPGR mutation. Genomics 80, 166–171.

Azaiez, H., Chamberlin, G.P., Fischer, S.M., Welp, C.L., Prasad, S.D., Taggart, R.T., del Castillo, I., Van Camp, G., and Smith, R.J.H. (2004). GJB2: the spectrum of deafness-causing allele variants and their phenotype. Hum Mutat 24, 305–311.

Azari, A.A., Aleman, T.S., Cideciyan, A.V., Schwartz, S.B., Windsor, E.A.M., Sumaroka, A., Cheung, A.Y., Steinberg, J.D., Roman, A.J., Stone, E.M., et al. (2006). Retinal disease expression in Bardet-Biedl syndrome-1 (BBS1) is a spectrum from maculopathy to retina-wide degeneration. Invest Ophthalmol Vis Sci 47, 5004–5010.

Azie, N., and Vincent, J. (2012). Rare diseases: the bane of modern society and the quest for cures. Clin Pharmacol Ther 92, 135–139.

Azuma, N., Yamaguchi, Y., Handa, H., Hayakawa, M., Kanai, A., and Yamada, M. (1999). Missense mutation in the alternative splice region of the PAX6 gene in eye anomalies. Am J Hum Genet 65, 656–663.

Badano, J.L., Ansley, S.J., Leitch, C.C., Lewis, R.A., Lupski, J.R., and Katsanis, N. (2003). Identification of a novel Bardet-Biedl syndrome protein, BBS7, that shares structural features with BBS1 and BBS2. Am J Hum Genet 72, 650–658.

Badano, J.L., Leitch, C.C., Ansley, S.J., May-Simera, H., Lawson, S., Lewis, R.A., Beales, P.L., Dietz, H.C., Fisher, S., and Katsanis, N. (2006). Dissection of epistasis in oligogenic Bardet-Biedl syndrome. Nature 439, 326–330.

Baker, K., Northam, G.B., Chong, W.K., Banks, T., Beales, P., and Baldeweg, T. (2011). Neocortical and hippocampal volume loss in a human ciliopathy: A quantitative MRI study in Bardet-Biedl syndrome. Am J Med Genet A 155A, 1–8.

van Balkom, B.W.M., Pisitkun, T., Verhaar, M.C., and Knepper, M.A. (2011). Exosomes and the kidney: prospects for diagnosis and therapy of renal diseases. Kidney Int 80, 1138–1145.

Barbelanne, M., Song, J., Ahmadzai, M., and Tsang, W.Y. (2013). Pathogenic NPHP5 mutations impair protein interaction with Cep290, a prerequisite for ciliogenesis. Hum Mol Genet 22, 2482–2494.

Barbelanne, M., Hossain, D., Chan, D.P., Peränen, J., and Tsang, W.Y. (2015). Nephrocystin proteins NPHP5 and Cep290 regulate BBSome integrity, ciliary trafficking and cargo delivery. Hum Mol Genet 24, 2185–2200.

Barber, R.D., Harmer, D.W., Coleman, R.A., and Clark, B.J. (2005). GAPDH as a housekeeping gene: analysis of GAPDH mRNA expression in a panel of 72 human tissues. Physiol Genomics 21, 389–395.

Bardet, G. (1920). Sur un syndrome d'obésité infantile avec polydactylie et rétinite pigmentaire (contribution à l'étude des formes cliniques de l'obésité hypophysaire). Doctoral dissertation. Université de Paris.

Barrero, M.J., Berdasco, M., Paramonov, I., Bilic, J., Vitaloni, M., Esteller, M., and Izpisua Belmonte, J.C. (2012). DNA hypermethylation in somatic cells correlates with higher reprogramming efficiency. Stem Cells 30, 1696–1702.

Basten, S.G., and Giles, R.H. (2013). Functional aspects of primary cilia in signaling, cell cycle and tumorigenesis. Cilia 2, 6.

Basto, R., Lau, J., Vinogradova, T., Gardiol, A., Woods, C.G., Khodjakov, A., and Raff, J.W. (2006). Flies without centrioles. Cell 125, 1375–1386.

Bauman, M.L., and Hogan, G.R. (1973). Laurence-Moon-Biedl syndrome. Am J Dis Child 119.

Beales, P.L., Elcioglu, N., Woolf, A.S., Parker, D., and Flinter, F.A. (1999). New criteria for improved diagnosis of Bardet-Biedl syndrome: results of a population survey. J Med Genet 36, 437–446.

Beales, P.L., Badano, J.L., Ross, A.J., Ansley, S.J., Hoskins, B.E., Kirsten, B., Mein, C.A., Froguel, P., Scambler, P.J., Lewis, R.A., et al. (2003). Genetic interaction of BBS1 mutations with alleles at other BBS loci can result in non-Mendelian Bardet-Biedl syndrome. Am J Hum Genet 72, 1187–1199.

Berbari, N.F., Lewis, J.S., Bishop, G.A., Askwith, C.C., and Mykytyn, K. (2008). Bardet-Biedl syndrome proteins are required for the localization of G protein-coupled receptors to primary cilia. Proc Natl Acad Sci U S A 105, 4242–4246.

Besharse, J.C., Hollyfield, J.G., and Rayborn, M.E. (1977). Turnover of rod photoreceptor outer segments. J Cell Biol 75, 507–527.

Besharse, J.C., Forestner, D.M., and Defoe, D.M. (1985). Membrane assembly in retinal photoreceptors. III. Distinct membrane domains of the connecting cilium of developing rods. J Neurosci 5, 1035–1048.

Biedl, A. (1922). Ein Geschwisterpaar mit adiposo-genitaler Dystrophie. Deutsche Medicinische Wochenschrift 1630.

Biomarkers Definitions Working Group. (2001). Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther 69, 89–95.

Bizet, A.A., Becker-Heck, A., Ryan, R., Weber, K., Filhol, E., Krug, P., Halbritter, J., Delous, M., Lasbennes, M.-C., Linghu, B., et al. (2015). Mutations in TRAF3IP1/IFT54 reveal a new role for IFT proteins in microtubule stabilization. Nat Commun 6, 8666.

Blacque, O.E., Reardon, M.J., Li, C., McCarthy, J., Mahjoub, M.R., Ansley, S.J., Badano, J.L., Mah, A.K., Beales, P.L., Davidson, W.S., et al. (2004). Loss of C. elegans BBS-7 and BBS-8 protein function results in cilia defects and compromised intraflagellar transport. Genes Dev 18, 1630–1642.

Boldt, K., van Reeuwijk, J., Lu, Q., Koutroumpas, K., Nguyen, T.-M.T., Texier, Y., van Beersum, S.E.C., Horn, N., Willer, J.R., Mans, D.A., et al. (2016). An organelle-specific protein landscape identifies novel diseases and molecular mechanisms. Nat Commun 7, 11491.

Boubakri, M., Chaya, T., Hirata, H., Kajimura, N., Kuwahara, R., Ueno, A., Malicki, J., Furukawa, T., and Omori, Y. (2016). Loss of ift122, a Retrograde Intraflagellar Transport (IFT) Complex Component, Leads to Slow, Progressive Photoreceptor Degeneration Due to Inefficient Opsin Transport. J Biol Chem 291, 24465–24474.

Boucher, C., and Sandford, R. (2004). Autosomal dominant polycystic kidney disease (ADPKD, MIM 173900, PKD1 and PKD2 genes, protein products known as polycystin-1 and polycystin-2). Eur J Hum Genet 12, 347–354.

Boucherie, C., Mukherjee, S., Henckaerts, E., Thrasher, A.J., Sowden, J.C., and Ali, R.R. (2013). Brief report: self-organizing neuroepithelium from human pluripotent stem cells facilitates derivation of photoreceptors. Stem Cells 31, 408–414.

Boycott, K.M., Vanstone, M.R., Bulman, D.E., and MacKenzie, A.E. (2013). Rare-disease genetics in the era of next-generation sequencing: discovery to translation. Nat Rev Genet 14, 681–691.

Boycott, K.M., Rath, A., Chong, J.X., Hartley, T., Alkuraya, F.S., Baynam, G., Brookes, A.J., Brudno, M., Carracedo, A., den Dunnen, J.T., et al. (2017). International cooperation to enable the diagnosis of all rare genetic diseases. Am J Hum Genet 100, 695–705.

Boyle, M.P. (2007). Strategies for identifying modifier genes in cystic fibrosis. Proc Am Thorac Soc 4, 52–57.

Brancati, F., Barrano, G., Silhavy, J.L., Marsh, S.E., Travaglini, L., Bielas, S.L., Amorini, M., Zablocka, D., Kayserili, H., Al-Gazali, L., et al. (2007). CEP290 mutations are frequently identified in the oculo-renal form of Joubert syndrome-related disorders. Am J Hum Genet 81, 104–113.

Brodlie, M., Haq, I.J., Roberts, K., and Elborn, J.S. (2015). Targeted therapies to improve CFTR function in cystic fibrosis. Genome Med 7, 101.

Brody, E.N., and Gold, L. (2000). Aptamers as therapeutic and diagnostic agents. Reviews in Molecular Biotechnology 74, 5–13.

Brown, N.J., Marjanović, M., Lüders, J., Stracker, T.H., and Costanzo, V. (2013). Cep63 and cep152 cooperate to ensure centriole duplication. PLoS ONE 8, e69986.

Čajánek, L., and Nigg, E.A. (2014). Cep164 triggers ciliogenesis by recruiting Tau tubulin kinase 2 to the mother centriole. Proc Natl Acad Sci U S A 111, E2841-50.

Cannon, P.S., Clayton-Smith, J., Beales, P.L., and Lloyd, I.C. (2008). Bardet-biedl syndrome: an atypical phenotype in brothers with a proven BBS1 mutation. Ophthalmic Genet 29, 128–132.

Cao, J., Shen, Y., Zhu, L., Xu, Y., Zhou, Y., Wu, Z., Li, Y., Yan, X., and Zhu, X. (2012). miR-129-3p controls cilia assembly by regulating CP110 and actin dynamics. Nat Cell Biol 14, 697–706.

Capowski, E.E., Simonett, J.M., Clark, E.M., Wright, L.S., Howden, S.E., Wallace, K.A., Petelinsek, A.M., Pinilla, I., Phillips, M.J., Meyer, J.S., et al. (2014). Loss of MITF expression during human embryonic stem cell differentiation disrupts retinal pigment epithelium development and optic vesicle cell proliferation. Hum Mol Genet 23, 6332–6344.

Cardenas-Rodriguez, M., Osborn, D.P.S., Irigoín, F., Graña, M., Romero, H., Beales, P.L., and Badano, J.L. (2013). Characterization of CCDC28B reveals its role in ciliogenesis and provides insight to understand its modifier effect on Bardet-Biedl syndrome. Hum Genet 132, 91–105.

Carmi, R., Elbedour, K., Stone, E.M., and Sheffield, V.C. (1995). Phenotypic differences among patients with Bardet-Biedl syndrome linked to three different chromosome loci. Am J Med Genet 59, 199–203.

Caulfield, M., Davies, J., Dennys, M., Elbahy, L., Fowler, T., Hill, S., Hubbard, T., Jostins, L., Maltby, N., Mahon-Pearson, J., et al. (2017). The 100,000 Genomes Project Protocol. Figshare.

Cayouette, M., Smith, S.B., Becerra, S.P., and Gravel, C. (1999). Pigment epithelium-derived factor delays the death of photoreceptors in mouse models of inherited retinal degenerations. Neurobiol Dis 6, 523–532.

Chaki, M., Airik, R., Ghosh, A.K., Giles, R.H., Chen, R., Slaats, G.G., Wang, H., Hurd, T.W., Zhou, W., Cluckey, A., et al. (2012). Exome capture reveals ZNF423 and CEP164 mutations, linking renal ciliopathies to DNA damage response signaling. Cell 150, 533–548.

Challa, T.D., Straub, L.G., Balaz, M., Kiehlmann, E., Donze, O., Rudofsky, G., Ukropec, J., Ukropcova, B., and Wolfrum, C. (2015). Regulation of de novo adipocyte differentiation through cross talk between adipocytes and preadipocytes. Diabetes 64, 4075–4087.

Chamling, X., Seo, S., Searby, C.C., Kim, G., Slusarski, D.C., and Sheffield, V.C. (2014). The centriolar satellite protein AZI1 interacts with BBS4 and regulates ciliary trafficking of the BBSome. PLoS Genet 10, e1004083.

Chavan, S.S., Hudson, L.K., Li, J.H., Ochani, M., Harris, Y., Patel, N.B., Katz, D., Scheinerman, J.A., Pavlov, V.A., and Tracey, K.J. (2012). Identification of pigment epithelium-derived factor as an adipocyte-derived inflammatory factor. Mol Med 18, 1161–1168.

Chen, G., Gulbranson, D.R., Hou, Z., Bolin, J.M., Ruotti, V., Probasco, M.D., Smuga-Otto, K., Howden, S.E., Diol, N.R., Propson, N.E., et al. (2011). Chemically defined conditions for human iPSC derivation and culture. Nat Methods 8, 424–429.

Chen, R., Mias, G.I., Li-Pook-Than, J., Jiang, L., Lam, H.Y.K., Chen, R., Miriami, E., Karczewski, K.J., Hariharan, M., Dewey, F.E., et al. (2012). Personal omics profiling reveals dynamic molecular and medical phenotypes. Cell 148, 1293–1307.

Chiang, A.P., Beck, J.S., Yen, H.-J., Tayeh, M.K., Scheetz, T.E., Swiderski, R.E., Nishimura, D.Y., Braun, T.A., Kim, K.-Y.A., Huang, J., et al. (2006). Homozygosity mapping with SNP arrays identifies TRIM32, an E3 ubiquitin ligase, as a Bardet-Biedl syndrome gene (BBS11). Proc Natl Acad Sci U S A 103, 6287–6292.

Choi, Y., and Chan, A.P. (2015). PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics 31, 2745–2747.

Choi, Y., Sims, G.E., Murphy, S., Miller, J.R., and Chan, A.P. (2012). Predicting the functional effect of amino acid substitutions and indels. PLoS ONE 7, e46688.

Christensen, S.T., Pedersen, S.F., Satir, P., Veland, I.R., and Schneider, L. (2008). The primary cilium coordinates signaling pathways in cell cycle control and migration during development and tissue repair. Curr Top Dev Biol 85, 261–301.

Christensen, S.T., Veland, I.R., Schwab, A., Cammer, M., and Satir, P. (2013). Analysis of primary cilia in directional cell migration in fibroblasts. Meth Enzymol 525, 45–58.

Christiansen, M.S., Hommel, E., Magid, E., and Feldt-Rasmussen, B. (2002). Orosomucoid in urine predicts cardiovascular and over-all mortality in patients with Type II diabetes. Diabetologia 45, 115–120.

Chuang, J.-Z., Zhao, Y., and Sung, C.-H. (2007). SARA-regulated vesicular targeting underlies formation of the light-sensing organelle in mammalian rods. Cell 130, 535–547.

Chung, D.C., and Traboulsi, E.I. (2009). Leber congenital amaurosis: clinical correlations with genotypes, gene therapy trials update, and future directions. J AAPOS 13, 587–592.

Churchill, D.N., McManamon, P., and Hurley, R.M. (1981). Renal disease-a sixth cardinal feature of the Laurence-Moon-Biedl syndrome. Clin Nephrol 16, 151–154.

Cizmecioglu, O., Arnold, M., Bahtz, R., Settele, F., Ehret, L., Haselmann-Weiss, U., Antony, C., and Hoffmann, I. (2010). Cep152 acts as a scaffold for recruitment of Plk4 and CPAP to the centrosome. J Cell Biol 191, 731–739.

Coene, K.L.M., Roepman, R., Doherty, D., Afroze, B., Kroes, H.Y., Letteboer, S.J.F., Ngu, L.H., Budny, B., van Wijk, E., Gorden, N.T., et al. (2009). OFD1 is mutated in X-linked Joubert syndrome and interacts with LCA5-encoded lebercilin. Am J Hum Genet 85, 465–481.

Collin, G.B., Marshall, J.D., Ikeda, A., So, W.V., Russell-Eggitt, I., Maffei, P., Beck, S., Boerkoel, C.F., Sicolo, N., Martin, M., et al. (2002). Mutations in ALMS1 cause obesity, type 2 diabetes and neurosensory degeneration in Alström syndrome. Nat Genet 31, 74–78.

Collin, R.W.J., Safieh, C., Littink, K.W., Shalev, S.A., Garzozi, H.J., Rizel, L., Abbasi, A.H., Cremers, F.P.M., den Hollander, A.I., Klevering, B.J., et al. (2010). Mutations in C2ORF71 cause autosomal-recessive retinitis pigmentosa. Am J Hum Genet 86, 783–788.

Conrad, D.F., Bird, C., Blackburne, B., Lindsay, S., Mamanova, L., Lee, C., Turner, D.J., and Hurles, M.E. (2010). Mutation spectrum revealed by breakpoint sequencing of human germline CNVs. Nat Genet 42, 385–391.

