# 5 Discussion

In this dissertation I have described three projects that use next generation sequencing (NGS) to identify variants that can cause rare developmental disorders, along with statistical or functional follow-up approaches. The aim of the project described in chapter 2 was to explore how well exome sequencing performs as a method for identifying variants that cause abnormal fetal development. Exome sequencing of 30 parent-fetus trios was performed, where the fetuses had structural abnormalities. I identified single nucleotide variants (SNVs), insertion deletions (indels), and copy number variants (CNVs) with *de novo*, autosomal recessive, or X-linked (for male fetuses) inheritance in this cohort. I investigated various methods of variant prioritisation and interpretation, and concluded that for 3/30 fetuses (10%) a causal mutation had been identified. All of these were *de novo*, emphasising the importance of sequencing trios, and showing that there is a low recurrence risk for future pregnancies of these couples. Only one of these three mutations was a CNV and could therefore have been detected by microarray, the highest resolution genome-wide method currently used in prenatal genetic diagnostics. No novel disease-associated genes were identified during this study, because it was underpowered for this due to the small cohort size, and diversity of the fetal phenotypes. Nevertheless, this study demonstrates the utility of exome sequencing for prenatal genetic diagnosis, and paves the way for similar, larger studies. Issues that would need to be addressed before exome sequencing could become widely used for prenatal genetic diagnostics include the development and implementation of a primarily computational variant interpretation pipeline, and resolution of some contentious ethical issues. Based on this project, it seems clear that NGS is the future of prenatal diagnostics.

In chapter 3, I described a targeted resequencing study that was performed on a cohort of patients with intellectual disability (ID) as part of the UK10K project. I designed and implemented an analytical pipeline to identify variants that were likely to be causative. The first aim of this project was to identify causal variants in known ID-associated genes in the cohort. Using my pipeline, and further interpretation of variants by clinical collaborators, likely causative variants in known ID-associated genes were found for 14% of the cohort. The second aim was to attempt to identify any novel ID-associated

genes. We found causative *de novo* loss of function mutations in the putative histone methyltransferase gene *SETD5* in seven patients with ID, and showed that loss of function of *SETD5* is probably responsible for many features of 3p25 microdeletion syndrome, as well as being a relatively common cause of sporadic ID. This finding also emphasises the importance of methyltransferases in the pathology of ID. The final aim of the targeted resequencing study described in chapter 3 was to ascertain whether there is a burden of variants in ID-associated genes in ID patients compared to controls. I used the cohort allelic sums test to demonstrate that there is a burden of both loss of function variants, and some categories of missense variants, in ID-associated genes in ID patients compared to controls. This project demonstrates the importance of rigorous statistical methods in assigning causality to a gene associated with a rare developmental disorder. It also shows how case-control enrichment analyses can be a valuable statistical follow-up approaches to NGS, as it can focus attention on specific classes of variant with a higher likelihood of being pathogenic.

The aims of the project described in chapter 4 were to make zebrafish models of dystroglycanopathy using morpholino oligonucleotides to inhibit the expression of the two candidate dystroglycanopathy-associated genes *B3GALNT2* and *GMPPB*, and to determine the extent to which the phenotype of these models recapitulated the phenotypes of the patients. I first showed that zebrafish embryos are appropriate models for *B3GALNT2* and *GMPPB*, and that morpholinos do inhibit their expression. I next used several assays including immunofluorescence staining, Evans blue dye, and immunoblotting to determine the phenotype of the models compared to wildtype embryos. I found that there were similarities between the zebrafish models and the patients in terms of gross appearance and behaviour (such as movement defects), muscle structure (such as disordered fibres), and molecular level (hypoglycosylation of α-DG). This phenotype data from these two zebrafish models, together with clinical patient data and cellular models, led to the conclusion that variants in *B3GALNT2* and *GMPPB* can indeed cause dystroglycanopathy.

The work described in this dissertation has four important outcomes that I think directly or indirectly could improve the lives of patients affected by rare developmental disorders. All of this work was done in association with colleagues, collaborators and supervisors, as described in the Acknowledgements section and at the relevant portions of the text. First, a genetic cause was identified for 10% of the cohort of 30 fetuses with structural abnormalities, and for 14% of the UK10K ID cohort. Where the results were returned to the families, this ended the ‘diagnostic odyssey’ for them, and

could allow their clinicians to estimate recurrence risk for future pregnancies. Furthermore, it revealed some insights into the pathology of rare developmental disorders, for example the importance of *de novo* mutations in abnormal fetal development. Second, we demonstrated that exome sequencing is a promising tool for prenatal genetic diagnostics, and may be better than the current gold-standard method. This cohort of 30 fetuses with structural abnormalities is the largest such cohort published to date, nonetheless, my findings represent a 'proof-of-principle' study that paves the way for the even larger-scale evaluations of NGS for prenatal genetic diagnosis that are needed to both accurately quantify the diagnostic yield and identify novel genetic causes of fetal abnormalities. Third, the case-control enrichment analyses of ID patients in the UK10K study revealed some interesting findings into the genetic architecture of ID, many of which support findings that have previously been shown by other methods. For example, *de novo* mutations are an important cause of ID, and there is a burden of variants in some candidate genes that have not yet been conclusively associated with ID. The most interesting and novel finding from these analyses was that there is an enrichment of certain categories of missense variant, such as predicted-damaging X-linked variants, in ID-associated genes in ID patients compared to controls. Finally, and perhaps most importantly, the work described in this dissertation has contributed towards the discovery of three novel developmental disease-associated genes: *B3GALNT2* and *GMPPB* in dystroglycanopathy, and *SETD5* in ID. This will have a direct impact upon patients who have disease caused by damaging variants in these genes, as now they are more likely to receive a genetic diagnosis. Additionally, it improves understanding of the disease, for example *SETD5* emphasises the importance of appropriate histone methylation in normal cognitive functioning.

In conclusion, the three projects described in this dissertation highlight the importance of NGS for understanding rare developmental disorders. NGS, whether it is exome sequencing, whole genome sequencing, or targeted resequencing of candidate genes, has proved to be a valuable tool for clinical diagnosis of rare developmental disorders, and for the discovery of novel disease-associated genes. Often, statistical or functional follow-up approaches are required to confirm that variants in a particular gene do cause the disorder. An increasing number of genes that are associated with rare developmental disorders are being identified through the use of NGS. As progress continues to be made in this area, the focus of the research community is likely to shift towards understanding the precise mechanisms by which variants in a given gene

cause a rare developmental disorder, so that ultimately therapies for these disorders might be developed. It is likely that statistical follow-up approaches such as case-control enrichment analyses, and functional follow-up approaches such as modelling candidate genes in an organism such as the zebrafish, will be valuable for increasing this understanding. It is clear that NGS, along with supplementary and follow-up approaches, are both directly and indirectly improving the lives of patients with rare developmental disorders, and will continue to do so for the foreseeable future.