Coon, B.G., Hernandez, V., Madhivanan, K., Mukherjee, D., Hanna, C.B., Barinaga-Rementeria Ramirez, I., Lowe, M., Beales, P.L., and Aguilar, R.C. (2012). The Lowe syndrome protein OCRL1 is involved in primary cilia assembly. Hum Mol Genet 21, 1835–1847.

Coppieters, F., Lefever, S., Leroy, B.P., and De Baere, E. (2010a). CEP290, a gene with many faces: mutation overview and presentation of CEP290base. Hum Mutat 31, 1097–1108.

Coppieters, F., Casteels, I., Meire, F., De Jaegere, S., Hooghe, S., van Regemorter, N., Van Esch, H., Matuleviciene, A., Nunes, L., Meersschaut, V., et al. (2010b). Genetic screening of LCA in Belgium: predominance of CEP290 and identification of potential modifier alleles in AHI1 of CEP290-related phenotypes. Hum Mutat 31, E1709-66.

Cosgrove, D., and Zallocchi, M. (2014). Usher protein functions in hair cells and photoreceptors. Int J Biochem Cell Biol 46, 80–89.

Cox, K.F., Kerr, N.C., Kedrov, M., Nishimura, D., Jennings, B.J., Stone, E.M., Sheffield, V.C., and Iannaccone, A. (2012). Phenotypic expression of Bardet-Biedl syndrome in patients homozygous for the common M390R mutation in the BBS1 gene. Vision Res 75, 77–87.

Cravatt, B.F., Simon, G.M., and Yates, J.R. (2007). The biological impact of mass-spectrometry-based proteomics. Nature 450, 991–1000.

Croft, D., O'Kelly, G., Wu, G., Haw, R., Gillespie, M., Matthews, L., Caudy, M., Garapati, P., Gopinath, G., Jassal, B., et al. (2011). Reactome: a database of reactions, pathways and biological processes. Nucleic Acids Res 39, D691-7.

Croft, J.B., Morrell, D., Chase, C.L., and Swift, M. (1995). Obesity in heterozygous carriers of the gene for the Bardet-Biedl syndrome. Am J Med Genet 55, 12–15.

Crowe, S., Wu, L.E., Economou, C., Turpin, S.M., Matzaris, M., Hoehn, K.L., Hevener, A.L., James, D.E., Duh, E.J., and Watt, M.J. (2009). Pigment epithelium-derived factor contributes to insulin resistance in obesity. Cell Metab 10, 40–47.

Crutchfield, C.A., Thomas, S.N., Sokoll, L.J., and Chan, D.W. (2016). Advances in mass spectrometry-based clinical biomarker discovery. Clin Proteomics 13, 1.

Cui, C., Chatterjee, B., Lozito, T.P., Zhang, Z., Francis, R.J., Yagi, H., Swanhart, L.M., Sanker, S., Francis, D., Yu, Q., et al. (2013). Wdpcp, a PCP protein required for ciliogenesis, regulates directional cell migration and cell polarity by direct modulation of the actin cytoskeleton. PLoS Biol 11, e1001720.

Cummings, B.B., Marshall, J.L., Tukiainen, T., Lek, M., Donkervoort, S., Foley, A.R., Bolduc, V., Waddell, L., Sandaradura, S., O'Grady, G.L., et al. (2016). Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. BioRxiv.

Cummings, B.B., Marshall, J.L., Tukiainen, T., Lek, M., Donkervoort, S., Foley, A.R., Bolduc, V., Waddell, L.B., Sandaradura, S.A., O'Grady, G.L., et al. (2017). Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. Sci Transl Med 9.

Cutting, G.R. (2010). Modifier genes in Mendelian disorders: the example of cystic fibrosis. Ann N Y Acad Sci 1214, 57–69.

Czyz, W., Morahan, J.M., Ebers, G.C., and Ramagopalan, S.V. (2012). Genetic, environmental and stochastic factors in monozygotic twin discordance with a focus on epigenetic differences. BMC Med 10, 93.

Daiger, S.P., Rossiter, B.J.F., Greenberg, J., Christoffels, A., and Hide, W. (1998). Data services and software for identifying genes and mutations causing retinal degeneration.

D'Alessandro, L.C.A., Al Turki, S., Manickaraj, A.K., Manase, D., Mulder, B.J.M., Bergin, L., Rosenberg, H.C., Mondal, T., Gordon, E., Lougheed, J., et al. (2016). Exome sequencing identifies rare variants in multiple genes in atrioventricular septal defect. Genet Med 18, 189–198.

van Dam, T.J., Wheway, G., Slaats, G.G., SYSCILIA Study Group, Huynen, M.A., and Giles, R.H. (2013). The SYSCILIA gold standard (SCGSv1) of known ciliary components and its applications within a systems biology consortium. Cilia 2, 7.

van Dam, T.J.P., Kennedy, J., van der Lee, R., de Vrieze, E., Wunderlich, K.A., Rix, S., Dougherty, G.W., Lambacher, N.J., Li, C., Jensen, V.L., et al. (2017). Ciliacarta: an integrated and validated compendium of ciliary genes. BioRxiv.

Daniels, A.B., Sandberg, M.A., Chen, J., Weigel-DiFranco, C., Fielding Hejtmancic, J., and Berson, E.L. (2012). Genotype-phenotype correlations in Bardet-Biedl syndrome. Arch Ophthalmol 130, 901–907.

Danno, H., Michiue, T., Hitachi, K., Yukita, A., Ishiura, S., and Asashima, M. (2008). Molecular links among the causative genes for ocular malformation: Otx2 and Sox2 coregulate Rax expression. Proc Natl Acad Sci U S A 105, 5408–5413.

Datta, P., Allamargot, C., Hudson, J.S., Andersen, E.K., Bhattarai, S., Drack, A.V., Sheffield, V.C., and Seo, S. (2015). Accumulation of non-outer segment proteins in the outer segment underlies photoreceptor degeneration in Bardet-Biedl syndrome. Proc Natl Acad Sci U S A 112, E4400-9.

Davis, E.E., and Katsanis, N. (2012). The ciliopathies: a transitional model into systems biology of human genetic disease. Curr Opin Genet Dev 22, 290–303.

Davis, E.E., Zhang, Q., Liu, Q., Diplas, B.H., Davey, L.M., Hartley, J., Stoetzel, C., Szymanska, K., Ramaswami, G., Logan, C.V., et al. (2011). TTC21B contributes both causal and modifying alleles across the ciliopathy spectrum. Nat Genet 43, 189–196.

Davis, R.E., Swiderski, R.E., Rahmouni, K., Nishimura, D.Y., Mullins, R.F., Agassandian, K., Philp, A.R., Searby, C.C., Andrews, M.P., Thompson, S., et al. (2007). A knockin mouse model of the Bardet-Biedl syndrome 1 M390R mutation has cilia defects, ventriculomegaly, retinopathy, and obesity. Proc Natl Acad Sci U S A 104, 19422–19427.

Dawson, D.W., Volpert, O.V., Gillis, P., Crawford, S.E., Xu, H., Benedict, W., and Bouck, N.P. (1999). Pigment epithelium-derived factor: a potent inhibitor of angiogenesis. Science 285, 245–248.

Dear, J.W., Street, J.M., and Bailey, M.A. (2013). Urinary exosomes: a reservoir for biomarker discovery and potential mediators of intrarenal signalling. Proteomics 13, 1572–1580.

Delous, M., Baala, L., Salomon, R., Laclef, C., Vierkotten, J., Tory, K., Golzio, C., Lacoste, T., Besse, L., Ozilou, C., et al. (2007). The ciliary gene RPGRIP1L is mutated in cerebello-oculo-renal syndrome (Joubert syndrome type B) and Meckel syndrome. Nat Genet 39, 875–881.

Deveault, C., Billingsley, G., Duncan, J.L., Bin, J., Theal, R., Vincent, A., Fieggen, K.J., Gerth, C., Noordeh, N., Traboulsi, E.I., et al. (2011). BBS genotype-phenotype assessment of a multiethnic patient cohort calls for a revision of the disease definition. Hum Mutat 32, 610–619.

Dewey, F.E., Grove, M.E., Pan, C., Goldstein, B.A., Bernstein, J.A., Chaib, H., Merker, J.D., Goldfeder, R.L., Enns, G.M., David, S.P., et al. (2014). Clinical interpretation and implications of whole-genome sequencing. JAMA 311, 1035–1045.

Di Marcotullio, L., Ferretti, E., Greco, A., De Smaele, E., Po, A., Sico, M.A., Alimandi, M., Giannini, G., Maroder, M., Screpanti, I., et al. (2006). Numb is a suppressor of Hedgehog signalling and targets Gli1 for Itch-dependent ubiquitination. Nat Cell Biol 8, 1415–1423.

Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., et al. (2012). Landscape of transcription in human cells. Nature 489, 101–108.

Domire, J.S., Green, J.A., Lee, K.G., Johnson, A.D., Askwith, C.C., and Mykytyn, K. (2011). Dopamine receptor 1 localizes to neuronal cilia in a dynamic process that requires the Bardet-Biedl syndrome proteins. Cell Mol Life Sci 68, 2951–2960.

Drivas, T.G., Wojno, A.P., Tucker, B.A., Stone, E.M., and Bennett, J. (2015). Basal exon skipping and genetic pleiotropy: A predictive model of disease pathogenesis. Sci Transl Med 7, 291ra97.

Dullaart, R.P.F., Plomgaard, P., de Vries, R., Dahlbäck, B., and Nielsen, L.B. (2009). Plasma apolipoprotein M is reduced in metabolic syndrome but does not predict intima media thickness. Clin Chim Acta 406, 129–133.

Dummer, A., Poelma, C., DeRuiter, M.C., Goumans, M.-J.T.H., and Hierck, B.P. (2016). Measuring the primary cilium length: improved method for unbiased high-throughput analysis. Cilia 5, 7.

Dunn, W.B., Broadhurst, D., Begley, P., Zelena, E., Francis-McIntyre, S., Anderson, N., Brown, M., Knowles, J.D., Halsall, A., Haselden, J.N., et al. (2011). Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. Nat Protoc 6, 1060–1083.

Dutta, N., and Seo, S. (2016). RPGR, a prenylated retinal ciliopathy protein, is targeted to cilia in a prenylation- and PDE6D-dependent manner. Biol Open 5, 1283–1289.

Eggenschwiler, J.T., and Anderson, K.V. (2007). Cilia and developmental signaling. Annu Rev Cell Dev Biol 23, 345–373.

Eguether, T., San Agustin, J.T., Keady, B.T., Jonassen, J.A., Liang, Y., Francis, R., Tobita, K., Johnson, C.A., Abdelhamed, Z.A., Lo, C.W., et al. (2014). IFT27 links the BBSome to IFT for maintenance of the ciliary signaling compartment. Dev Cell 31, 279–290.

Eiraku, M., Takata, N., Ishibashi, H., Kawada, M., Sakakura, E., Okuda, S., Sekiguchi, K., Adachi, T., and Sasai, Y. (2011). Self-organizing optic-cup morphogenesis in three-dimensional culture. Nature 472, 51–56.

Elshal, M.F., and McCoy, J.P. (2006). Multiplex bead array assays: performance evaluation and comparison of sensitivity to ELISA. Methods 38, 317–323.

Estrada-Cuzcano, A., Koenekoop, R.K., Senechal, A., De Baere, E.B.W., de Ravel, T., Banfi, S., Kohl, S., Ayuso, C., Sharon, D., Hoyng, C.B., et al. (2012). BBS1 mutations in a wide spectrum of phenotypes ranging from nonsyndromic retinitis pigmentosa to Bardet-Biedl syndrome. Arch Ophthalmol 130, 1425–1432.

EURORDIS (2017). Rare Disease Day 2018 article - What is a rare disease?

Ezratty, E.J., Stokes, N., Chai, S., Shah, A.S., Williams, S.E., and Fuchs, E. (2011). A role for the primary cilium in Notch signaling and epidermal differentiation during skin development. Cell 145, 1129–1141.

Fabregat, A., Sidiropoulos, K., Garapati, P., Gillespie, M., Hausmann, K., Haw, R., Jassal, B., Jupe, S., Korninger, F., McKay, S., et al. (2016). The Reactome pathway Knowledgebase. Nucleic Acids Res 44, D481-7.

Famulla, S., Lamers, D., Hartwig, S., Passlack, W., Horrighs, A., Cramer, A., Lehr, S., Sell, H., and Eckel, J. (2011). Pigment epithelium-derived factor (PEDF) is one of the most abundant proteins secreted by human adipocytes and induces insulin resistance and inflammatory signaling in muscle and fat cells. Int J Obes (Lond) 35, 762–772.

Fan, Y., Esmail, M.A., Ansley, S.J., Blacque, O.E., Boroevich, K., Ross, A.J., Moore, S.J., Badano, J.L., May-Simera, H., Compton, D.S., et al. (2004). Mutations in a member of the Ras superfamily of small GTP-binding proteins causes Bardet-Biedl syndrome. Nat Genet 36, 989–993.

Fang, D.Y., King, H.W., Li, J.Y., and Gleadle, J.M. (2013). Exosomes and the kidney: blaming the messenger. Nephrology (Carlton) 18, 1–10.

Farag, T.I., and Teebi, A.S. (2008). High incidence of Bardet Biedl syndrome among the Bedouin. Clin Genet 36, 463–464.

Farmer, A., Aymé, S., de Heredia, M.L., Maffei, P., McCafferty, S., Młynarski, W., Nunes, V., Parkinson, K., Paquis-Flucklinger, V., Rohayem, J., et al. (2013). EURO-WABB: an EU rare diseases registry for Wolfram syndrome, Alström syndrome and Bardet-Biedl syndrome. BMC Pediatr 13, 130.

Feng, S., Knödler, A., Ren, J., Zhang, J., Zhang, X., Hong, Y., Huang, S., Peränen, J., and Guo, W. (2012). A Rab8 guanine nucleotide exchange factor-effector interaction network regulates primary ciliogenesis. J Biol Chem 287, 15602–15609.

Festuccia, W.T., Blanchard, P.-G., and Deshaies, Y. (2011). Control of brown adipose tissue glucose and lipid metabolism by pparγ. Front Endocrinol (Lausanne) 2, 84.

Feuillan, P.P., Ng, D., Han, J.C., Sapp, J.C., Wetsch, K., Spaulding, E., Zheng, Y.C., Caruso, R.C., Brooks, B.P., Johnston, J.J., et al. (2011). Patients with Bardet-Biedl syndrome have hyperleptinemia suggestive of leptin resistance. J Clin Endocrinol Metab 96, E528-35.

Finnemann, S.C., Bonilha, V.L., Marmorstein, A.D., and Rodriguez-Boulan, E. (1997). Phagocytosis of rod outer segments by retinal pigment epithelial cells requires alpha(v)beta5 integrin for binding but not for internalization. Proc Natl Acad Sci U S A 94, 12932–12937.

Forsythe, E., and Beales, P.L. (2013). Bardet-Biedl syndrome. Eur J Hum Genet 21, 8–13.

Forsythe, E., Sparks, K., Hoskins, B.E., Bagkeris, E., McGowan, B.M., Carroll, P.V., Huda, M.S.B., Mujahid, S., Peters, C., Barrett, T., et al. (2015). Genetic predictors of cardiovascular morbidity in Bardet-Biedl syndrome. Clin Genet 87, 343–349.

Fortunel, N.O., Otu, H.H., Ng, H.-H., Chen, J., Mu, X., Chevassut, T., Li, X., Joseph, M., Bailey, C., Hatzfeld, J.A., et al. (2003). Comment on " 'Stemness': transcriptional profiling of embryonic and adult stem cells" and "a stem cell molecular signature". Science 302, 393; author reply 393.

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

Freytag, S.O., Paielli, D.L., and Gilbert, J.D. (1994). Ectopic expression of the CCAAT/enhancer-binding protein alpha promotes the adipogenic program in a variety of mouse fibroblastic cells. Genes Dev 8, 1654–1663.

Friedenstein, A.J. (1976). Precursor cells of mechanocytes. Int Rev Cytol 47, 327–359.

Furukawa, T., Morrow, E.M., Li, T., Davis, F.C., and Cepko, C.L. (1999). Retinopathy and attenuated circadian entrainment in Crx-deficient mice. Nat Genet 23, 466–470.

Gabay, C., and Kushner, I. (1999). Acute-phase proteins and other systemic responses to inflammation. N Engl J Med 340, 448–454.

Gao, F., and Keinan, A. (2014). High burden of private mutations due to explosive human population growth and purifying selection. BMC Genomics 15 Suppl 4, S3.

Garcia-Gonzalo, F.R., Corbit, K.C., Sirerol-Piquer, M.S., Ramaswami, G., Otto, E.A., Noriega, T.R., Seol, A.D., Robinson, J.F., Bennett, C.L., Josifova, D.J., et al. (2011). A transition zone complex regulates mammalian ciliogenesis and ciliary membrane composition. Nat Genet 43, 776–784.

Gascue, C., Tan, P.L., Cardenas-Rodriguez, M., Libisch, G., Fernandez-Calero, T., Liu, Y.P., Astrada, S., Robello, C., Naya, H., Katsanis, N., et al. (2012). Direct role of Bardet-Biedl syndrome proteins in transcriptional regulation. J Cell Sci 125, 362–375.

Gazave, E., Chang, D., Clark, A.G., and Keinan, A. (2013). Population growth inflates the per-individual number of deleterious mutations and reduces their mean effect. Genetics 195, 969–978.

Ge, H., Walhout, A.J.M., and Vidal, M. (2003). Integrating "omic" information: a bridge between genomics and systems biology. Trends Genet 19, 551–560.

Geiger, T., Wehner, A., Schaab, C., Cox, J., and Mann, M. (2012). Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. Mol Cell Proteomics 11, M111.014050.

Gene Ontology Consortium (2015). Gene Ontology Consortium: going forward. Nucleic Acids Res 43, D1049-56.

Génin, E., Feingold, J., and Clerget-Darpoux, F. (2008). Identifying modifier genes of monogenic disease: strategies and difficulties. Hum Genet 124, 357–368.

Gerdes, J.M., Liu, Y., Zaghloul, N.A., Leitch, C.C., Lawson, S.S., Kato, M., Beachy, P.A., Beales, P.L., DeMartino, G.N., Fisher, S., et al. (2007). Disruption of the basal body compromises proteasomal function and perturbs intracellular Wnt response. Nat Genet 39, 1350–1360.

Gerdes, J.M., Davis, E.E., and Katsanis, N. (2009). The vertebrate primary cilium in development, homeostasis, and disease. Cell 137, 32–45.

Gergely, F., Draviam, V.M., and Raff, J.W. (2003). The ch-TOG/XMAP215 protein is essential for spindle pole organization in human somatic cells. Genes Dev 17, 336–341.

Gerhardt, C., Lier, J.M., Burmühl, S., Struchtrup, A., Deutschmann, K., Vetter, M., Leu, T., Reeg, S., Grune, T., and Rüther, U. (2015). The transition zone protein Rpgrip1l regulates proteasomal activity at the primary cilium. J Cell Biol 210, 115–133.

Gerhardt, C., Wiegering, A., Leu, T., and Rüther, U. (2016). Control of Hedgehog Signalling by the Cilia-Regulated Proteasome. J Dev Biol 4, 27.

German, O.L., Buzzi, E., Rotstein, N.P., Rodríguez-Boulan, E., and Politi, L.E. (2008). Retinal pigment epithelial cells promote spatial reorganization and differentiation of retina photoreceptors. J Neurosci Res 86, 3503–3514.

Gerstein, M.B., Kundaje, A., Hariharan, M., Landt, S.G., Yan, K.-K., Cheng, C., Mu, X.J., Khurana, E., Rozowsky, J., Alexander, R., et al. (2012). Architecture of the human regulatory network derived from ENCODE data. Nature 489, 91–100.

Gherman, A., Davis, E.E., and Katsanis, N. (2006). The ciliary proteome database: an integrated community resource for the genetic and functional dissection of cilia. Nat Genet 38, 961–962.

Girard, S.L., Dion, P.A., Bourassa, C.V., Geoffroy, S., Lachance-Touchette, P., Barhdadi, A., Langlois, M., Joober, R., Krebs, M.-O., Dubé, M.-P., et al. (2015). Mutation burden of rare variants in schizophrenia candidate genes. PLoS ONE 10, e0128988.

Glish, G.L., and Vachet, R.W. (2003). The basics of mass spectrometry in the twenty-first century. Nat Rev Drug Discov 2, 140–150.

de Godoy, L.M.F., Olsen, J.V., Cox, J., Nielsen, M.L., Hubner, N.C., Fröhlich, F., Walther, T.C., and Mann, M. (2008). Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. Nature 455, 1251–1254.

Goeman, J.J., and Bühlmann, P. (2007). Analyzing gene expression data in terms of gene sets: methodological issues. Bioinformatics 23, 980–987.

Goetz, S.C., and Anderson, K.V. (2010). The primary cilium: a signalling centre during vertebrate development. Nat Rev Genet 11, 331–344.

Gold, L. (1995). Oligonucleotides as research, diagnostic, and therapeutic agents. J Biol Chem 270, 13581–13584.

Gold, L., Ayers, D., Bertino, J., Bock, C., Bock, A., Brody, E.N., Carter, J., Dalby, A.B., Eaton, B.E., Fitzwater, T., et al. (2010). Aptamer-based multiplexed proteomic technology for biomarker discovery. PLoS ONE 5, e15004.

Goldstein, D.B., Tate, S.K., and Sisodiya, S.M. (2003). Pharmacogenetics goes genomic. Nat Rev Genet 4, 937–947.

Gomez-Cabrero, D., Abugessaisa, I., Maier, D., Teschendorff, A., Merkenschlager, M., Gisel, A., Ballestar, E., Bongcam-Rudloff, E., Conesa, A., and Tegnér, J. (2014). Data integration in the era of omics: current and future challenges. BMC Syst Biol 8 Suppl 2, I1.

Gonzaga-Jauregui, C., Harel, T., Gambin, T., Kousi, M., Griffin, L.B., Francescatto, L., Ozes, B., Karaca, E., Jhangiani, S.N., Bainbridge, M.N., et al. (2015). Exome Sequence Analysis Suggests that Genetic Burden Contributes to Phenotypic Variability and Complex Neuropathy. Cell Rep 12, 1169–1183.

Gonzales, P.A., Pisitkun, T., Hoffert, J.D., Tchapyjnikov, D., Star, R.A., Kleta, R., Wang, N.S., and Knepper, M.A. (2009). Large-scale proteomics and phosphoproteomics of urinary exosomes. J Am Soc Nephrol 20, 363–379.

Gonzalez-Cordero, A., West, E.L., Pearson, R.A., Duran, Y., Carvalho, L.S., Chu, C.J., Naeem, A., Blackford, S.J.I., Georgiadis, A., Lakowski, J., et al. (2013). Photoreceptor precursors derived from three-dimensional embryonic stem cell cultures integrate and mature within adult degenerate retina. Nat Biotechnol 31, 741–747.

Goodwin, S., McPherson, J.D., and McCombie, W.R. (2016). Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet 17, 333–351.

Gorden, N.T., Arts, H.H., Parisi, M.A., Coene, K.L.M., Letteboer, S.J.F., van Beersum, S.E.C., Mans, D.A., Hikida, A., Eckert, M., Knutzen, D., et al. (2008). CC2D2A is mutated in Joubert syndrome and interacts with the ciliopathy-associated basal body protein CEP290. Am J Hum Genet 83, 559–571.

Gore, A., Li, Z., Fung, H.-L., Young, J.E., Agarwal, S., Antosiewicz-Bourget, J., Canto, I., Giorgetti, A., Israel, M.A., Kiskinis, E., et al. (2011). Somatic coding mutations in human induced pluripotent stem cells. Nature 471, 63–67.

Graser, S., Stierhof, Y.-D., Lavoie, S.B., Gassner, O.S., Lamla, S., Le Clech, M., and Nigg, E.A. (2007). Cep164, a novel centriole appendage protein required for primary cilium formation. J Cell Biol 179, 321–330.

Graw, J. (2003). The genetic and molecular basis of congenital eye defects. Nat Rev Genet 4, 876–888.

Graw, J. (2010). Eye development. Curr Top Dev Biol 90, 343–386.

Grayson, B.L., Wang, L., and Aune, T.M. (2011). Peripheral blood gene expression profiles in metabolic syndrome, coronary artery disease and type 2 diabetes. Genes Immun 12, 341–351.

Green, J.S., Parfrey, P.S., Harnett, J.D., Farid, N.R., Cramer, B.C., Johnson, G., Heath, O., McManamon, P.J., O'Leary, E., and Pryse-Phillips, W. (1989). The cardinal manifestations of Bardet-Biedl syndrome, a form of Laurence-Moon-Biedl syndrome. N Engl J Med 321, 1002–1009.

Griffin, N.M., Yu, J., Long, F., Oh, P., Shore, S., Li, Y., Koziol, J.A., and Schnitzer, J.E. (2010). Label-free, normalized quantification of complex mass spectrometry data for proteomic analysis. Nat Biotechnol 28, 83–89.

GTEx Portal (2017). Gene expression for PKHD1.

Gu, S.M., Thompson, D.A., Srikumari, C.R., Lorenz, B., Finckh, U., Nicoletti, A., Murthy, K.R., Rathmann, M., Kumaramanickavel, G., Denton, M.J., et al. (1997). Mutations in RPE65 cause autosomal recessive childhood-onset severe retinal dystrophy. Nat Genet 17, 194–197.

Guemez-Gamboa, A., Coufal, N.G., and Gleeson, J.G. (2014). Primary cilia in the developing and mature brain. Neuron 82, 511–521.

Guo, D.-F., and Rahmouni, K. (2011). Molecular basis of the obesity associated with Bardet-Biedl syndrome. Trends Endocrinol Metab 22, 286–293.

Guo, M.H., Dauber, A., Lippincott, M.F., Chan, Y.-M., Salem, R.M., and Hirschhorn, J.N. (2016). Determinants of Power in Gene-Based Burden Testing for Monogenic Disorders. Am J Hum Genet 99, 527–539.

Han, T.S., and Lean, M.E. (2016). A clinical perspective of obesity, metabolic syndrome and cardiovascular disease. JRSM Cardiovasc Dis 5, 2048004016633371.

Händel, M., Schulz, S., Stanarius, A., Schreff, M., Erdtmann-Vourliotis, M., Schmidt, H., Wolf, G., and Höllt, V. (1999). Selective targeting of somatostatin receptor 3 to neuronal cilia. Neuroscience 89, 909–926.

Happé, H., de Heer, E., and Peters, D.J.M. (2011). Polycystic kidney disease: the complexity of planar cell polarity and signaling during tissue regeneration and cyst formation. Biochim Biophys Acta 1812, 1249–1255.

Harnett, J.D., Green, J.S., Cramer, B.C., Johnson, G., Chafe, L., McManamon, P., Farid, N.R., Pryse-Phillips, W., and Parfrey, P.S. (1988). The spectrum of renal disease in Laurence-Moon-Biedl syndrome. N Engl J Med 319, 615–618.

Harper, A.R., Nayee, S., and Topol, E.J. (2015). Protective alleles and modifier variants in human health and disease. Nat Rev Genet 16, 689–701.

Hartong, D.T., Berson, E.L., and Dryja, T.P. (2006). Retinitis pigmentosa. The Lancet 368, 1795–1809.

Hathout, Y., Brody, E., Clemens, P.R., Cripe, L., DeLisle, R.K., Furlong, P., Gordish-Dressman, H., Hache, L., Henricson, E., Hoffman, E.P., et al. (2015). Large-scale serum protein biomarker discovery in Duchenne muscular dystrophy. Proc Natl Acad Sci U S A 112, 7153–7158.

Hattula, K., Furuhjelm, J., Arffman, A., and Peränen, J. (2002). A Rab8-specific GDP/GTP exchange factor is involved in actin remodeling and polarized membrane transport. Mol Biol Cell 13, 3268–3280.

Hawkridge, A.M., and Muddiman, D.C. (2009). Mass spectrometry-based biomarker discovery: toward a global proteome index of individuality. Annu Rev Anal Chem (Palo Alto Calif) 2, 265–277.

Haycraft, C.J., Zhang, Q., Song, B., Jackson, W.S., Detloff, P.J., Serra, R., and Yoder, B.K. (2007). Intraflagellar transport is essential for endochondral bone formation. Development 134, 307–316.

Heavner, W., and Pevny, L. (2012). Eye development and retinogenesis. Cold Spring Harb Perspect Biol 4.

Hendrickson, A., Bumsted-O'Brien, K., Natoli, R., Ramamurthy, V., Possin, D., and Provis, J. (2008). Rod photoreceptor differentiation in fetal and infant human retina. Exp Eye Res 87, 415–426.

Héon, E., Westall, C., Carmi, R., Elbedour, K., Panton, C., Mackeen, L., Stone, E.M., and Sheffield, V.C. (2005). Ocular phenotypes of three genetic variants of Bardet-Biedl syndrome. Am J Med Genet A 132A, 283–287.

Heon, E., Kim, G., Qin, S., Garrison, J.E., Tavares, E., Vincent, A., Nuangchamnong, N., Scott, C.A., Slusarski, D.C., and Sheffield, V.C. (2016). Mutations in C8ORF37 cause Bardet Biedl syndrome (BBS21). Hum Mol Genet 25, 2283–2294.

Hernandez-Hernandez, V., Pravincumar, P., Diaz-Font, A., May-Simera, H., Jenkins, D., Knight, M., and Beales, P.L. (2013). Bardet-Biedl syndrome proteins control the cilia length through regulation of actin polymerization. Hum Mol Genet 22, 3858–3868.

Hilgendorf, K.I., Johnson, C.T., and Jackson, P.K. (2016). The primary cilium as a cellular receiver: organizing ciliary GPCR signaling. Curr Opin Cell Biol 39, 84–92.

Hirami, Y., Osakada, F., Takahashi, K., Okita, K., Yamanaka, S., Ikeda, H., Yoshimura, N., and Takahashi, M. (2009). Generation of retinal cells from mouse and human induced pluripotent stem cells. Neurosci Lett 458, 126–131.

Hjortshøj, T.D., Grønskov, K., Philp, A.R., Nishimura, D.Y., Riise, R., Sheffield, V.C., Rosenberg, T., and Brøndum-Nielsen, K. (2010). Bardet-Biedl syndrome in Denmark--report of 13 novel sequence variations in six genes. Hum Mutat 31, 429–436.

Ho, C.S., Lam, C.W.K., Chan, M.H.M., Cheung, R.C.K., Law, L.K., Lit, L.C.W., Ng, K.F., Suen, M.W.M., and Tai, H.L. (2003). Electrospray ionisation mass spectrometry: principles and clinical applications. Clin Biochem Rev 24, 3–12.

Hoefele, J., Sudbrak, R., Reinhardt, R., Lehrack, S., Hennig, S., Imm, A., Muerb, U., Utsch, B., Attanasio, M., O'Toole, J.F., et al. (2005). Mutational analysis of the NPHP4 gene in 250 patients with nephronophthisis. Hum Mutat 25, 411.

Homma, K., Okamoto, S., Mandai, M., Gotoh, N., Rajasimha, H.K., Chang, Y.-S., Chen, S., Li, W., Cogliati, T., Swaroop, A., et al. (2013). Developing rods transplanted into the degenerating retina of Crx-knockout mice exhibit neural activity similar to native photoreceptors. Stem Cells 31, 1149–1159.

Horsford, D.J., Nguyen, M.-T.T., Sellar, G.C., Kothary, R., Arnheiter, H., and McInnes, R.R. (2005). Chx10 repression of Mitf is required for the maintenance of mammalian neuroretinal identity. Development 132, 177–187.

Hou, L., Li, F., Zeng, Q., Su, L., Chen, P., Xu, Z., Zhu, D., Liu, C., and Xu, D. (2014). Excretion of urinary orosomucoid 1 protein is elevated in patients with chronic heart failure. PLoS ONE 9, e107550.

Hu, Q., Friedrich, A.M., Johnson, L.V., and Clegg, D.O. (2010). Memory in induced pluripotent stem cells: reprogrammed human retinal-pigmented epithelial cells show tendency for spontaneous redifferentiation. Stem Cells 28, 1981–1991.

Huangfu, D., and Anderson, K.V. (2005). Cilia and Hedgehog responsiveness in the mouse. Proc Natl Acad Sci U S A 102, 11325–11330.

Huangfu, D., Liu, A., Rakeman, A.S., Murcia, N.S., Niswander, L., and Anderson, K.V. (2003). Hedgehog signalling in the mouse requires intraflagellar transport proteins. Nature 426, 83–87.

Hughes, C.S., Postovit, L.M., and Lajoie, G.A. (2010). Matrigel: a complex protein mixture required for optimal growth of cell culture. Proteomics 10, 1886–1890.

Human Protein Atlas (2017a). Tissue expression of C2orf71 - Staining in kidney - The Human Protein Atlas.

Human Protein Atlas (2017b). Tissue expression of PKHD1 - Staining in kidney - The Human Protein Atlas.

Hunter, D.J. (2005). Gene-environment interactions in human diseases. Nat Rev Genet 6, 287–298.

Huppke, P., Wegener, E., Böhrer-Rabel, H., Bolz, H.J., Zoll, B., Gärtner, J., and Bergmann, C. (2015). Tectonic gene mutations in patients with Joubert syndrome. Eur J Hum Genet 23, 616–620.

Ibraghimov-Beskrovnaya, O., and Natoli, T.A. (2011). mTOR signaling in polycystic kidney disease. Trends Mol Med 17, 625–633.

Imhoff, O., Marion, V., Stoetzel, C., Durand, M., Holder, M., Sigaudy, S., Sarda, P., Hamel, C.P., Brandt, C., Dollfus, H., et al. (2011). Bardet-Biedl syndrome: a study of the renal and cardiovascular phenotypes in a French cohort. Clin J Am Soc Nephrol 6, 22–29.

Inglis, P.N., Boroevich, K.A., and Leroux, M.R. (2006). Piecing together a ciliome. Trends Genet 22, 491–500.

Insinna, C., and Besharse, J.C. (2008). Intraflagellar transport and the sensory outer segment of vertebrate photoreceptors. Dev Dyn 237, 1982–1992.

Ionita-Laza, I., Buxbaum, J.D., Laird, N.M., and Lange, C. (2011). A new testing strategy to identify rare variants with either risk or protective effect on disease. PLoS Genet 7, e1001289.

Irizarry, R.A., Wang, C., Zhou, Y., and Speed, T.P. (2009). Gene set enrichment analysis made simple. Stat Methods Med Res 18, 565–575.

Ishikawa, H., and Marshall, W.F. (2011). Ciliogenesis: building the cell's antenna. Nat Rev Mol Cell Biol 12, 222–234.

Izawa, I., Goto, H., Kasahara, K., and Inagaki, M. (2015). Current topics of functional links between primary cilia and cell cycle. Cilia 4, 12.

Jafari, M., Primo, V., Smejkal, G.B., Moskovets, E.V., Kuo, W.P., and Ivanov, A.R. (2012). Comparison of in-gel protein separation techniques commonly used for fractionation in mass spectrometry-based proteomic profiling. Electrophoresis 33, 2516–2526.

Jaffe, A.S., Babuin, L., and Apple, F.S. (2006). Biomarkers in acute cardiac disease: the present and the future. J Am Coll Cardiol 48, 1–11.

James, P. (1997). Protein identification in the post-genome era: the rapid rise of proteomics. Q Rev Biophys 30, 279–331.

Jaros, J.A.J., Guest, P.C., Bahn, S., and Martins-de-Souza, D. (2013). Affinity depletion of plasma and serum for mass spectrometry-based proteome analysis. Methods Mol Biol 1002, 1–11.

Jenkins, A., Zhang, S.X., Gosmanova, A., Aston, C., Dashti, A., Baker, M.Z., Lyons, T., and Ma, J.-X. (2008). Increased serum pigment epithelium derived factor levels in Type 2 diabetes patients. Diabetes Res Clin Pract 82, e5-7.

Jeong, Y., Leskow, F.C., El-Jaick, K., Roessler, E., Muenke, M., Yocum, A., Dubourg, C., Li, X., Geng, X., Oliver, G., et al. (2008). Regulation of a remote Shh forebrain enhancer by the Six3 homeoprotein. Nat Genet 40, 1348–1353.

Ji, X., Kember, R.L., Brown, C.D., and Bućan, M. (2016). Increased burden of deleterious variants in essential genes in autism spectrum disorder. Proc Natl Acad Sci U S A 113, 15054–15059.

Jin, H., White, S.R., Shida, T., Schulz, S., Aguiar, M., Gygi, S.P., Bazan, J.F., and Nachury, M.V. (2010). The conserved Bardet-Biedl syndrome proteins assemble a coat that traffics membrane proteins to cilia. Cell 141, 1208–1219.

Jurczyk, A., Gromley, A., Redick, S., San Agustin, J., Witman, G., Pazour, G.J., Peters, D.J.M., and Doxsey, S. (2004). Pericentrin forms a complex with intraflagellar transport proteins and polycystin-2 and is required for primary cilia assembly. J Cell Biol 166, 637–643.

Käll, L., Canterbury, J.D., Weston, J., Noble, W.S., and MacCoss, M.J. (2007). Semi-supervised learning for peptide identification from shotgun proteomics datasets. Nat Methods 4, 923–925.

Kamińska, A., Platt, M., Kasprzyk, J., Kuśnierz-Cabala, B., Gala-Błądzińska, A., Woźnicka, O., Jany, B.R., Krok, F., Piekoszewski, W., Kuźniewski, M., et al. (2016). Urinary extracellular vesicles: potential biomarkers of renal function in diabetic patients. J Diabetes Res 2016, 5741518.

Kasahara, K., Kawakami, Y., Kiyono, T., Yonemura, S., Kawamura, Y., Era, S., Matsuzaki, F., Goshima, N., and Inagaki, M. (2014). Ubiquitin-proteasome system controls ciliogenesis at the initial step of axoneme extension. Nat Commun 5, 5081.

Katsanis, N. (2004). The oligogenic properties of Bardet-Biedl syndrome. Hum Mol Genet 13 Spec No 1, R65-71.

Katsanis, N., Beales, P.L., Woods, M.O., Lewis, R.A., Green, J.S., Parfrey, P.S., Ansley, S.J., Davidson, W.S., and Lupski, J.R. (2000). Mutations in MKKS cause obesity, retinal dystrophy and renal malformations associated with Bardet-Biedl syndrome. Nat Genet 26, 67–70.

Katsanis, N., Lupski, J.R., and Beales, P.L. (2001a). Exploring the molecular basis of Bardet-Biedl syndrome. Hum Mol Genet 10, 2293–2299.

Katsanis, N., Ansley, S.J., Badano, J.L., Eichers, E.R., Lewis, R.A., Hoskins, B.E., Scambler, P.J., Davidson, W.S., Beales, P.L., and Lupski, J.R. (2001b). Triallelic inheritance in Bardet-Biedl syndrome, a Mendelian recessive disorder. Science 293, 2256–2259.

Katsanis, N., Eichers, E.R., Ansley, S.J., Lewis, R.A., Kayserili, H., Hoskins, B.E., Scambler, P.J., Beales, P.L., and Lupski, J.R. (2002). BBS4 is a minor contributor to Bardet-Biedl syndrome and may also participate in triallelic inheritance. Am J Hum Genet 71, 22–29.

Keady, B.T., Le, Y.Z., and Pazour, G.J. (2011). IFT20 is required for opsin trafficking and photoreceptor outer segment development. Mol Biol Cell 22, 921–930.

Kenney, A.M., Cole, M.D., and Rowitch, D.H. (2003). Nmyc upregulation by sonic hedgehog signaling promotes proliferation in developing cerebellar granule neuron precursors. Development 130, 15–28.

Kentsis, A., Monigatti, F., Dorff, K., Campagne, F., Bachur, R., and Steen, H. (2009). Urine proteomics for profiling of human disease using high accuracy mass spectrometry. Proteomics Clin Appl 3, 1052–1061.

Khanna, H., Davis, E.E., Murga-Zamalloa, C.A., Estrada-Cuzcano, A., Lopez, I., den Hollander, A.I., Zonneveld, M.N., Othman, M.I., Waseem, N., Chakarova, C.F., et al. (2009). A common allele in RPGRIP1L is a modifier of retinal degeneration in ciliopathies. Nat Genet 41, 739–745.

Khatri, P., Sirota, M., and Butte, A.J. (2012). Ten years of pathway analysis: current approaches and outstanding challenges. PLoS Comput Biol 8, e1002375.

Kilpinen, H., Goncalves, A., Leha, A., Afzal, V., Alasoo, K., Ashford, S., Bala, S., Bensaddek, D., Casale, F.P., Culley, O.J., et al. (2017). Common genetic variation drives molecular heterogeneity in human iPSCs. Nature 546, 370–375.

Kim, D.-S., Ross, P.J., Zaslavsky, K., and Ellis, J. (2014a). Optimizing neuronal differentiation from induced pluripotent stem cells to model ASD. Front Cell Neurosci 8, 109.

Kim, H.S., Bernitz, J.M., Lee, D.-F., and Lemischka, I.R. (2014b). Genomic editing tools to model human diseases with isogenic pluripotent stem cells. Stem Cells Dev 23, 2673–2686.

Kim, J., Lee, J.E., Heynen-Genel, S., Suyama, E., Ono, K., Lee, K., Ideker, T., Aza-Blanc, P., and Gleeson, J.G. (2010a). Functional genomic screen for modulators of ciliogenesis and cilium length. Nature 464, 1048–1051.

Kim, J., Jo, H., Hong, H., Kim, M.H., Kim, J.M., Lee, J.-K., Heo, W.D., and Kim, J. (2015a). Actin remodelling factors control ciliogenesis by regulating YAP/TAZ activity and vesicle trafficking. Nat Commun 6, 6781.

Kim, J.C., Badano, J.L., Sibold, S., Esmail, M.A., Hill, J., Hoskins, B.E., Leitch, C.C., Venner, K., Ansley, S.J., Ross, A.J., et al. (2004). The Bardet-Biedl protein BBS4 targets cargo to the pericentriolar region and is required for microtubule anchoring and cell cycle progression. Nat Genet 36, 462–470.

Kim, K., Doi, A., Wen, B., Ng, K., Zhao, R., Cahan, P., Kim, J., Aryee, M.J., Ji, H., Ehrlich, L.I.R., et al. (2010b). Epigenetic memory in induced pluripotent stem cells. Nature 467, 285–290.

Kim, S., Lee, K., Choi, J.-H., Ringstad, N., and Dynlacht, B.D. (2015b). Nek2 activation of Kif24 ensures cilium disassembly during the cell cycle. Nat Commun 6, 8087.

Kim, S.K., Shindo, A., Park, T.J., Oh, E.C., Ghosh, S., Gray, R.S., Lewis, R.A., Johnson, C.A., Attie-Bittach, T., Katsanis, N., et al. (2010c). Planar cell polarity acts through septins to control collective cell movement and ciliogenesis. Science 329, 1337–1340.

Kimura, M., Maruyama, T., and Crow, J.F. (1963). The mutation load in small populations. Genetics 48, 1303–1312.

King, S.M. (2016). Axonemal Dynein Arms. Cold Spring Harb Perspect Biol 8.

Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet 46, 310–315.

Klein, D., and Ammann, F. (1969). The syndrome of Laurence-Moon-Bardet-Biedl and allied diseases in Switzerland. Clinical, genetic and epidemiological studies. J Neurol Sci 9, 479–513.

Knödler, A., Feng, S., Zhang, J., Zhang, X., Das, A., Peränen, J., and Guo, W. (2010). Coordination of Rab8 and Rab11 in primary ciliogenesis. Proc Natl Acad Sci U S A 107, 6346–6351.

Kobayashi, T., Tsang, W.Y., Li, J., Lane, W., and Dynlacht, B.D. (2011). Centriolar kinesin Kif24 interacts with CP110 to remodel microtubules and regulate ciliogenesis. Cell 145, 914–925.

Kodani, A., Yu, T.W., Johnson, J.R., Jayaraman, D., Johnson, T.L., Al-Gazali, L., Sztriha, L., Partlow, J.N., Kim, H., Krup, A.L., et al. (2015). Centriolar satellites assemble centrosomal microcephaly proteins to recruit CDK2 and promote centriole duplication. Elife 4.

Koemeter-Cox, A.I., Sherwood, T.W., Green, J.A., Steiner, R.A., Berbari, N.F., Yoder, B.K., Kauffman, A.S., Monsma, P.C., Brown, A., Askwith, C.C., et al. (2014). Primary cilia enhance kisspeptin receptor signaling on gonadotropin-releasing hormone neurons. Proc Natl Acad Sci U S A 111, 10335–10340.

Koo, J., Tammam, S., Ku, S.-Y., Sampaleanu, L.M., Burrows, L.L., and Howell, P.L. (2008). PilF is an outer membrane lipoprotein required for multimerization and localization of the Pseudomonas aeruginosa Type IV pilus secretin. J Bacteriol 190, 6961–6969.

Kopinke, D., Roberson, E.C., and Reiter, J.F. (2017). Ciliary Hedgehog Signaling Restricts Injury-Induced Adipogenesis. Cell 170, 340–351.e12.

Kousi, M., and Katsanis, N. (2015). Genetic modifiers and oligogenic inheritance. Cold Spring Harb Perspect Med 5.

Kozera, B., and Rapacz, M. (2013). Reference genes in real-time PCR. J Appl Genet 54, 391–406.

Kraemer, S., Vaught, J.D., Bock, C., Gold, L., Katilius, E., Keeney, T.R., Kim, N., Saccomano, N.A., Wilcox, S.K., Zichi, D., et al. (2011). From SOMAmer-based biomarker discovery to diagnostic and clinical applications: a SOMAmer-based, streamlined multiplex proteomic assay. PLoS ONE 6, e26332.

Krock, B.L., Mills-Henry, I., and Perkins, B.D. (2009). Retrograde intraflagellar transport by cytoplasmic dynein-2 is required for outer segment extension in vertebrate photoreceptors but not arrestin translocation. Invest Ophthalmol Vis Sci 50, 5463–5471.

Kuharev, J., Navarro, P., Distler, U., Jahn, O., and Tenzer, S. (2015). In-depth evaluation of software tools for data-independent acquisition based label-free quantification. Proteomics 15, 3140–3151.

Kulaga, H.M., Leitch, C.C., Eichers, E.R., Badano, J.L., Lesemann, A., Hoskins, B.E., Lupski, J.R., Beales, P.L., Reed, R.R., and Katsanis, N. (2004). Loss of BBS proteins causes anosmia in humans and defects in olfactory cilia structure and function in the mouse. Nat Genet 36, 994–998.

Kwitek-Black, A.E., Carmi, R., Duyk, G.M., Buetow, K.H., Elbedour, K., Parvari, R., Yandava, C.N., Stone, E.M., and Sheffield, V.C. (1993). Linkage of Bardet-Biedl syndrome to chromosome 16q and evidence for non-allelic genetic heterogeneity. Nat Genet 5, 392–396.

Lai, C.K., Gupta, N., Wen, X., Rangell, L., Chih, B., Peterson, A.S., Bazan, J.F., Li, L., and Scales, S.J. (2011). Functional characterization of putative cilia genes by high-content analysis. Mol Biol Cell 22, 1104–1119.

Lakowski, J., Han, Y.T., Pearson, R.A., Gonzalez-Cordero, A., West, E.L., Gualdoni, S., Barber, A.C., Hubank, M., Ali, R.R., and Sowden, J.C. (2011). Effective transplantation of photoreceptor precursor cells selected via cell surface antigen expression. Stem Cells 29, 1391–1404.

Lamba, D.A., Karl, M.O., Ware, C.B., and Reh, T.A. (2006). Efficient generation of retinal progenitor cells from human embryonic stem cells. Proc Natl Acad Sci U S A 103, 12769–12774.

Lamba, D.A., McUsic, A., Hirata, R.K., Wang, P.-R., Russell, D., and Reh, T.A. (2010). Generation, purification and transplantation of photoreceptors derived from human induced pluripotent stem cells. PLoS ONE 5, e8763.

Larance, M., and Lamond, A.I. (2015). Multidimensional proteomics for cell biology. Nat Rev Mol Cell Biol 16, 269–280.

Lawo, S., Bashkurov, M., Mullin, M., Ferreria, M.G., Kittler, R., Habermann, B., Tagliaferro, A., Poser, I., Hutchins, J.R.A., Hegemann, B., et al. (2009). HAUS, the 8-subunit human Augmin complex, regulates centrosome and spindle integrity. Curr Biol 19, 816–826.

Layer, R.M., Chiang C., Quinlan, A.R., Hall, I.M. (2014). LUMPY: a probabilistic framework for structural variant discovery. Genome Biol 15, R84.

Lechtreck, K.F. (2015). IFT-Cargo Interactions and Protein Transport in Cilia. Trends Biochem Sci 40, 765–778.

Lee, K.L., Guevarra, M.D., Nguyen, A.M., Chua, M.C., Wang, Y., and Jacobs, C.R. (2015). The primary cilium functions as a mechanical and calcium signaling nexus. Cilia 4, 7.

Lee, S., Abecasis, G.R., Boehnke, M., and Lin, X. (2014). Rare-variant association analysis: study designs and statistical tests. Am J Hum Genet 95, 5–23.

Leitch, C.C., Zaghloul, N.A., Davis, E.E., Stoetzel, C., Diaz-Font, A., Rix, S., Alfadhel, M., Lewis, R.A., Eyaid, W., Banin, E., et al. (2008). Hypomorphic mutations in syndromic encephalocele genes are associated with Bardet-Biedl syndrome. Nat Genet 40, 443–448.

Leitch, C.C., Lodh, S., Prieto-Echagüe, V., Badano, J.L., and Zaghloul, N.A. (2014). Basal body proteins regulate Notch signaling through endosomal trafficking. J Cell Sci 127, 2407–2419.

Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature 536, 285–291.

Leppert, M., Baird, L., Anderson, K.L., Otterud, B., Lupski, J.R., and Lewis, R.A. (1994). Bardet-Biedl syndrome is linked to DNA markers on chromosome 11q and is genetically heterogeneous. Nat Genet 7, 108–112.

Li, B., and Leal, S.M. (2008). Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. Am J Hum Genet 83, 311–321.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923–930.

Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P., and Mesirov, J.P. (2011). Molecular signatures database (MSigDB) 3.0. Bioinformatics 27, 1739–1740.

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst 1, 417–425.

Liem, K.F., He, M., Ocbina, P.J.R., and Anderson, K.V. (2009). Mouse Kif7/Costal2 is a cilia-associated protein that regulates Sonic hedgehog signaling. Proc Natl Acad Sci U S A 106, 13377–13382.

Liew, G.M., Ye, F., Nager, A.R., Murphy, J.P., Lee, J.S., Aguiar, M., Breslow, D.K., Gygi, S.P., and Nachury, M.V. (2014). The intraflagellar transport protein IFT27 promotes BBSome exit from cilia through the GTPase ARL6/BBS3. Dev Cell 31, 265–278.

Lilienbaum, A. (2013). Relationship between the proteasomal system and autophagy. Int J Biochem Mol Biol 4, 1–26.

Lin, W.-Y. (2016). Beyond Rare-Variant Association Testing: Pinpointing Rare Causal Variants in Case-Control Sequencing Study. Sci Rep 6, 21824.

Lindstrand, A., Davis, E.E., Carvalho, C.M.B., Pehlivan, D., Willer, J.R., Tsai, I.-C., Ramanathan, S., Zuppan, C., Sabo, A., Muzny, D., et al. (2014). Recurrent CNVs and SNVs at the NPHP1 locus contribute pathogenic alleles to Bardet-Biedl syndrome. Am J Hum Genet 94, 745–754.

Lindstrand, A., Frangakis, S., Carvalho, C.M.B., Richardson, E.B., McFadden, K.A., Willer, J.R., Pehlivan, D., Liu, P., Pediaditakis, I.L., Sabo, A., et al. (2016). Copy-Number Variation Contributes to the Mutational Load of Bardet-Biedl Syndrome. Am J Hum Genet 99, 318–336.

Lister, R., Pelizzola, M., Kida, Y.S., Hawkins, R.D., Nery, J.R., Hon, G., Antosiewicz-Bourget, J., O'Malley, R., Castanon, R., Klugman, S., et al. (2011). Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. Nature 471, 68–73.

Liu, A., Wang, B., and Niswander, L.A. (2005). Mouse intraflagellar transport proteins regulate both the activator and repressor functions of Gli transcription factors. Development 132, 3103–3111.

Liu, Q., Tan, G., Levenkova, N., Li, T., Pugh, E.N., Rux, J.J., Speicher, D.W., and Pierce, E.A. (2007). The proteome of the mouse photoreceptor sensory cilium complex. Mol Cell Proteomics 6, 1299–1317.

Liu, Y.P., Tsai, I.-C., Morleo, M., Oh, E.C., Leitch, C.C., Massa, F., Lee, B.-H., Parker, D.S., Finley, D., Zaghloul, N.A., et al. (2014). Ciliopathy proteins regulate paracrine signaling by modulating proteasomal degradation of mediators. J Clin Invest 124, 2059–2070.

Loktev, A.V., and Jackson, P.K. (2013). Neuropeptide Y family receptors traffic via the Bardet-Biedl syndrome pathway to signal in neuronal primary cilia. Cell Rep 5, 1316–1329.

Loktev, A.V., Zhang, Q., Beck, J.S., Searby, C.C., Scheetz, T.E., Bazan, J.F., Slusarski, D.C., Sheffield, V.C., Jackson, P.K., and Nachury, M.V. (2008). A BBSome subunit links ciliogenesis, microtubule stability, and acetylation. Dev Cell 15, 854–865.

Lollo, B., Steele, F., and Gold, L. (2014). Beyond antibodies: new affinity reagents to unlock the proteome. Proteomics 14, 638–644.

Louie, C.M., Caridi, G., Lopes, V.S., Brancati, F., Kispert, A., Lancaster, M.A., Schlossman, A.M., Otto, E.A., Leitges, M., Gröne, H.-J., et al. (2010). AHI1 is required for photoreceptor outer segment development and is a modifier for retinal degeneration in nephronophthisis. Nat Genet 42, 175–180.

Lowe, G., and Gold, G.H. (1991). The spatial distributions of odorant sensitivity and odorant-induced currents in salamander olfactory receptor cells. J Physiol (Lond) 442, 147–168.

Lower, K.M., Turner, G., Kerr, B.A., Mathews, K.D., Shaw, M.A., Gedeon, A.K., Schelley, S., Hoyme, H.E., White, S.M., Delatycki, M.B., et al. (2002). Mutations in PHF6 are associated with Börjeson-Forssman-Lehmann syndrome. Nat Genet 32, 661–665.

Lu, R., Ji, H., Chang, Z., Su, S., and Yang, G. (2010). Mitochondrial development and the influence of its dysfunction during rat adipocyte differentiation. Mol Biol Rep 37, 2173–2182.

Luby-Phelps, K., Fogerty, J., Baker, S.A., Pazour, G.J., and Besharse, J.C. (2008). Spatial distribution of intraflagellar transport proteins in vertebrate photoreceptors. Vision Res 48, 413–423.

Lüddecke, S., Ertych, N., Stenzinger, A., Weichert, W., Beissbarth, T., Dyczkowski, J., Gaedcke, J., Valerius, O., Braus, G.H., Kschischo, M., et al. (2016). The putative oncogene CEP72 inhibits the mitotic function of BRCA1 and induces chromosomal instability. Oncogene 35, 2398–2406.

Ludwig, T., and A Thomson, J. (2007). Defined, feeder-independent medium for human embryonic stem cell culture. Curr Protoc Stem Cell Biol Chapter 1, Unit 1C.2.

Lüftner, D., Lüke, C., and Possinger, K. (2003). Serum HER-2/neu in the management of breast cancer patients. Clin Biochem 36, 233–240.

Luo, D.-G., Xue, T., and Yau, K.-W. (2008). How vision begins: an odyssey. Proc Natl Acad Sci U S A 105, 9855–9862.

Lupski, J.R. (2010). New mutations and intellectual function. Nat Genet 42, 1036–1038.

Lupski, J.R., Belmont, J.W., Boerwinkle, E., and Gibbs, R.A. (2011). Clan genomics and the complex architecture of human disease. Cell 147, 32–43.

Ma, X., Peterson, R., and Turnbull, J. (2011). Adenylyl cyclase type 3, a marker of primary cilia, is reduced in primary cell culture and in lumbar spinal cord in situ in G93A SOD1 mice. BMC Neurosci 12, 71.

MacLaren, R.E., Pearson, R.A., MacNeil, A., Douglas, R.H., Salt, T.E., Akimoto, M., Swaroop, A., Sowden, J.C., and Ali, R.R. (2006). Retinal repair by transplantation of photoreceptor precursors. Nature 444, 203–207.

Maherali, N., Ahfeldt, T., Rigamonti, A., Utikal, J., Cowan, C., and Hochedlinger, K. (2008). A high-efficiency system for the generation and study of human induced pluripotent stem cells. Cell Stem Cell 3, 340–345.

Mahley, R. (1988). Apolipoprotein E: Cholesterol Transport Protein with Expanding Role in Cell Biology. *Science, 240*(4852), 622-630.

Mahoney, N.M., Goshima, G., Douglass, A.D., and Vale, R.D. (2006). Making microtubules and mitotic spindles in cells without functional centrosomes. Curr Biol 16, 564–569.

Malicki, J.J., and Johnson, C.A. (2017). The cilium: cellular antenna and central processing unit. Trends Cell Biol 27, 126–140.

Marion, V., Stoetzel, C., Schlicht, D., Messaddeq, N., Koch, M., Flori, E., Danse, J.M., Mandel, J.-L., and Dollfus, H. (2009). Transient ciliogenesis involving Bardet-Biedl syndrome proteins is a fundamental characteristic of adipogenic differentiation. Proc Natl Acad Sci U S A 106, 1820–1825.

Marion, V., Stutzmann, F., Gérard, M., De Melo, C., Schaefer, E., Claussmann, A., Hellé, S., Delague, V., Souied, E., Barrey, C., et al. (2012a). Exome sequencing identifies mutations in LZTFL1, a BBSome and smoothened trafficking regulator, in a family with Bardet--Biedl syndrome with situs inversus and insertional polydactyly. J Med Genet 49, 317–321.

Marion, V., Mockel, A., De Melo, C., Obringer, C., Claussmann, A., Simon, A., Messaddeq, N., Durand, M., Dupuis, L., Loeffler, J.-P., et al. (2012b). BBS-induced ciliary defect enhances adipogenesis, causing paradoxical higher-insulin sensitivity, glucose usage, and decreased inflammatory response. Cell Metab 16, 363–377.

Marmorstein, A.D., Finnemann, S.C., Bonilha, V.L., and Rodriguez-Boulan, E. (1998). Morphogenesis of the retinal pigment epithelium: toward understanding retinal degenerative diseases. Ann N Y Acad Sci 857, 1–12.

Marszalek, J.R., Liu, X., Roberts, E.A., Chui, D., Marth, J.D., Williams, D.S., and Goldstein, L.S. (2000). Genetic evidence for selective transport of opsin and arrestin by kinesin-II in mammalian photoreceptors. Cell 102, 175–187.

Masyuk, A.I., Gradilone, S.A., Banales, J.M., Huang, B.Q., Masyuk, T.V., Lee, S.-O., Splinter, P.L., Stroope, A.J., and Larusso, N.F. (2008). Cholangiocyte primary cilia are chemosensory organelles that detect biliary nucleotides via P2Y12 purinergic receptors. Am J Physiol Gastrointest Liver Physiol 295, G725-34.

Matthews, L., Gopinath, G., Gillespie, M., Caudy, M., Croft, D., de Bono, B., Garapati, P., Hemish, J., Hermjakob, H., Jassal, B., et al. (2009). Reactome knowledgebase of human biological pathways and processes. Nucleic Acids Res 37, D619-22.

Mazelova, J., Ransom, N., Astuto-Gribble, L., Wilson, M.C., and Deretic, D. (2009). Syntaxin 3 and SNAP-25 pairing, regulated by omega-3 docosahexaenoic acid, controls the delivery of rhodopsin for the biogenesis of cilia-derived sensory organelles, the rod outer segments. J Cell Sci 122, 2003–2013.

McCarthy, D.J., Chen, Y., and Smyth, G.K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res 40, 4288–4297.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20, 1297–1303.

Medina-Martinez, O., Amaya-Manzanares, F., Liu, C., Mendoza, M., Shah, R., Zhang, L., Behringer, R.R., Mahon, K.A., and Jamrich, M. (2009). Cell-autonomous requirement for rx function in the mammalian retina and posterior pituitary. PLoS ONE 4, e4513.

Meissner, F., and Mann, M. (2014). Quantitative shotgun proteomics: considerations for a high-quality workflow in immunology. Nat Immunol 15, 112–117.

Mellough, C.B., Sernagor, E., Moreno-Gimeno, I., Steel, D.H.W., and Lako, M. (2012). Efficient stage-specific differentiation of human pluripotent stem cells toward retinal photoreceptor cells. Stem Cells 30, 673–686.

Metzker, M.L. (2010). Sequencing technologies - the next generation. Nat Rev Genet 11, 31–46.

Metzler, M. (2011). Mutations in a novel cilia-centrosome protein cause a cystic kidney disease associated with retinal degeneration. Clin Genet 79, 222–224.

Meyer, J.S., Shearer, R.L., Capowski, E.E., Wright, L.S., Wallace, K.A., McMillan, E.L., Zhang, S.-C., and Gamm, D.M. (2009). Modeling early retinal development with human embryonic and induced pluripotent stem cells. Proc Natl Acad Sci U S A 106, 16698–16703.

Meyer, J.S., Howden, S.E., Wallace, K.A., Verhoeven, A.D., Wright, L.S., Capowski, E.E., Pinilla, I., Martin, J.M., Tian, S., Stewart, R., et al. (2011). Optic vesicle-like structures derived from human pluripotent stem cells facilitate a customized approach to retinal disease treatment. Stem Cells 29, 1206–1218.

Mi, H., Muruganujan, A., Casagrande, J.T., and Thomas, P.D. (2013). Large-scale gene function analysis with the PANTHER classification system. Nat Protoc 8, 1551–1566.

Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D., and Thomas, P.D. (2017). PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. Nucleic Acids Res 45, D183–D189.

Mick, D.U., Rodrigues, R.B., Leib, R.D., Adams, C.M., Chien, A.S., Gygi, S.P., and Nachury, M.V. (2015). Proteomics of primary cilia by proximity labeling. Dev Cell 35, 497–512.

Milligan, G., and Kostenis, E. (2006). Heterotrimeric G-proteins: a short history. Br J Pharmacol 147 Suppl 1, S46-55.

Millioni, R., Tolin, S., Puricelli, L., Sbrignadello, S., Fadini, G.P., Tessari, P., and Arrigoni, G. (2011). High abundance proteins depletion vs low abundance proteins enrichment: comparison of methods to reduce the plasma proteome complexity. PLoS ONE 6, e19603.

Mirhafez, S.R., Ebrahimi, M., Saberi Karimian, M., Avan, A., Tayefi, M., Heidari-Bakavoli, A., Parizadeh, M.R., Moohebati, M., Azarpazhooh, M.R., Esmaily, H., et al. (2016). Serum high-sensitivity C-reactive protein as a biomarker in patients with metabolic syndrome: evidence-based study with 7284 subjects. Eur J Clin Nutr 70, 1298–1304.

Mirza, S.P., Halligan, B.D., Greene, A.S., and Olivier, M. (2007). Improved method for the analysis of membrane proteins by mass spectrometry. Physiol Genomics 30, 89–94.

Mitchison, H.M., and Valente, E.M. (2017). Motile and non-motile cilia in human pathology: from function to phenotypes. J Pathol 241, 294–309.

Miyamoto, T., Hosoba, K., Ochiai, H., Royba, E., Izumi, H., Sakuma, T., Yamamoto, T., Dynlacht, B.D., and Matsuura, S. (2015). The Microtubule-Depolymerizing Activity of a Mitotic Kinesin Protein KIF2A Drives Primary Cilia Disassembly Coupled with Cell Proliferation. Cell Rep.

Mockel, A., Perdomo, Y., Stutzmann, F., Letsch, J., Marion, V., and Dollfus, H. (2011). Retinal dystrophy in Bardet-Biedl syndrome and related syndromic ciliopathies. Prog Retin Eye Res 30, 258–274.

Mockel, A., Obringer, C., Hakvoort, T.B.M., Seeliger, M., Lamers, W.H., Stoetzel, C., Dollfus, H., and Marion, V. (2012). Pharmacological modulation of the retinal unfolded protein response in Bardet-Biedl syndrome reduces apoptosis and preserves light detection ability. J Biol Chem 287, 37483–37494.

Mootha, V.K., Lindgren, C.M., Eriksson, K.-F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstråle, M., Laurila, E., et al. (2003). PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet 34, 267–273.

Moritz, O.L., Tam, B.M., Hurd, L.L., Peränen, J., Deretic, D., and Papermaster, D.S. (2001). Mutant rab8 Impairs docking and fusion of rhodopsin-bearing post-Golgi membranes and causes cell death of transgenic Xenopus rods. Mol Biol Cell 12, 2341–2351.

Morley, M., Molony, C.M., Weber, T.M., Devlin, J.L., Ewens, K.G., Spielman, R.S., and Cheung, V.G. (2004). Genetic analysis of genome-wide variation in human gene expression. Nature 430, 743-747.

Moseti, D., Regassa, A., and Kim, W.-K. (2016). Molecular Regulation of Adipogenesis and Potential Anti-Adipogenic Bioactive Molecules. Int J Mol Sci 17.

Motulsky, H. (2014). Intuitive biostatistics: A nonmathematical guide to statistical thinking (New York: Oxford University Press).

Mourão, A., Nager, A.R., Nachury, M.V., and Lorentzen, E. (2014). Structural basis for membrane targeting of the BBSome by ARL6. Nat Struct Mol Biol 21, 1035–1041.

Müller, F.-J., Schuldt, B.M., Williams, R., Mason, D., Altun, G., Papapetrou, E.P., Danner, S., Goldmann, J.E., Herbst, A., Schmidt, N.O., et al. (2011). A bioinformatic assay for pluripotency in human cells. Nat Methods 8, 315–317.

Mykytyn, K., Braun, T., Carmi, R., Haider, N.B., Searby, C.C., Shastri, M., Beck, G., Wright, A.F., Iannaccone, A., Elbedour, K., et al. (2001). Identification of the gene that, when mutated, causes the human obesity syndrome BBS4. Nat Genet 28, 188–191.

Mykytyn, K., Nishimura, D.Y., Searby, C.C., Shastri, M., Yen, H., Beck, J.S., Braun, T., Streb, L.M., Cornier, A.S., Cox, G.F., et al. (2002). Identification of the gene (BBS1) most commonly involved in Bardet-Biedl syndrome, a complex human obesity syndrome. Nat Genet 31, 435–438.

Mykytyn, K., Nishimura, D.Y., Searby, C.C., Beck, G., Bugge, K., Haines, H.L., Cornier, A.S., Cox, G.F., Fulton, A.B., Carmi, R., et al. (2003). Evaluation of complex inheritance involving the most common Bardet-Biedl syndrome locus (BBS1). Am J Hum Genet 72, 429–437.

Mykytyn, K., Mullins, R.F., Andrews, M., Chiang, A.P., Swiderski, R.E., Yang, B., Braun, T., Casavant, T., Stone, E.M., and Sheffield, V.C. (2004). Bardet-Biedl syndrome type 4 (BBS4)-null mice implicate Bbs4 in flagella formation but not global cilia assembly. Proc Natl Acad Sci U S A 101, 8664–8669.

Nachury, M.V., Loktev, A.V., Zhang, Q., Westlake, C.J., Peränen, J., Merdes, A., Slusarski, D.C., Scheller, R.H., Bazan, J.F., Sheffield, V.C., et al. (2007). A core complex of BBS proteins cooperates with the GTPase Rab8 to promote ciliary membrane biogenesis. Cell 129, 1201–1213.

Nagalakshmi, U., Waern, K., and Snyder, M. (2010). RNA-Seq: a method for comprehensive transcriptome analysis. Curr Protoc Mol Biol Chapter 4, Unit 4.11.1-13.

Nagaraj, N., Wisniewski, J.R., Geiger, T., Cox, J., Kircher, M., Kelso, J., Pääbo, S., and Mann, M. (2011). Deep proteome and transcriptome mapping of a human cancer cell line. Mol Syst Biol 7, 548.

Nakamura, K., Yamagishi, S.-I., Adachi, H., Kurita-Nakamura, Y., Matsui, T., and Inoue, H. (2009). Serum levels of pigment epithelium-derived factor (PEDF) are positively associated with visceral adiposity in Japanese patients with type 2 diabetes. Diabetes Metab Res Rev 25, 52–56.

Nakano, T., Ando, S., Takata, N., Kawada, M., Muguruma, K., Sekiguchi, K., Saito, K., Yonemura, S., Eiraku, M., and Sasai, Y. (2012). Self-formation of optic cups and storable stratified neural retina from human ESCs. Cell Stem Cell 10, 771–785.

Nathans, J., Thomas, D., and Hogness, D.S. (1986). Molecular genetics of human color vision: the genes encoding blue, green, and red pigments. Science 232, 193–202.

Nauli, S.M., and Zhou, J. (2004). Polycystins and mechanosensation in renal and nodal cilia. Bioessays 26, 844–856.

Nauli, S.M., Alenghat, F.J., Luo, Y., Williams, E., Vassilev, P., Li, X., Elia, A.E.H., Lu, W., Brown, E.M., Quinn, S.J., et al. (2003). Polycystins 1 and 2 mediate mechanosensation in the primary cilium of kidney cells. Nat Genet 33, 129–137.

Ng, P.C., and Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res 31, 3812–3814.

Nigg, E.A., and Stearns, T. (2011). The centrosome cycle: Centriole biogenesis, duplication and inherent asymmetries. Nat Cell Biol 13, 1154–1160.

Nilsson, T., Mann, M., Aebersold, R., Yates, J.R., Bairoch, A., and Bergeron, J.J.M. (2010). Mass spectrometry in high-throughput proteomics: ready for the big time. Nat Methods 7, 681–685.

Nishimura, D.Y., Searby, C.C., Carmi, R., Elbedour, K., Van Maldergem, L., Fulton, A.B., Lam, B.L., Powell, B.R., Swiderski, R.E., Bugge, K.E., et al. (2001). Positional cloning of a novel gene on chromosome 16q causing Bardet-Biedl syndrome (BBS2). Hum Mol Genet 10, 865–874.

Nishimura, D.Y., Fath, M., Mullins, R.F., Searby, C., Andrews, M., Davis, R., Andorf, J.L., Mykytyn, K., Swiderski, R.E., Yang, B., et al. (2004). Bbs2-null mice have neurosensory deficits, a defect in social dominance, and retinopathy associated with mislocalization of rhodopsin. Proc Natl Acad Sci U S A 101, 16588–16593.

Nishimura, D.Y., Swiderski, R.E., Searby, C.C., Berg, E.M., Ferguson, A.L., Hennekam, R., Merin, S., Weleber, R.G., Biesecker, L.G., Stone, E.M., et al. (2005). Comparative genomics and gene expression analysis identifies BBS9, a new Bardet-Biedl syndrome gene. Am J Hum Genet 77, 1021–1033.

Nishimura, D.Y., Baye, L.M., Perveen, R., Searby, C.C., Avila-Fernandez, A., Pereiro, I., Ayuso, C., Valverde, D., Bishop, P.N., Manson, F.D.C., et al. (2010). Discovery and functional analysis of a retinitis pigmentosa gene, C2ORF71. Am J Hum Genet 86, 686–695.

Oishi, I., Kawakami, Y., Raya, A., Callol-Massot, C., and Izpisúa Belmonte, J.C. (2006). Regulation of primary cilia formation and left-right patterning in zebrafish by a noncanonical Wnt signaling mediator, duboraya. Nat Genet 38, 1316–1322.

Omori, Y., Chaya, T., Yoshida, S., Irie, S., Tsujii, T., and Furukawa, T. (2015). Identification of G Protein-Coupled Receptors (GPCRs) in Primary Cilia and Their Possible Involvement in Body Weight Control. PLoS ONE 10, e0128422.

Omran, H., Sasmaz, G., Häffner, K., Volz, A., Olbrich, H., Melkaoui, R., Otto, E., Wienker, T.F., Korinthenberg, R., Brandis, M., et al. (2002). Identification of a gene locus for Senior-Løken syndrome in the region of the nephronophthisis type 3 gene. J Am Soc Nephrol 13, 75–79.

Osakada, F., Ikeda, H., Mandai, M., Wataya, T., Watanabe, K., Yoshimura, N., Akaike, A., Sasai, Y., and Takahashi, M. (2008). Toward the generation of rod and cone photoreceptors from mouse, monkey and human embryonic stem cells. Nat Biotechnol 26, 215–224.

O'Toole, J.F., Liu, Y., Davis, E.E., Westlake, C.J., Attanasio, M., Otto, E.A., Seelow, D., Nurnberg, G., Becker, C., Nuutinen, M., et al. (2010). Individuals with mutations in XPNPEP3, which encodes a mitochondrial protein, develop a nephronophthisis-like nephropathy. J Clin Invest 120, 791–802.

Ou, G., Blacque, O.E., Snow, J.J., Leroux, M.R., and Scholey, J.M. (2005). Functional coordination of intraflagellar transport motors. Nature 436, 583–587.

Ozsolak, F., and Milos, P.M. (2011). RNA sequencing: advances, challenges and opportunities. Nat Rev Genet 12, 87–98.

Pan, J., You, Y., Huang, T., and Brody, S.L. (2007). RhoA-mediated apical actin enrichment is required for ciliogenesis and promoted by Foxj1. J Cell Sci 120, 1868–1876.

Parfitt, D.A., Lane, A., Ramsden, C.M., Carr, A.-J.F., Munro, P.M., Jovanovic, K., Schwarz, N., Kanuga, N., Muthiah, M.N., Hull, S., et al. (2016). Identification and Correction of Mechanisms Underlying Inherited Blindness in Human iPSC-Derived Optic Cups. Cell Stem Cell 18, 769–781.

Parisi, M.A., Bennett, C.L., Eckert, M.L., Dobyns, W.B., Gleeson, J.G., Shaw, D.W.W., McDonald, R., Eddy, A., Chance, P.F., and Glass, I.A. (2004). The NPHP1 gene deletion associated with juvenile nephronophthisis is present in a subset of individuals with Joubert syndrome. Am J Hum Genet 75, 82–91.

Parisi, M.A., Doherty, D., Eckert, M.L., Shaw, D.W.W., Ozyurek, H., Aysun, S., Giray, O., Al Swaid, A., Al Shahwan, S., Dohayan, N., et al. (2006). AHI1 mutations cause both retinal dystrophy and renal cystic disease in Joubert syndrome. J Med Genet 43, 334–339.

Park, Y.-J., Yoo, S.-A., Hwang, D., Cho, C.-S., and Kim, W.-U. (2016). Identification of novel urinary biomarkers for assessing disease activity and prognosis of rheumatoid arthritis. Exp Mol Med 48, e211.

Parker, C.E., and Borchers, C.H. (2014). Mass spectrometry based biomarker discovery, verification, and validation--quality assurance and control of protein biomarker assays. Mol Oncol 8, 840–858.

Patil, S.B., Verma, R., Venkatareddy, M., and Khanna, H. (2010). Expression and localization of the ciliary disease protein retinitis pigmentosa GTPase regulator in mammalian kidney. Kidney Int 78, 622–623.

Pazour, G.J., Wilkerson, C.G., and Witman, G.B. (1998). A dynein light chain is essential for the retrograde particle movement of intraflagellar transport (IFT). J Cell Biol 141, 979–992.

Pazour, G.J., Baker, S.A., Deane, J.A., Cole, D.G., Dickert, B.L., Rosenbaum, J.L., Witman, G.B., and Besharse, J.C. (2002). The intraflagellar transport protein, IFT88, is essential for vertebrate photoreceptor assembly and maintenance. J Cell Biol 157, 103–113.

Pedersen, L.B., Veland, I.R., Schrøder, J.M., and Christensen, S.T. (2008). Assembly of primary cilia. Dev Dyn 237, 1993–2006.

Petrova, R., and Joyner, A.L. (2014). Roles for Hedgehog signaling in adult organ homeostasis and repair. Development 141, 3445–3457.

Petrovski, S., Wang, Q., Heinzen, E.L., Allen, A.S., and Goldstein, D.B. (2013). Genic intolerance to functional variation and the interpretation of personal genomes. PLoS Genet 9, e1003709.

Pietiläinen, K.H., Kaprio, J., Borg, P., Plasqui, G., Yki-Järvinen, H., Kujala, U.M., Rose, R.J., Westerterp, K.R., and Rissanen, A. (2008). Physical inactivity and obesity: a vicious circle. Obesity (Silver Spring) 16, 409–414.

Pirkmajer, S., and Chibalin, A.V. (2011). Serum starvation: caveat emptor. Am J Physiol, Cell Physiol 301, C272-9.

Plotnikova, O.V., Pugacheva, E.N., and Golemis, E.A. (2009). Primary cilia and the cell cycle. Methods Cell Biol 94, 137–160.

Plumb, R.S., Johnson, K.A., Rainville, P., Smith, B.W., Wilson, I.D., Castro-Perez, J.M., and Nicholson, J.K. (2006). UPLC/MS(E); a new approach for generating molecular fragment information for biomarker structure elucidation. Rapid Commun Mass Spectrom 20, 1989–1994.

Pocsfalvi, G., Raj, D.A.A., Fiume, I., Vilasi, A., Trepiccione, F., and Capasso, G. (2015). Urinary extracellular vesicles as reservoirs of altered proteins during the pathogenesis of polycystic kidney disease. Proteomics Clin Appl 9, 552–567.

Polo, J.M., Liu, S., Figueroa, M.E., Kulalert, W., Eminli, S., Tan, K.Y., Apostolou, E., Stadtfeld, M., Li, Y., Shioda, T., et al. (2010). Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. Nat Biotechnol 28, 848–855.

Poste, G. (2011). Bring on the biomarkers. Nature 469, 156–157.

Praetorius, H.A., and Spring, K.R. (2001). Bending the MDCK cell primary cilium increases intracellular calcium. J Membr Biol 184, 71–79.

Praetorius, H.A., and Spring, K.R. (2005). A physiological view of the primary cilium. Annu Rev Physiol 67, 515–529.

Qi, D., Brownridge, P., Xia, D., Mackay, K., Gonzalez-Galarza, F.F., Kenyani, J., Harman, V., Beynon, R.J., and Jones, A.R. (2012). A software toolkit and interface for performing stable isotope labeling and top3 quantification using Progenesis LC-MS. OMICS 16, 489–495.

Qiao, L., Zou, C., Shao, P., Schaack, J., Johnson, P.F., and Shao, J. (2008). Transcriptional regulation of fatty acid translocase/CD36 expression by CCAAT/enhancer-binding protein alpha. J Biol Chem 283, 8788–8795.

Quiring, R., Walldorf, U., Kloter, U., and Gehring, W.J. (1994). Homology of the eyeless gene of Drosophila to the Small eye gene in mice and Aniridia in humans. Science 265, 785–789.

Rachel, R.A., Li, T., and Swaroop, A. (2012a). Photoreceptor sensory cilia and ciliopathies: focus on CEP290, RPGR and their interacting proteins. Cilia 1, 22.

Rachel, R.A., May-Simera, H.L., Veleri, S., Gotoh, N., Choi, B.Y., Murga-Zamalloa, C., McIntyre, J.C., Marek, J., Lopez, I., Hackett, A.N., et al. (2012b). Combining Cep290 and Mkks ciliopathy alleles in mice rescues sensory defects and restores ciliogenesis. J Clin Invest 122, 1233–1245.

Ramachandran, H., Schäfer, T., Kim, Y., Herfurth, K., Hoff, S., Lienkamp, S.S., Kramer-Zucker, A., and Walz, G. (2014). Interaction with the Bardet-Biedl gene product TRIM32/BBS11 modifies the half-life and localization of Glis2/NPHP7. J Biol Chem 289, 8390–8401.

Ramsbottom, S., Miles, C., and Sayer, J. (2015). Murine Cep290 phenotypes are modified by genetic backgrounds and provide an impetus for investigating disease modifier alleles. [version 1; referees: 2 approved]. F1000Res 4, 590.

Ramsey, B.W., Davies, J., McElvaney, N.G., Tullis, E., Bell, S.C., Dřevínek, P., Griese, M., McKone, E.F., Wainwright, C.E., Konstan, M.W., et al. (2011). A CFTR potentiator in patients with cystic fibrosis and the G551D mutation. N Engl J Med 365, 1663–1672.

Rao, P.V., Lu, X., Standley, M., Pattee, P., Neelima, G., Girisesh, G., Dakshinamurthy, K.V., Roberts, C.T., and Nagalla, S.R. (2007). Proteomic identification of urinary biomarkers of diabetic nephropathy. Diabetes Care 30, 629–637.

Raposo, G., and Stoorvogel, W. (2013). Extracellular vesicles: exosomes, microvesicles, and friends. J Cell Biol 200, 373–383.

Redin, C., Le Gras, S., Mhamdi, O., Geoffroy, V., Stoetzel, C., Vincent, M.-C., Chiurazzi, P., Lacombe, D., Ouertani, I., Petit, F., et al. (2012). Targeted high-throughput sequencing for diagnosis of genetically heterogeneous diseases: efficient mutation detection in Bardet-Biedl and Alström syndromes. J Med Genet 49, 502–512.

Reichman, S., Terray, A., Slembrouck, A., Nanteau, C., Orieux, G., Habeler, W., Nandrot, E.F., Sahel, J.-A., Monville, C., and Goureau, O. (2014). From confluent human iPS cells to self-forming neural retina and retinal pigmented epithelium. Proc Natl Acad Sci U S A 111, 8518–8523.

Reiter, J.F., and Skarnes, W.C. (2006). Tectonic, a novel regulator of the Hedgehog pathway required for both activation and inhibition. Genes Dev 20, 22–27.

Reiter, J.F., Blacque, O.E., and Leroux, M.R. (2012). The base of the cilium: roles for transition fibres and the transition zone in ciliary formation, maintenance and compartmentalization. EMBO Rep 13, 608–618.

Riaz, S. (2015). Study of Protein Biomarkers of Diabetes Mellitus Type 2 and Therapy with Vitamin B1. J Diabetes Res 2015, 150176.

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., et al. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus

recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 17, 405–424.

Rifai, N., Gillette, M.A., and Carr, S.A. (2006). Protein biomarker discovery and validation: the long and uncertain path to clinical utility. Nat Biotechnol 24, 971–983.

Robert, A., Margall-Ducos, G., Guidotti, J.-E., Brégerie, O., Celati, C., Bréchot, C., and Desdouets, C. (2007). The intraflagellar transport component IFT88/polaris is a centrosomal protein regulating G1-S transition in non-ciliated cells. J Cell Sci 120, 628–637.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140.

Rohatgi, R., Milenkovic, L., and Scott, M.P. (2007). Patched1 regulates hedgehog signaling at the primary cilium. Science 317, 372–376.

Rosen, E.D., Hsu, C.-H., Wang, X., Sakai, S., Freeman, M.W., Gonzalez, F.J., and Spiegelman, B.M. (2002). C/EBPalpha induces adipogenesis through PPARgamma: a unified pathway. Genes Dev 16, 22–26.

Rosenbaum, M., and Leibel, R.L. (2014). 20 years of leptin: role of leptin in energy homeostasis in humans. J Endocrinol 223, T83-96.

Ross, A.J., May-Simera, H., Eichers, E.R., Kai, M., Hill, J., Jagger, D.J., Leitch, C.C., Chapple, J.P., Munro, P.M., Fisher, S., et al. (2005). Disruption of Bardet-Biedl syndrome ciliary proteins perturbs planar cell polarity in vertebrates. Nat Genet 37, 1135–1140.

Ross, S.E., Hemati, N., Longo, K.A., Bennett, C.N., Lucas, P.C., Erickson, R.L., and MacDougald, O.A. (2000). Inhibition of adipogenesis by Wnt signaling. Science 289, 950–953.

Ruiz i Altaba, A., Palma, V., and Dahmane, N. (2002). Hedgehog-Gli signalling and the growth of the brain. Nat Rev Neurosci 3, 24–33.

Ruiz-Perez, V.L., Blair, H.J., Rodriguez-Andres, M.E., Blanco, M.J., Wilson, A., Liu, Y.-N., Miles, C., Peters, H., and Goodship, J.A. (2007). Evc is a positive mediator of Ihh-regulated bone growth that localises at the base of chondrocyte cilia. Development 134, 2903–2912.

Rungarunlert, S., Techakumphu, M., Pirity, M.K., and Dinnyes, A. (2009). Embryoid body formation from embryonic and induced pluripotent stem cells: Benefits of bioreactors. World J Stem Cells 1, 11–21.

Ruotolo, G., Zoppo, A., Parlavecchia, M., Giberti, B., and Micossi, P. (1991). Apolipoprotein (a) levels in type 1 and type 2 diabetes mellitus. Acta Diabetol 28, 158–161.

Saga, T., Scheurer, D., and Adler, R. (1996). Development and maintenance of outer segments by isolated chick embryo photoreceptor cells in culture. Invest Ophthalmol Vis Sci 37, 561–573.

Sajic, T., Varesio, E., Szanto, I., and Hopfgartner, G. (2015). Comparison of fractionation strategies for offline two-dimensional liquid chromatography tandem mass spectrometry analysis of proteins from mouse adipose tissue. Anal Biochem 484, 122–132.

Salonen, R. (1984). The Meckel syndrome: clinicopathological findings in 67 patients. Am J Med Genet 18, 671–689.

Santos, N., and Reiter, J.F. (2008). Building it up and taking it down: the regulation of vertebrate ciliogenesis. Dev Dyn 237, 1972–1981.

Schachat, A.P., and Maumenee, I.H. (1982). Bardet-Biedl syndrome and related disorders. Arch Ophthalmol 100, 285–288.

Schaefer, E., Durand, M., Stoetzel, C., Doray, B., Viville, B., Hellé, S., Danse, J.-M., Hamel, C., Bitoun, P., Goldenberg, A., et al. (2011a). Molecular diagnosis reveals genetic heterogeneity for the overlapping MKKS and BBS phenotypes. Eur J Med Genet 54, 157–160.

Schaefer, E., Zaloszyc, A., Lauer, J., Durand, M., Stutzmann, F., Perdomo-Trujillo, Y., Redin, C., Bennouna Greene, V., Toutain, A., Perrin, L., et al. (2011b). Mutations in SDCCAG8/NPHP10 Cause Bardet-Biedl Syndrome and Are Associated with Penetrant Renal Disease and Absent Polydactyly. Mol Syndromol 1, 273–281.

Schaefer, E., Stoetzel, C., Scheidecker, S., Geoffroy, V., Prasad, M.K., Redin, C., Missotte, I., Lacombe, D., Mandel, J.-L., Muller, J., et al. (2016). Identification of a novel mutation confirms the implication of IFT172 (BBS20) in Bardet-Biedl syndrome. J Hum Genet 61, 447–450.

Scheidecker, S., Etard, C., Pierce, N.W., Geoffroy, V., Schaefer, E., Muller, J., Chennen, K., Flori, E., Pelletier, V., Poch, O., et al. (2014). Exome sequencing of Bardet-Biedl syndrome patient identifies a null mutation in the BBSome subunit BBIP1 (BBS18). J Med Genet 51, 132–136.

Schneider, L., Clement, C.A., Teilmann, S.C., Pazour, G.J., Hoffmann, E.K., Satir, P., and Christensen, S.T. (2005). PDGFRalphaalpha signaling is regulated through the primary cilium in fibroblasts. Curr Biol 15, 1861–1866.

Schou, K.B., Pedersen, L.B., and Christensen, S.T. (2015). Ins and outs of GPCR signaling in primary cilia. EMBO Rep 16, 1099–1113.

Schulze, W.X., and Usadel, B. (2010). Quantitation in mass-spectrometry-based proteomics. Annu Rev Plant Biol 61, 491–516.

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. Nature 473, 337–342.

Seo, S., Guo, D.-F., Bugge, K., Morgan, D.A., Rahmouni, K., and Sheffield, V.C. (2009). Requirement of Bardet-Biedl syndrome proteins for leptin receptor signaling. Hum Mol Genet 18, 1323–1331.

Seo, S., Baye, L.M., Schulz, N.P., Beck, J.S., Zhang, Q., Slusarski, D.C., and Sheffield, V.C. (2010). BBS6, BBS10, and BBS12 form a complex with CCT/TRiC family chaperonins and mediate BBSome assembly. Proc Natl Acad Sci U S A 107, 1488–1493.

Seo, S., Zhang, Q., Bugge, K., Breslow, D.K., Searby, C.C., Nachury, M.V., and Sheffield, V.C. (2011). A novel protein LZTFL1 regulates ciliary trafficking of the BBSome and Smoothened. PLoS Genet 7, e1002358.

Shaheen, R., Szymanska, K., Basu, B., Patel, N., Ewida, N., Faqeih, E., Al Hashem, A., Derar, N., Alsharif, H., Aldahmesh, M.A., et al. (2016). Characterizing the morbid genome of ciliopathies. Genome Biol 17, 242.

Shawky, R.M. (2014). Reduced penetrance in human inherited disease. Egyptian Journal of Medical Human Genetics 15, 103–111.

Shawky, R.M., Elsayed, S.M., Zaki, M.E., Nour El-Din, S.M., and Kamal, F.M. (2013). Consanguinity and its relevance to clinical genetics. Egyptian Journal of Medical Human Genetics 14, 157–164.

Sheffield, V.C., Carmi, R., Kwitek-Black, A., Rokhlina, T., Nishimura, D., Duyk, G.M., Elbedour, K., Sunden, S.L., and Stone, E.M. (1994). Identification of a Bardet-Biedl syndrome locus on chromosome 3 and evaluation of an efficient approach to homozygosity mapping. Hum Mol Genet 3, 1331–1335.

Shi, M., Movius, J., Dator, R., Aro, P., Zhao, Y., Pan, C., Lin, X., Bammler, T.K., Stewart, T., Zabetian, C.P., et al. (2015). Cerebrospinal fluid peptides as potential Parkinson disease biomarkers: a staged pipeline for discovery and validation. Mol Cell Proteomics 14, 544–555.

Shi, X., Burkart, A., Nicoloro, S.M., Czech, M.P., Straubhaar, J., and Corvera, S. (2008). Paradoxical effect of mitochondrial respiratory chain impairment on insulin signaling and glucose transport in adipose cells. J Biol Chem 283, 30658–30667.

Silva, S., Martins, Y., Matias, A., and Blickstein, I. (2011). Why are monozygotic twins different? J Perinat Med 39, 195–202.

Simons, M., Gloy, J., Ganner, A., Bullerkotte, A., Bashkurov, M., Krönig, C., Schermer, B., Benzing, T., Cabello, O.A., Jenny, A., et al. (2005). Inversin, the gene product mutated in nephronophthisis type II, functions as a molecular switch between Wnt signaling pathways. Nat Genet 37, 537–543.

Singla, V., and Reiter, J.F. (2006). The primary cilium as the cell's antenna: signaling at a sensory organelle. Science 313, 629–633.

Sinha, S., and Chen, J.K. (2006). Purmorphamine activates the Hedgehog pathway by targeting Smoothened. Nat Chem Biol 2, 29–30.

Slaats, G.G., Ghosh, A.K., Falke, L.L., Le Corre, S., Shaltiel, I.A., van de Hoek, G., Klasson, T.D., Stokman, M.F., Logister, I., Verhaar, M.C., et al. (2014). Nephronophthisis-associated CEP164 regulates cell cycle progression, apoptosis and epithelial-to-mesenchymal transition. PLoS Genet 10, e1004594.

Smith, T.S., Spitzbarth, B., Li, J., Dugger, D.R., Stern-Schneider, G., Sehn, E., Bolch, S.N., McDowell, J.H., Tipton, J., Wolfrum, U., et al. (2013). Light-dependent phosphorylation of Bardet-Biedl syndrome 5 in photoreceptor cells modulates its interaction with arrestin1. Cell Mol Life Sci 70, 4603–4616.

Soggiu, A., Piras, C., Bonizzi, L., Hussein, H.A., Pisanu, S., and Roncada, P. (2012). A discovery-phase urine proteomics investigation in type 1 diabetes. Acta Diabetol 49, 453–464.

Sokolov, M., Lyubarsky, A.L., Strissel, K.J., Savchenko, A.B., Govardovskii, V.I., Pugh, E.N., and Arshavsky, V.Y. (2002). Massive light-driven translocation of transducin between the two major compartments of rod cells: a novel mechanism of light adaptation. Neuron 34, 95–106.

Song, Z., Zhang, X., Jia, S., Yelick, P.C., and Zhao, C. (2016). Zebrafish as a model for human ciliopathies. J Genet Genomics 43, 107–120.

Soni, A., Amisten, S., Rorsman, P., and Salehi, A. (2013). GPRC5B a putative glutamate-receptor candidate is negative modulator of insulin secretion. Biochem Biophys Res Commun 441, 643–648.

Sorrell, J.M., and Caplan, A.I. (2004). Fibroblast heterogeneity: more than skin deep. J Cell Sci 117, 667–675.

Sorusch, N., Wunderlich, K., Bauss, K., Nagel-Wolfrum, K., and Wolfrum, U. (2014). Usher syndrome protein network functions in the retina and their relation to other retinal ciliopathies. Adv Exp Med Biol 801, 527–533.

de Sousa Abreu, R., Penalva, L.O., Marcotte, E.M., and Vogel, C. (2009). Global signatures of protein and mRNA expression levels. Mol Biosyst 5, 1512–1526.

Spektor, A., Tsang, W.Y., Khoo, D., and Dynlacht, B.D. (2007). Cep97 and CP110 suppress a cilia assembly program. Cell 130, 678–690.

Sridhar, A., Steward, M.M., and Meyer, J.S. (2013). Nonxenogeneic growth and retinal differentiation of human induced pluripotent stem cells. Stem Cells Translational Medicine 2, 255–264.

Stamey, T.A., Yang, N., Hay, A.R., McNeal, J.E., Freiha, F.S., and Redwine, E. (1987). Prostate-specific antigen as a serum marker for adenocarcinoma of the prostate. N Engl J Med 317, 909–916.

Steele, F.R., Chader, G.J., Johnson, L.V., and Tombran-Tink, J. (1993). Pigment epithelium-derived factor: neurotrophic activity and identification as a member of the serine protease inhibitor gene family. Proc Natl Acad Sci U S A 90, 1526–1530.

Stenson, P.D., Mort, M., Ball, E.V., Evans, K., Hayden, M., Heywood, S., Hussain, M., Phillips, A.D., and Cooper, D.N. (2017). The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. Hum Genet 136, 665–677.

Sterneckert, J.L., Reinhardt, P., and Schöler, H.R. (2014). Investigating human disease using stem cell models. Nat Rev Genet 15, 625–639.

Stoetzel, C., Laurier, V., Davis, E.E., Muller, J., Rix, S., Badano, J.L., Leitch, C.C., Salem, N., Chouery, E., Corbani, S., et al. (2006). BBS10 encodes a vertebrate-specific chaperonin-like protein and is a major BBS locus. Nat Genet 38, 521–524.

Stone, D.L., Slavotinek, A., Bouffard, G.G., Banerjee-Basu, S., Baxevanis, A.D., Barr, M., and Biesecker, L.G. (2000). Mutation of a gene encoding a putative chaperonin causes McKusick-Kaufman syndrome. Nat Genet 25, 79–82.

Streeter, I., Harrison, P.W., Faulconbridge, A., The HipSci Consortium, Flicek, P., Parkinson, H., and Clarke, L. (2017). The human-induced pluripotent stem cell initiative-data resources for cellular genetics. Nucleic Acids Res 45, D691–D697.

Strutz, F., Okada, H., Lo, C.W., Danoff, T., Carone, R.L., Tomaszewski, J.E., and Neilson, E.G. (1995). Identification and characterization of a fibroblast marker: FSP1. J Cell Biol 130, 393–405.

Su, X., Driscoll, K., Yao, G., Raed, A., Wu, M., Beales, P.L., and Zhou, J. (2014). Bardet-Biedl syndrome proteins 1 and 3 regulate the ciliary trafficking of polycystic kidney disease 1 protein. Hum Mol Genet 23, 5441–5451.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 102, 15545–15550.

Sun, X., Haley, J., Bulgakov, O.V., Cai, X., McGinnis, J., and Li, T. (2012). Tubby is required for trafficking G protein-coupled receptors to neuronal cilia. Cilia 1, 21.

Swaroop, A., Kim, D., and Forrest, D. (2010). Transcriptional regulation of photoreceptor development and homeostasis in the mammalian retina. Nat Rev Neurosci 11, 563–576.

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., et al. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res 43, D447-52.

Szymanska, K., and Johnson, C.A. (2012). The transition zone: an essential functional compartment of cilia. Cilia 1, 10.

Takacs, Z., and Proikas-Cezanne, T. (2016). Primary cilia mechanosensing triggers autophagy-regulated cell volume control. Nat Cell Biol 18, 591–592.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126, 663–676.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell 131, 861–872.

Theodoridis, G., Gika, H.G., and Wilson, I.D. (2008). LC-MS-based methodology for global metabolite profiling in metabonomics/metabolomics. TrAC Trends in Analytical Chemistry 27, 251–260.

Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S., and Jones, J.M. (1998). Embryonic stem cell lines derived from human blastocysts. Science 282, 1145–1147.

Timms, J.F., and Cutillas, P.R. (2010). Overview of quantitative LC-MS techniques for proteomics and activitomics. Methods Mol Biol 658, 19–45.

Tobin, J.L., and Beales, P.L. (2008). Restoration of renal function in zebrafish models of ciliopathies. Pediatr Nephrol 23, 2095–2099.

Tobin, J.L., and Beales, P.L. (2009). The nonmotile ciliopathies. Genet Med 11, 386–402.

Tombran-Tink, J., Chader, G.G., and Johnson, L.V. (1991). PEDF: A pigment epithelium-derived factor with potent neuronal differentiative activity. Exp Eye Res 53, 411–414.

Tong, C.K., Han, Y.-G., Shah, J.K., Obernier, K., Guinto, C.D., and Alvarez-Buylla, A. (2014). Primary cilia are required in a unique subpopulation of neural progenitors. Proc Natl Acad Sci U S A 111, 12438–12443.

Tonge, R., Fadgen, K., Stapels, M., Geromanos, S., Langridge, J., and Waters Corporation (2011). Reduction of co-fragmented peptides by ion mobility to improve peptide identification and quantitation.

Tontonoz, P., Hu, E., and Spiegelman, B.M. (1994). Stimulation of adipogenesis in fibroblasts by PPARγ2, a lipid-activated transcription factor. Cell 79, 1147–1156.

Topol, E.J. (2014). Individualized medicine from prewomb to tomb. Cell 157, 241–253.

Tory, K., Lacoste, T., Burglen, L., Morinière, V., Boddaert, N., Macher, M.-A., Llanas, B., Nivet, H., Bensman, A., Niaudet, P., et al. (2007). High NPHP1 and NPHP6 mutation rate in patients with Joubert syndrome and nephronophthisis: potential epistatic effect of NPHP6 and AHI1 mutations in patients with NPHP1 mutations. J Am Soc Nephrol 18, 1566–1575.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25, 1105–1111.

Tsang, W.Y., and Dynlacht, B.D. (2013). CP110 and its network of partners coordinately regulate cilia assembly. Cilia 2, 9.

Tsang, W.Y., Bossard, C., Khanna, H., Peränen, J., Swaroop, A., Malhotra, V., and Dynlacht, B.D. (2008). CP110 suppresses primary cilia formation through its interaction with CEP290, a protein deficient in human ciliary disease. Dev Cell 15, 187–197.

Tucker, B.A., Park, I.-H., Qi, S.D., Klassen, H.J., Jiang, C., Yao, J., Redenti, S., Daley, G.Q., and Young, M.J. (2011). Transplantation of adult mouse iPS cell-derived photoreceptor precursors restores retinal structure and function in degenerative mice. PLoS ONE 6, e18992.

Tucker, B.A., Mullins, R.F., Streb, L.M., Anfinson, K., Eyestone, M.E., Kaalberg, E., Riker, M.J., Drack, A.V., Braun, T.A., and Stone, E.M. (2013). Patient-specific iPSC-derived photoreceptor precursor cells as a means to investigate retinitis pigmentosa. Elife 2, e00824.

Tucker, R.W., Pardee, A.B., and Fujiwara, K. (1979). Centriole ciliation is related to quiescence and DNA synthesis in 3T3 cells. Cell 17, 527–535.

UCSC Genome Browser (2017). UCSC Genome Browser Home.

Uehara, R., Nozawa, R., Tomioka, A., Petry, S., Vale, R.D., Obuse, C., and Goshima, G. (2009). The augmin complex plays a critical role in spindle microtubule generation for mitotic progression and cytokinesis in human cells. Proc Natl Acad Sci U S A 106, 6998–7003.

Uetake, Y., Loncarek, J., Nordberg, J.J., English, C.N., La Terra, S., Khodjakov, A., and Sluder, G. (2007). Cell cycle progression and de novo centriole assembly after centrosomal removal in untransformed human cells. J Cell Biol 176, 173–182.

Uhlén, M., Fagerberg, L., Hallström, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., et al. (2015). Proteomics. Tissue-based map of the human proteome. Science 347, 1260419.

Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., and Leunissen, J.A.M. (2007). Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Res 35, W71-4.

US National Research Council (2011). Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease (Washington (DC): National Academies Press (US)).

Vajzovic, L., Hendrickson, A.E., O'Connell, R.V., Clark, L.A., Tran-Viet, D., Possin, D., Chiu, S.J., Farsiu, S. and Toth, C.A. 2012. Maturation of the human fovea: correlation of spectral-domain optical coherence tomography findings with histology. American Journal of Ophthalmology 154(5), pp. 779–789.e2.

Valente, E.M., Brancati, F., Silhavy, J.L., Castori, M., Marsh, S.E., Barrano, G., Bertini, E., Boltshauser, E., Zaki, M.S., Abdel-Aleem, A., et al. (2006). AHI1 gene mutations cause specific forms of Joubert syndrome-related disorders. Ann Neurol 59, 527–534.

Wagner, M.K., and Yost, H.J. (2000). Left-right development: the roles of nodal cilia. Curr Biol 10, R149-51.

Wallingford, J.B., and Mitchell, B. (2011). Strange as it may seem: the many links between Wnt signaling, planar cell polarity, and cilia. Genes Dev 25, 201–213.

Wang, C., Rüther, U., and Wang, B. (2007). The Shh-independent activator function of the full-length Gli3 protein and its role in vertebrate limb digit patterning. Dev Biol 305, 460–469.

Wang, H., Alvarez, S., and Hicks, L.M. (2012a). Comprehensive comparison of iTRAQ and label-free LC-based quantitative proteomics approaches using two Chlamydomonas reinhardtii strains of interest for biofuels engineering. J Proteome Res 11, 487–501.

Wang, L., Wang, S., and Li, W. (2012b). RSeQC: quality control of RNA-seq experiments. Bioinformatics 28, 2184–2185.

Wang, L., Meece, K., Williams, D.J., Lo, K.A., Zimmer, M., Heinrich, G., Martin Carli, J., Leduc, C.A., Sun, L., Zeltser, L.M., et al. (2015). Differentiation of hypothalamic-like neurons from human pluripotent stem cells. J Clin Invest 125, 796–808.

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10, 57–63.

Ward, L.D., and Kellis, M. (2012). Interpreting noncoding genetic variation in complex traits and human disease. Nat Biotechnol 30, 1095–1106.

Warnes, Gregory & Bolker, B & Bonebakker, L & Gentleman, R & Huber, W & Liaw, A & Lumley, T & Mächler, Martin & Magnusson, Arni & Möller, Steffen. (2005). gplots: Various R programming tools for plotting data. R package version. 2.

Waters, A.M., and Beales, P.L. (2011). Ciliopathies: an expanding disease spectrum. Pediatr Nephrol 26, 1039–1056.

Wätzlich, D., Vetter, I., Gotthardt, K., Miertzschke, M., Chen, Y.-X., Wittinghofer, A., and Ismail, S. (2013). The interplay between RPGR, PDEδ and Arl2/3 regulate the ciliary targeting of farnesylated cargo. EMBO Rep 14, 465–472.

Webber, W.A., and Lee, J. (1975). Fine structure of mammalian renal cilia. Anat Rec 182, 339–343.

Wei, Q., Zhang, Y., Li, Y., Zhang, Q., Ling, K., and Hu, J. (2012). The BBSome controls IFT assembly and turnaround in cilia. Nat Cell Biol 14, 950–957.

Wei, Q., Ling, K., and Hu, J. (2015). The essential roles of transition fibers in the context of cilia. Curr Opin Cell Biol 35, 98–105.

Wen, B., Zhou, R., Feng, Q., Wang, Q., Wang, J., and Liu, S. (2014). IQuant: an automated pipeline for quantitative proteomics based upon isobaric tags. Proteomics 14, 2280–2285.

Westfall, J.E., Hoyt, C., Liu, Q., Hsiao, Y.-C., Pierce, E.A., Page-McCaw, P.S., and Ferland, R.J. (2010). Retinal degeneration and failure of photoreceptor outer segment formation in mice with targeted deletion of the Joubert syndrome gene, Ahi1. J Neurosci 30, 8759–8768.

Westlake, C.J., Baye, L.M., Nachury, M.V., Wright, K.J., Ervin, K.E., Phu, L., Chalouni, C., Beck, J.S., Kirkpatrick, D.S., Slusarski, D.C., et al. (2011). Primary cilia membrane assembly is initiated by Rab11 and transport protein particle II (TRAPPII) complex-dependent trafficking of Rabin8 to the centrosome. Proc Natl Acad Sci U S A 108, 2759–2764.

Wetterstrand, K.A. (2016). DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP).

Wheway, G., Abdelhamed, Z., Natarajan, S., and Johnson, C.A. (2012). MKS1 interacts with components of the ubiquitin-proteasome pathway to regulate ciliogenesis and multiple signalling pathways. Cilia 1, P108.

Wheway, G., Schmidts, M., Mans, D.A., Szymanska, K., Nguyen, T.-M.T., Racher, H., Phelps, I.G., Toedt, G., Kennedy, J., Wunderlich, K.A., et al. (2015). An siRNA-based functional genomics screen for the identification of regulators of ciliogenesis and ciliopathy genes. Nat Cell Biol 17, 1074–1087.

Wiese, S., Reidegeld, K.A., Meyer, H.E., and Warscheid, B. (2007). Protein labeling by iTRAQ: a new tool for quantitative mass spectrometry in proteome research. Proteomics 7, 340–350.

Wilhelm, M., Schlegl, J., Hahne, H., Gholami, A.M., Lieberenz, M., Savitski, M.M., Ziegler, E., Butzmann, L., Gessulat, S., Marx, H., et al. (2014). Mass-spectrometry-based draft of the human proteome. Nature 509, 582–587.

Wolf, M.T.F., Saunier, S., O'Toole, J.F., Wanner, N., Groshong, T., Attanasio, M., Salomon, R., Stallmach, T., Sayer, J.A., Waldherr, R., et al. (2007). Mutational analysis of the RPGRIP1L gene in patients with Joubert syndrome and nephronophthisis. Kidney Int 72, 1520–1526.

Wolfrum, U., and Schmitt, A. (2000). Rhodopsin transport in the membrane of the connecting cilium of mammalian photoreceptor cells. Cell Motil Cytoskeleton 46, 95–107.

World Health Organisation (2008). Global Health Observatory data for obesity from 1980 - 2008.

Wray, N.R., Goddard, M.E., and Visscher, P.M. (2007). Prediction of individual genetic risk to disease from genome-wide association studies. Genome Res 17, 1520–1528.

Wright, A.F., Chakarova, C.F., Abd El-Aziz, M.M., and Bhattacharya, S.S. (2010). Photoreceptor degeneration: genetic and mechanistic dissection of a complex trait. Nat Rev Genet 11, 273–284.

Wright, F.A., Strug, L.J., Doshi, V.K., Commander, C.W., Blackman, S.M., Sun, L., Berthiaume, Y., Cutler, D., Cojocaru, A., Collaco, J.M., et al. (2011). Genome-wide association and linkage identify modifier loci of lung disease severity in cystic fibrosis at 11p13 and 20q13.2. Nat Genet 43, 539–546.

Wright, L.S., Phillips, M.J., Pinilla, I., Hei, D., and Gamm, D.M. (2014). Induced pluripotent stem cells as custom therapeutics for retinal repair: progress and rationale. Exp Eye Res 123, 161–172.

Xia, Y., Nivet, E., Sancho-Martinez, I., Gallegos, T., Suzuki, K., Okamura, D., Wu, M.-Z., Dubova, I., Esteban, C.R., Montserrat, N., et al. (2013). Directed differentiation of human pluripotent cells to ureteric bud kidney progenitor-like cells. Nat Cell Biol 15, 1507–1515.

Xu, Y., Cao, J., Huang, S., Feng, D., Zhang, W., Zhu, X., and Yan, X. (2015). Characterization of tetratricopeptide repeat-containing proteins critical for cilia formation and function. PLoS ONE 10, e0124378.

Xue, Y., Chen, Y., Ayub, Q., Huang, N., Ball, E.V., Mort, M., Phillips, A.D., Shaw, K., Stenson, P.D., Cooper, D.N., et al. (2012). Deleterious- and disease-allele prevalence in healthy individuals: insights from current predictions, mutation databases, and population-scale resequencing. Am J Hum Genet 91, 1022–1032.

Yamagishi, S.-I., Adachi, H., Abe, A., Yashiro, T., Enomoto, M., Furuki, K., Hino, A., Jinnouchi, Y., Takenaka, K., Matsui, T., et al. (2006). Elevated serum levels of pigment epithelium-derived factor in the metabolic syndrome. J Clin Endocrinol Metab 91, 2447–2450.

Yang, C., Yang, Y., Brennan, L., Bouhassira, E.E., Kantorow, M., and Cvekl, A. (2010). Efficient generation of lens progenitor cells and lentoid bodies from human embryonic stem cells in chemically defined conditions. FASEB J 24, 3274–3283.

Yoshida, Y., and Yamanaka, S. (2010). Recent stem cell advances: induced pluripotent stem cells for disease modeling and stem cell-based regeneration. Circulation 122, 80–87.

Young, T.L., Woods, M.O., Parfrey, P.S., Green, J.S., Hefferton, D., and Davidson, W.S. (1999a). A founder effect in the newfoundland population reduces the Bardet-Biedl syndrome I (BBS1) interval to 1 cM. Am J Hum Genet 65, 1680–1687.

Young, T.L., Penney, L., Woods, M.O., Parfrey, P.S., Green, J.S., Hefferton, D., and Davidson, W.S. (1999b). A fifth locus for Bardet-Biedl syndrome maps to chromosome 2q31. Am J Hum Genet 64, 900–904.

Youssoufian, H., and Pyeritz, R.E. (2002). Mechanisms and consequences of somatic mosaicism in humans. Nat Rev Genet 3, 748–758.

Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R., et al. (2007). Induced pluripotent stem cell lines derived from human somatic cells. Science 318, 1917–1920.

Yun, S., Saijoh, Y., Hirokawa, K.E., Kopinke, D., Murtaugh, L.C., Monuki, E.S., and Levine, E.M. (2009). Lhx2 links the intrinsic and extrinsic factors that control optic cup formation. Development 136, 3895–3906.

Zaghloul, N.A., and Katsanis, N. (2009). Mechanistic insights into Bardet-Biedl syndrome, a model ciliopathy. J Clin Invest 119, 428–437.

Zaghloul, N.A., and Katsanis, N. (2011). Zebrafish assays of ciliopathies. Methods Cell Biol 105, 257–272.

Zaghloul, N.A., Liu, Y., Gerdes, J.M., Gascue, C., Oh, E.C., Leitch, C.C., Bromberg, Y., Binkley, J., Leibel, R.L., Sidow, A., et al. (2010). Functional analyses of variants reveal a significant role for dominant negative and common alleles in oligogenic Bardet-Biedl syndrome. Proc Natl Acad Sci U S A 107, 10602–10607.

Zaki, M.S., Sattar, S., Massoudi, R.A., and Gleeson, J.G. (2011). Co-occurrence of distinct ciliopathy diseases in single families suggests genetic modifiers. Am J Med Genet A 155A, 3042–3049.

Zhang, M.-Z., Mai, W., Li, C., Cho, S., Hao, C., Moeckel, G., Zhao, R., Kim, I., Wang, J., Xiong, H., et al. (2004). PKHD1 protein encoded by the gene for autosomal recessive polycystic kidney disease associates with basal bodies and primary cilia in renal epithelial cells. Proc Natl Acad Sci U S A 101, 2311–2316.

Zhang, Q., Nishimura, D., Seo, S., Vogel, T., Morgan, D.A., Searby, C., Bugge, K., Stone, E.M., Rahmouni, K., and Sheffield, V.C. (2011). Bardet-Biedl syndrome 3 (Bbs3) knockout mouse model reveals common BBS-associated phenotypes and Bbs3 unique phenotypes. Proc Natl Acad Sci U S A 108, 20678–20683.

Zhang, Q., Yu, D., Seo, S., Stone, E.M., and Sheffield, V.C. (2012a). Intrinsic protein-protein interaction-mediated and chaperonin-assisted sequential assembly of stable bardet-biedl syndrome protein complex, the BBSome. J Biol Chem 287, 20625–20635.

Zhang, Q., Nishimura, D., Vogel, T., Shao, J., Swiderski, R., Yin, T., Searby, C., Carter, C.S., Kim, G., Bugge, K., et al. (2013a). BBS7 is required for BBSome formation and its absence in mice results in Bardet-Biedl syndrome phenotypes and selective abnormalities in membrane protein trafficking. J Cell Sci 126, 2372–2380.

Zhang, X., Bi, E., Novick, P., Du, L., Kozminski, K.G., Lipschutz, J.H., and Guo, W. (2001). Cdc42 interacts with the exocyst and regulates polarized secretion. J Biol Chem 276, 46745–46750.

Zhang, X., Xiao, Z., Liu, X., Du, L., Wang, L., Wang, S., Zheng, N., Zheng, G., Li, W., Zhang, X., et al. (2012b). The potential role of ORM2 in the development of colorectal cancer. PLoS ONE 7, e31868.

Zhang, Y., Marsboom, G., Toth, P.T., and Rehman, J. (2013b). Mitochondrial respiration regulates adipogenic differentiation of human mesenchymal stem cells. PLoS ONE 8, e77077.

Zhang, Y., Seo, S., Bhattarai, S., Bugge, K., Searby, C.C., Zhang, Q., Drack, A.V., Stone, E.M., and Sheffield, V.C. (2014). BBS mutations modify phenotypic expression of CEP290-related ciliopathies. Hum Mol Genet 23, 40–51.

Zhang, Z., Qin, X., Tong, N., Zhao, X., Gong, Y., Shi, Y., and Wu, X. (2012c). Valproic acid-mediated neuroprotection in retinal ischemia injury via histone deacetylase inhibition and transcriptional activation. Exp Eye Res 94, 98–108.

Zhao, J., Akinsanmi, I., Arafat, D., Cradick, T.J., Lee, C.M., Banskota, S., Marigorta, U.M., Bao, G., and Gibson, G. (2016). A Burden of Rare Variants Associated with Extremes of Gene Expression in Human Peripheral Blood. Am J Hum Genet 98, 299–309.

Zhao, Z., Wu, F., Ding, S., Sun, L., Liu, Z., Ding, K., and Lu, J. (2015). Label-free quantitative proteomic analysis reveals potential biomarkers and pathways in renal cell carcinoma. Tumour Biol 36, 939–951.

Zheng, L., Feng, Y., Shi, Y., Zhang, J., Mu, Q., Qin, L., Berggren-Söderlund, M., Nilsson-Ehle, P., Zhang, X., Luo, G., et al. (2014). Intralipid decreases apolipoprotein M levels and insulin sensitivity in rats. PLoS ONE 9, e105681.

Zhong, X., Gutierrez, C., Xue, T., Hampton, C., Vergara, M.N., Cao, L.-H., Peters, A., Park, T.S., Zambidis, E.T., Meyer, J.S., et al. (2014). Generation of three-dimensional retinal tissue with functional photoreceptors from human iPSCs. Nat Commun 5, 4047.

Zhou, H., Di Palma, S., Preisinger, C., Peng, M., Polat, A.N., Heck, A.J.R., and Mohammed, S. (2013). Toward a comprehensive characterization of a human cancer cell phosphoproteome. J Proteome Res 12, 260–271.

Zhu, W., Smith, J.W., and Huang, C.-M. (2010). Mass spectrometry-based label-free quantitative proteomics. J Biomed Biotechnol 2010, 840518.

Zimmerman, W.C., Sillibourne, J., Rosa, J., and Doxsey, S.J. (2004). Mitosis-specific anchoring of gamma tubulin complexes by pericentrin controls spindle organization and mitotic entry. Mol Biol Cell 15, 3642–3657.

Zuber, M.E., Gestri, G., Viczian, A.S., Barsacchi, G., and Harris, W.A. (2003). Specification of the vertebrate eye by a network of eye field transcription factors. Development 130, 5155–5167.